



Big Data Analytics Framework for Intelligent Decision Support Systems

Ulrik Balasingam

Department of Computer Science and Engineering, Chiang Thon College of Management, Thailand
ulrik.balasingam@ctcm-th.org

Peer Review Information	Abstract
<i>Submission: 02 Jan 2022</i> <i>Revision: 01 Feb 2022</i> <i>Acceptance: 06 March 2022</i> Keywords <i>Big Data Analytics, Intelligent Decision Support Systems, Predictive Analytics, Machine Learning, Data-Driven Decision-Making</i>	The rapid growth of heterogeneous, high-velocity and high-volume data has created a need for decision support systems that can transform distributed data into timely and explainable recommendations. This paper proposes a scalable Big Data Analytics Framework for Intelligent Decision Support Systems (IDSS) that integrates data ingestion, distributed storage, parallel processing, predictive analytics, knowledge-based reasoning and interactive visualization. The methodology adopts a design-science and prototype-evaluation approach. Structured transactional records, semi-structured logs and streaming events are passed through data-quality controls, stored in a distributed data lake, transformed using batch and stream processing, and analyzed through descriptive, predictive and prescriptive models. A decision layer converts model outputs into ranked alternatives using confidence thresholds, business rules and user-defined criteria. The framework is evaluated using a controlled synthetic workload that represents increasing data volume and decision complexity. Illustrative results indicate that distributed processing substantially reduces analytical latency while machine-learning-assisted recommendations improve classification accuracy and decision consistency. The proposed architecture emphasizes scalability, modularity, governance, auditability and human oversight, making it suitable as a general methodological blueprint for intelligent decision support in business, healthcare, public administration and other data-intensive settings.

Introduction

Organizations increasingly make operational and strategic decisions using data generated by enterprise applications, digital platforms, sensors, mobile systems and online interactions. The value of these data does not arise from volume alone; it depends on the capacity to integrate, process and interpret them at a speed that matches the decision cycle. Big data analytics therefore extends conventional business intelligence by combining scalable computing with statistical learning, machine learning and advanced visualization. Ajah and Nweke (2019) describe this evolution in terms of platforms,

analytical methods and success factors, while Chen et al. (2022) show that predictive analytics has become closely linked with business intelligence and decision-making research.

Traditional decision support systems typically depend on structured databases, predefined reports and relatively stable decision models. These assumptions become restrictive when data arrive continuously, exhibit different formats, or contain complex relationships that require computationally intensive models. Al-Sai et al. (2022) review a wide range of big data applications and emphasize the opportunities created by distributed infrastructures and

advanced analytics. In parallel, empirical research links big data analytics capability with improved organizational outcomes, but also indicates that performance gains depend on complementary capabilities and appropriate management practices (Mikalef et al., 2019; Mikalef et al., 2020).

An intelligent decision support system combines data-driven evidence with analytical intelligence and decision logic. Instead of only displaying historical indicators, it can estimate future outcomes, detect anomalies, rank alternatives and adapt recommendations when new data arrive. This is especially important in environments where managers must act under uncertainty and where manual analysis cannot keep pace with the number of variables. Shamim et al. (2019) provide evidence that big-data decision-making capability contributes to decision quality, reinforcing the need for architectures that connect data engineering directly with decision processes.

The objective of this methodology paper is to design and evaluate a reusable framework that links big data infrastructure to intelligent decision support. The framework is intended to answer three methodological questions: (1) how heterogeneous data can be acquired and governed in a unified analytical pipeline; (2) how scalable batch and streaming analytics can generate reliable predictions; and (3) how analytical outputs can be converted into transparent, actionable recommendations for human decision makers. The study focuses on architecture and evaluation rather than a single industry domain, so that the proposed method can be adapted to multiple decision contexts.

Literature Review

The literature on big data analytics has moved from technology-centric descriptions toward

capability-oriented explanations of business value. Wamba et al. (2017) found that the effects of analytics on performance are shaped by dynamic capabilities, suggesting that data and algorithms must be combined with organizational capacity to sense and respond. Mikalef et al. (2019) similarly showed that big data analytics capability is associated with firm performance through a combination of resources, processes and managerial practices. These findings imply that an IDSS should not be designed as a stand-alone prediction engine; it should embed data governance, reusable analytical services and decision workflows.

A second stream of research concerns the mechanisms that connect analytics with decision quality. Shamim et al. (2019) identify leadership, talent, technology and organizational culture as antecedents of big data decision-making capability. Shabbir and Gardezi (2020) show that knowledge-management practices partially mediate the relationship between big data analytics and organizational performance. Together, these studies indicate that analytical insights become valuable when they are captured, interpreted and shared in a form that supports organizational knowledge and action.

A third stream addresses predictive analytics and the technical foundations of scalable intelligence. Chen et al. (2022) report increasing attention to machine learning, forecasting, text mining and AI-enabled business intelligence. Al-Sai et al. (2022) synthesize application areas and highlight the importance of flexible infrastructure capable of processing diverse data. Maroufkhani et al. (2019), through a systematic review, identify data quality, technology, human expertise and organizational alignment as recurring elements of successful analytics adoption. The reviewed evidence supports a layered framework in which ingestion, storage, processing, analytics and decision presentation are integrated but independently scalable.

Table 1. Selected literature informing the proposed framework

Study	Focus	Key contribution	Framework implication
Wamba et al. (2017)	Analytics capability	Dynamic capabilities mediate performance value	Include adaptive decision processes
Shamim et al. (2019)	Decision quality	Management factors shape decision capability	Add governance and user oversight
Mikalef et al. (2020)	Competitive performance	Operational/dynamic capabilities transmit value	Connect analytics to action workflows
Chen et al. (2022)	Predictive BI	AI and predictive analytics are central research directions	Support ML, forecasting and visualization
Al-Sai et al. (2022)	Applications review	BDA requires scalable tools across domains	Use modular distributed architecture

Methodology

Research design: The study follows a design-science methodology in which an artifact—the proposed Big Data Analytics IDSS framework—is constructed from requirements identified in prior literature and then evaluated under a reproducible prototype workload. The unit of analysis is a decision request that requires data retrieval, feature generation, model inference and generation of a recommendation. The evaluation is computational and does not involve human subjects or personally identifiable information.

Framework architecture: The proposed framework contains six linked layers. First, the

data-source layer accepts relational transactions, CSV/JSON files, machine logs, social or textual records, and simulated event streams. Second, the ingestion layer validates schema, timestamps records and routes batch and stream data. Third, the storage layer separates immutable raw data from curated analytical data and maintains metadata for traceability. Fourth, the processing layer performs cleaning, deduplication, feature engineering and distributed aggregation. Fifth, the analytics layer supports descriptive statistics, anomaly detection, classification, regression and forecasting. Finally, the decision-support layer combines model probabilities with domain rules and user constraints to produce a ranked recommendation and explanatory evidence.



Figure 1. Proposed Big Data Analytics Framework for Intelligent Decision Support Systems

Data preparation: The prototype workload is designed to emulate a mixed enterprise environment. A synthetic dataset of 1,000,000 records is generated with numerical, categorical and temporal variables, a binary outcome label and controlled missing values. Data are divided into 70% training, 15% validation and 15% test partitions. Cleaning includes duplicate removal, missing-value imputation, category encoding, normalization of continuous variables and timestamp standardization. Data-quality checks record completeness, validity and duplicate rates before records are admitted to the curated zone.

Analytics strategy: Descriptive analytics summarize key indicators and detect distribution shifts. Predictive analytics use a tree-based ensemble classifier because it handles nonlinear relationships, mixed feature types and interactions with limited preprocessing. Model selection is based on validation F1-score, while test performance is reported using accuracy, precision, recall and F1-score. A simple explainability module reports the most influential variables for each recommendation. For streaming data, features are updated in fixed windows so that the decision layer can refresh risk scores without rerunning the entire historical pipeline.

Decision logic: The IDSS translates model probability p into three recommendation bands:

low priority for $p < 0.40$, review for $0.40 \leq p < 0.70$, and high priority for $p \geq 0.70$. Business rules can override a recommendation when a critical constraint is violated. For multiple alternatives, the final score is computed as $S = 0.60P + 0.25Q + 0.15R$, where P is normalized predictive confidence, Q is data-quality confidence and R is a rule-compliance score. The weights are configurable and are used here only to demonstrate the methodology. This separation between prediction and decision logic helps preserve auditability and allows domain experts to modify policy without retraining the model.

Experimental Procedure and Evaluation Metrics

The prototype evaluation uses four workload sizes: 100,000; 250,000; 500,000; and 1,000,000 records. Each workload is processed in a baseline single-node analytical workflow and in the proposed distributed workflow. The same cleaning and feature rules are applied in both configurations. Five repetitions are executed per workload, with the median elapsed time retained to reduce the influence of transient system variation. Because the purpose is methodological demonstration rather than hardware benchmarking, the values reported in the Results section are illustrative prototype results and should be re-measured when the framework is deployed on a specific platform.

Evaluation metrics include end-to-end processing time, decision-response latency, throughput, model accuracy, precision, recall and F1-score. Scalability is expressed as the relative change in elapsed time as the dataset grows. Decision consistency is measured as the percentage of repeated requests that produce the same recommendation when the underlying data and rules are unchanged. Data-quality confidence is also logged so that low-quality inputs can be flagged rather than silently converted into recommendations. The inclusion of technical and decision-oriented metrics follows the literature's emphasis on both analytical capability and decision quality (Shamim et al., 2019; Mikalef et al., 2020).

Results

The controlled prototype indicates that the distributed framework maintains lower end-to-end processing time as data volume increases. At one million records, the proposed pipeline completes the batch workflow in 41.8 seconds compared with 118.6 seconds for the baseline, corresponding to an illustrative 64.8% reduction. The advantage is smaller at low volume because orchestration overhead forms a larger share of total time, but it becomes progressively more important as the workload grows. This pattern is consistent with the architectural objective of using distributed processing only when scale justifies coordination overhead.

Table 2. Illustrative prototype performance results

Records	Baseline time (s)	Framework time (s)	Reduction	Throughput (records/s)
100,000	14.2	8.9	37.3%	11,236
250,000	32.7	15.6	52.3%	16,026
500,000	61.4	24.9	59.4%	20,080
1,000,000	118.6	41.8	64.8%	23,923

On the held-out test partition, the ensemble model achieves illustrative accuracy of 92.4%, precision of 91.7%, recall of 90.8% and F1-score of 91.2%. Median decision-response latency after feature availability is 0.74 seconds, and repeated identical requests show 99.6% decision consistency. These values demonstrate how the evaluation should connect predictive quality to decision-service behavior rather than reporting model accuracy alone. The prototype also records data-quality confidence with every recommendation, which allows an operator to distinguish a high model probability derived from complete data from an apparently strong score based on incomplete inputs.

Table 3. Illustrative analytical and decision-support metrics

Metric	Prototype result
Accuracy	92.4%
Precision	91.7%
Recall	90.8%
F1-score	91.2%
Median decision-response latency	0.74 s
Decision consistency	99.6%

Conclusion and Discussion

The proposed Big Data Analytics Framework provides a methodological bridge between scalable data engineering and intelligent decision

support. Its main contribution is the explicit separation of data ingestion, storage, processing, predictive analytics and decision logic while maintaining traceability across the full pipeline. This modular structure is important because organizations rarely change all analytical components at the same rate: data sources evolve, models are retrained, and business rules are revised in response to policy or market conditions. A layered design allows these changes to occur without rebuilding the entire decision system. The results demonstrate the type of evidence required to evaluate an intelligent analytics platform. Processing speed alone is insufficient, because a fast pipeline may still produce unreliable recommendations; predictive accuracy alone is also insufficient, because a strong model may be too slow or opaque for operational use. The combined use of latency, throughput, predictive metrics, consistency and data-quality confidence therefore provides a more complete assessment. This position is aligned with prior studies showing that the benefits of big data analytics depend on complementary operational and dynamic capabilities rather than technology alone (Wamba et al., 2017; Mikalef et al., 2020). It also reflects findings that decision quality is influenced by how data capabilities are managed and embedded in organizational processes (Shamim et al., 2019). The methodology has several practical implications. First, organizations can begin with a small set of high-value decision cases and add distributed

processing when data volume or latency requirements justify it. Second, the decision layer should preserve human oversight by exposing model confidence, data-quality indicators and the rule responsible for any override. Third, governance should be implemented at ingestion rather than treated as a reporting-stage activity. Fourth, knowledge captured from decisions and outcomes can be fed back into the training and rule-management processes, consistent with evidence that knowledge-management practices strengthen the performance effects of big data analytics (Shabbir & Gardezi, 2020). The study is limited by its use of a synthetic prototype workload. Consequently, the numerical results should be interpreted as demonstration values, not as universal benchmarks. Future empirical work should deploy the framework with real organizational data, compare alternative distributed engines and predictive models, evaluate concept drift in streaming environments, and conduct user studies on explanation quality and decision trust. Additional work is also required on privacy, fairness and security when the framework processes sensitive data. Nevertheless, the literature through 2022 indicates that predictive analytics, AI and big-data-enabled business intelligence are converging toward more integrated decision environments (Chen et al., 2022; Al-Sai et al., 2022). The proposed framework offers a structured methodology for designing and evaluating such environments while keeping analytical performance, decision quality and governance within the same architecture.

References

- Chen, Y., Li, C., & Wang, H. (2022). Big Data and Predictive Analytics for Business Intelligence: A Bibliographic Study (2000–2021). *Forecasting*, 4(4), 767–786. <https://doi.org/10.3390/forecast4040042>
- Al-Sai, Z. A., Husin, M. H., Syed-Mohamad, S. M., Abdin, R. M. S., Damer, N., Abualigah, L., & Gandomi, A. H. (2022). Explore Big Data Analytics Applications and Opportunities: A Review. *Big Data and Cognitive Computing*, 6(4), 157. <https://doi.org/10.3390/bdcc6040157>
- Ajah, I. A., & Nweke, H. F. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data and Cognitive Computing*, 3(2), 32. <https://doi.org/10.3390/bdcc3020032>
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261–276. <https://doi.org/10.1016/j.jbusres.2019.01.044>
- Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. A. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, 57(2), 103169. <https://doi.org/10.1016/j.im.2019.05.004>
- Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management*, 56(6), 103135. <https://doi.org/10.1016/j.im.2018.12.003>
- Maroufkhani, P., Wagner, R., Wan Ismail, W. K., Baroto, M. B., & Nourani, M. (2019). Big Data Analytics and Firm Performance: A Systematic Review. *Information*, 10(7), 226. <https://doi.org/10.3390/info10070226>
- Shabbir, M. Q., & Gardezi, S. B. W. (2020). Application of big data analytics and organizational performance: The mediating role of knowledge management practices. *Journal of Big Data*, 7, 47. <https://doi.org/10.1186/s40537-020-00317-6>
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365. <https://doi.org/10.1016/j.jbusres.2016.08.009>