SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 2 | February 2026 - Peer Reviewed Journal

TOWARDS INTELLIGENT AND DISTRIBUTED BUSINESS PROCESS ANALYTICS: A CLOUD-ENABLED BIG DATA FRAMEWORK

Mrs. Shaik Shabana Parveen[1], Prof. V. Tulasi Das[2]

[1]*Research Scholar, Dept. of MBA-HRM, Acharya Nagarjuna University, Guntur-522510, A.P.*

[2]*Professor, Dept. of MBA-HRM, Acharya Nagarjuna University, Guntur, A.P.*

Article DOI: https://doi.org/10.36713/epra26218
DOI No: 10.36713/epra26218

ABSTRACT

Business process improvement is a difficult endeavour that calls for sophisticated and reliable supporting technology. Behavioural pattern identification, predictive analytics, business activity monitoring, business intelligence systems, and "type simulations" are some of the new technologies and advanced analytics techniques that can help business users continuously improve their operations. However, business users are unable to effectively obtain timely analytics data due to the large amounts of event data generated by process execution during the business lifetime. This article offers a big data-based technology solution that gives business analysts insight into distributed processes and business performance. Users may quickly monitor business performance in highly distributed situation.

KEYWORDS: *Business Analytics, Big Data, Business Performance, Information Management, Data Storage Systems, Cloud Computing.*

1. INTRODUCTION

Real-time data analysis can be difficult due to the massive and complicated supply chains business process executions, which generate large volumes of unstructured event data. An architecture for incorporating big data, Seshathiri Dhanasekaran et. al., (2018) analytics into business performance management aids users in assessing and enhancing the effectiveness of business processes.

Businesses frequently find themselves managing very extensive process model repositories, which reflect important knowledge about their operations, as they advance in business process management (BPM) maturity. Because they dictate the method for creating value and delivering it to clients, business processes have grown in significance in many organizations. Moreover, these procedures are the main forces behind three crucial success factors: time, quality, and cost.

The Confederation of Indian Industry (CII) and ISO 9001 are two popular quality models that emphasize the significance of process orientation. Businesses frequently employ analytics, mining, or process intelligence, using a range of statistical and artificial intelligence methods to quantify and examine data pertaining to processes. The three categories of business process analysis (BPA) Vera-Baquero et.al., (2013) are validation, verification, and performance, according to Will van der Aalst and his associates. Each of these categories necessitates gathering and preserving substantial amounts of process and event data.

In this case, we concentrate on events, which, within the framework of a business process, indicate state changes in objects. Although BPA Vera-Baquero et.al., (2013) solutions frequently use the Business Process Analytics Format (BPAF) standard, no widely accepted format for sharing business events between distributed event producers and consumers has arisen, despite the significance of events for event-driven BPM and BPA. In a number of proposals, business process events and execution results are analysed using BPAF.

Big data-based approaches will be widely used to produce deep business insights, according to Bogdan Danov, Daniela Gotseva (2024). The recent surge in process event data and new business intelligence trends necessitate the adoption of new BPA methodologies. Big data opens up new avenues for BPM Bogdan Danov (2024) research, particularly evidence-based BPM, where actual data may be used to objectively assess study findings. On a bigger scale, process mining serves as the gap between big data analysis and business process management by attempting to relate event data to process models. Our suggested design will assist users in quickly analysing the results of business-process execution by combining big data analytics and BPM in a distributed setting.

The goal of our cloud-based architecture is to give business users more insight into corporate performance and processes. The ability to gather, aggregate, and store execution data results in a suitable format for subsequent measurement and analysis will enable them to keep an eye on business process executions from operational systems. In order to increase organizational effectiveness, users can enhance their processes and gain a deeper understanding of business performance by evaluating event data. Additionally, such data aids analysts in assessing the present situation, forecasting the behaviour of future process instances, and comprehending prior events.

However, employing traditional methods to manage company information successfully is difficult and difficult. Analytical applications require event data integration, which is challenging to accomplish in highly distributed environments where business activities are a part of intricate supply chains that are typically run under a number of heterogeneous systems. Additionally, because traditional systems are unable to handle the hundreds of millions of connected records, the continuous execution of distributed business processes generates a massive volume of event data that they are unable to manage effectively. Similarly, centralized solutions are unsuitable due to the substantial delay between the event's occurrence and its recording in central repositories.

Due to these limitations, current methods like the Framework for Business Process Analytics (F4BPA) cannot deliver real-time business analytics in environments with a significant degree of dispersion. Furthermore, rather than relying on a single centralized process owner to monitor and manage performance at individual supply chain nodes, we usually deal with highly distributed supply chains, where stakeholders are geographically separated and require a platform to perform BPM collaboratively. Therefore, we suggest expanding the framework with cloud-based infrastructure and utilizing data warehousing and distributed query processing to complement Stefano Rizzi's federative technique.

2. FRAMEWORK ARCHITECTURE

The local business analytics service unit (BASU) component, which is managed by each organizational unit, is connected to the big data-powered local event repository and other operational business systems. Apache Hadoop and HBase were used to create this repository, and Hive was added to offer data warehousing capabilities over huge data Thesnath A/L Velayudhan (2024).

Each organization may execute BPA both individually and cooperatively thanks to these local components, which conduct distributed queries over the network. Organizations can also assess the effectiveness of cross-functional business processes that transcend organizational boundaries by integrating BASU subsystems. The BASU components are integrated by the global business analytics service (GBAS), which serves as the hub for offering analytical services to applications developed by third parties.

Cloud computing services Srinivasa Rao Gundu et. al., (2023) can be delivered at extremely low latency response rates thanks to the entire architecture (see Figure 1). Through the provision of a rich, educational environment that promotes BPA and provides lucid insights into the efficacy and efficiency of organizational processes, these services can aid in the ongoing improvement of business processes. Additionally, a variety of analytical applications, including collaborative analytics, business activity monitoring, simulation engines, and real-time business intelligence systems, can make use of these services.

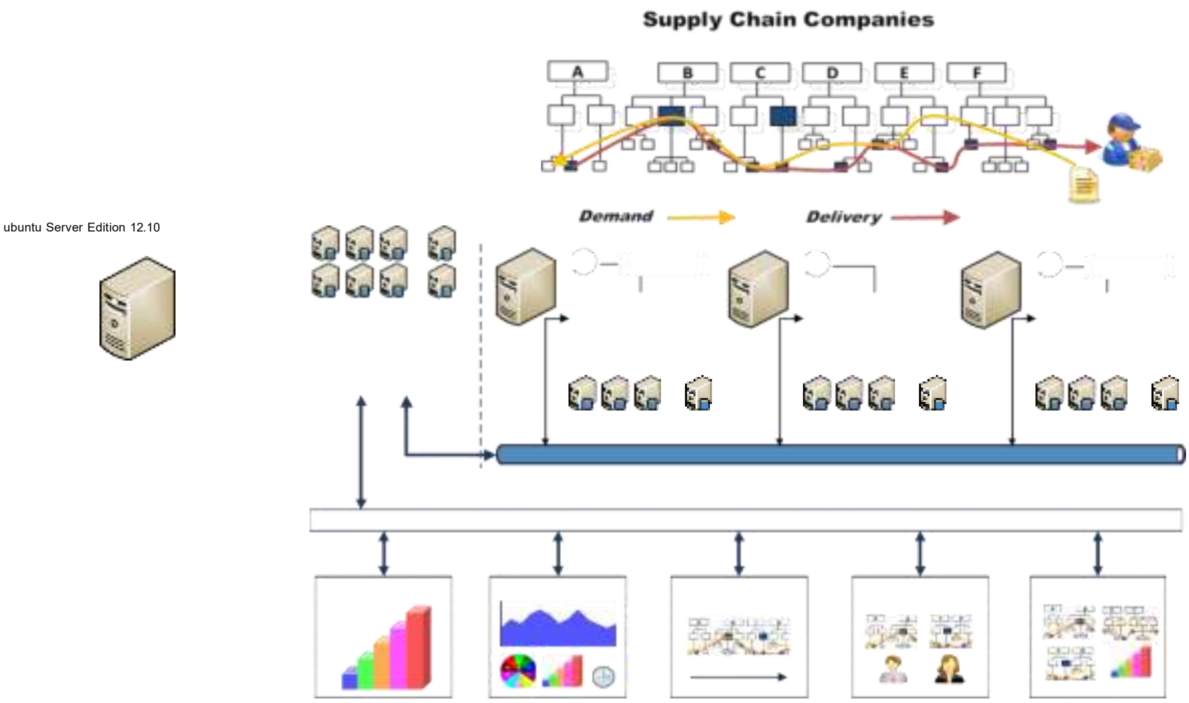

Figure 1. The distributed business analytics services architecture.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 2 | February 2026 - Peer Reviewed Journal

According to Rizzi, "the decision-making process beyond the company boundaries thanks to cooperation and data sharing with other companies and organizations" is how collaborative business intelligence settings expand business analytics functionality. Furthermore, by establishing a global schema that reflects the organization's shared business model, federated data warehouses enable transparent access to the dispersed analytical data across various functional groups. In order to accurately describe the business performance of organizations, we must create a generic model that is completely independent of any particular business area.

3. A MODEL BASED ON EVENTS

For the framework to give a clear knowledge of what has to be tracked, measured, and examined, an event model is required. The event structure must support the data needed to properly analyse the performance of business processes and must depict the data execution of any business process that passes through a variety of heterogeneous systems.

Events and activities that take place throughout the execution of a business process are represented by an event model. The information needed for the global system to conduct analytical procedures over these actions and events is provided by the suggested event model, which also represents any derived measurements made during the execution of business process flows. In order to construct this model, we coupled key elements from the intelligent Web Services Enterprise Integration Environment (iWise) paradigm with the BPAF standard.

BPM systems event streams are the main sources of BPAF data. We use the event structure that BPAF offers, which is agnostic of the underlying process model, to build a general process analytics system. This format aids in standardizing a model for auditing events in diverse environments and unifying criteria for analytical applications and business activity monitoring technologies. A BPAF extension serves as the foundation for the suggested event model, which has been covered elsewhere, in order to support the iWISE-defined event correlation characteristics. We changed the event format to accommodate distributed storage as part of this project.

4. BUSINESS ANALYTICS SERVICE UNIT

Local analyses are handled by the BASU component (see Figure 2), while the GBAS module integrates an arbitrary collection of BASU modules throughout the system and controls cross-organizational dependencies.

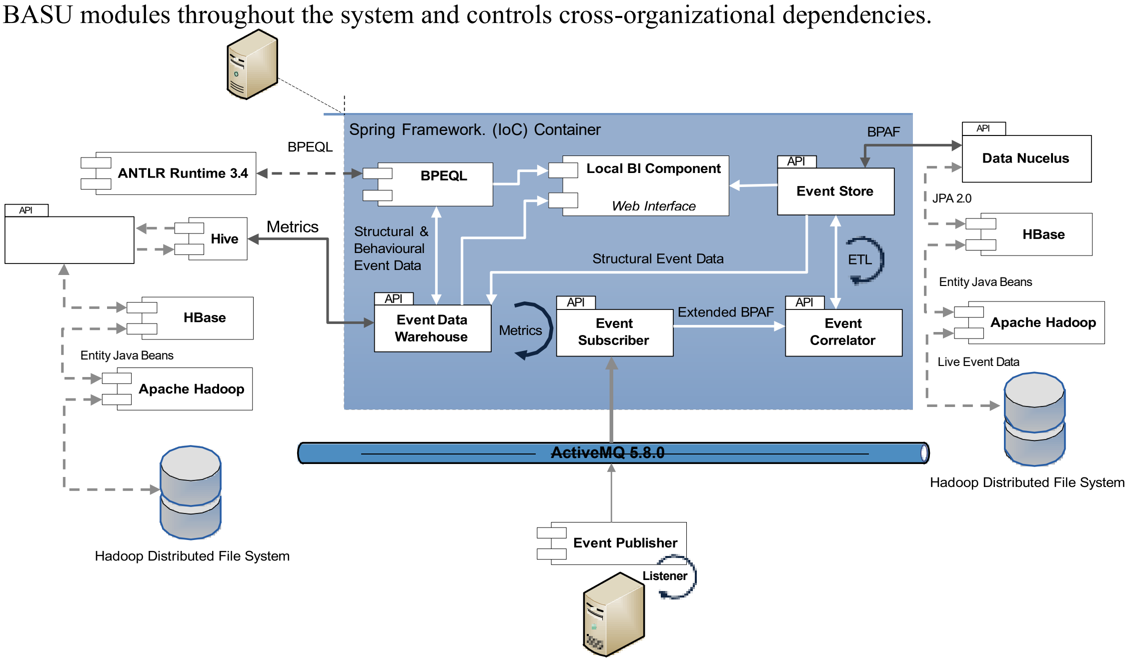

Figure 2. The architecture of the business analytics service unit (BASU)

The event publisher uses an ActiveMQ message broker instance to broadcast events to the network after capturing them from legacy business systems. Event streams are converted into XML messages using the enhanced BPAF format by the legacy listener, which then sends the enterprise events as they happen to a designated Java Message Service (JMS) queue.

In a designated JMS queue, the event subscriber is always on the lookout for incoming events. After that, each event is handled separately by converting its XML message content into a memory representation of an instance in an expanded BPAF format. Before being stored in big data Riccardo Rialti (2020) tables, each instance is sent to the event correlator, which determines the proper order for incoming events.

The process instance or activity linked to the event is found by the event correlator using the expanded BPAF data to query the local event storage for a process instance linked to the given correlation data. In order for the system to provide timely business activity monitoring, the latency for querying massive data tables must be kept to a minimum, which makes information retrieval crucial at this point.

The big data Nilesh Kumar Sahu et. al., (2023) store with the live enterprise event data is accessible through the event store's service interface. A collection of entity beans that represent business events in BPAF format, a collection of Spring components for managing event data across the Java Persistence API (JPA), and an additional collection of Spring components that offer the service interface to the data access methods make up the core of this module. The Data Nucleus open-source project includes an implementation of JPA over HBase, which we utilized to apply the JPA specification to the huge data tables for easy access and management.

In order to extract the event data from the subscriber module, convert the event data structured in the extended BPAF format into raw BPAF, and load the resulting data into the event store, this module's implementation of the extract, transform, and load methods is crucial. Metrics must be established in order to assist business analysts in comprehending the behaviour of the processes, even while real-time enterprise data provides insight into how business processes are being executed but does not offer quantifiable information on company performance. Thus, users can query business events for analytical reasons using the event data warehouse module using big data, which is made up of a data repository of metrics and a subset of event data. In order to support data warehouses, the underlying storage infrastructure is built around an HBase instance and the Hive product.

The suggested method notes the time at which events took place on the source system by locally capturing and recording the timestamp of those events. As the events come in, the event data warehouse module creates metrics for each process instance or activity by analysing the timestamp of a collection of linked events. This analytical data is obtained in a very short amount of time as events occur, and it is completely available via a query language that is similar to specific-purpose SQL.

Data warehouse implementations Seshathiri Dhanasekaran (2018) store and manage metrics and live event data together. This module puts this component into practice to assist analysts in retrieving and processing past events and analysing the performance of business processes using a set of suggested metrics.

5.THE GLOBAL BUSINESS ANALYTICS SERVICE

The next step is to identify supply chain sequences of interconnected processes that are components of a higher-level global business process. We have already discussed how to correlate events per process instance or activity. The sequence flow of a business process Venky Shankararaman (2015) that is operating along the involved systems must be identified as long as it crosses a variety of heterogeneous systems, such as workflow engines or BPEL engines.

The technique by which messages are uniquely identified across several process instances within the framework of an upper global business process is known as instance correlation. This is crucial from the standpoint of business analytics since it enables users to comprehend the relationship between business events, which in turn facilitates automated decision-making. This component correlates the process instances that are carried out across their organizational boundaries and integrates a collection of BASU components. An Enterprise Service Bus (ESB), which is a collaborative network with XML events and metrics data flowing through, connects these BASU subsystems. Due to two primary factors the high event-arrival rates in highly distributed environments and the high dependence of the event correlation mechanism on data access using the clustering capabilities of the Hadoop Distributed File System will be crucial in resolving potential performance issues for event correlation Riccardo Rialti (2020).

Its ability to store business performance data Venky Shankararaman (2015) and real-time enterprise data from cross-organizational business processes allows the GBAS component to offer analytical services of global operations on its own. Similarly, by executing distributed queries across the BASU components along the collaborative network, it enables users to delve into several levels of depth.

Numerical performance data of live event operations was gathered. We gathered data under different execution concurrencies from a dataset of more than 1,000,000 occurrences in a test setting using Jmeter. Write instructions were executed in a range of 5 to 9 milliseconds, whereas read operations took between 0.2 and 0.5 milliseconds (average 0.31 and standard deviation of 0.13). The most striking discovery, though, was that the timings did not increase as the dataset grew, and when comparing, for instance, the dataset populated with 700,000 versus 1,000,000 events, there was no statistical significance in such times.

The fact that distributed processing results in substantial overhead as compared to a centralized approach is one of the main drawbacks of the current strategy. As the number of nodes increases, the GBAS component's processing cost and network latency significantly increase. Furthermore, particularly on vast and complicated supply chain the scenarios we aim to monitor and improve business process instance correlation significantly impairs overall system performance and hinders systems from responding in near real time.

Another goal of assessing general system performance and query processing response times is to be able to forecast system performance in terms of data access for very large amounts of data. In a perfect world, BPA approaches would be applied to vast amounts of data Riccardo Rialti (2020), therefore system scalability has to be assessed in terms of data volume and business query demand. As a result, it will be a crucial case study in subsequent research. Regarding this, utilizing Hadoop due to the strong reliance of the event correlation mechanism on data access and the high event-arrival rates in highly distributed contexts, distributed file system clustering capabilities will be essential to resolving possible performance concerns for event correlation. Other possible study involves progressively adding services to support the sophisticated functionality that new technology requires, including optimization methods or behavioural pattern recognition Aparna Kumari (2020). Furthermore, using simulation methods would strengthen the cloud-based features. By reproducing what-if scenarios and conducting predictive analysis over enhanced data that forms a base of hypothetical information, structured data could be used as an input to simulation engines, enabling business users to anticipate actions. For diagnostic and root cause analysis purposes, this would also assist analysts in rerunning event streams in simulation mode and reproducing live process instances.

The last possible study topic is collaborative business analytics. Real-time visualization of interconnected corporate analytical data would be greatly enhanced by collaboration and data exchange amongst many firms employing big data Nilesh Kumar Sahu et. al., (2023). Additionally, it would facilitate the cooperative execution of diagnostics and root-cause analysis on noncompliant scenarios and bottleneck problems along extensive and intricate business processes that penetrate organizational borders.

6. CONCLUSION

A paradigm shift from traditional centralized analytics to scalable, cloud-enabled, big data-driven Business Process Analytics (BPA) is required due to the increasing scale, velocity, and heterogeneity of event data produced by distant supply chains and complicated business process executions. Organizations' vast process repositories become strategic assets as they grow in Business Process Management (BPM) maturity, impacting crucial performance metrics like time, quality, and cost. However, traditional frameworks are unable to provide real-time insights because of integration issues, large data volumes, and latency limitations. Near real-time monitoring, evidence-based BPM, and cooperative performance management among geographically scattered stakeholders are made possible by the suggested solution, which combines distributed data warehousing, scalable query processing, and process mining inside a cloud-based architecture. In order to improve transparency, trust, and proactive decision-making in next-generation BPM ecosystems, future research will concentrate on standardizing interoperable event formats beyond the current BPAF-based approaches, integrating real-time stream processing and AI-driven predictive and prescriptive analytics, enabling federated and privacy-preserving data sharing across organizational boundaries, and investigating digital process twins and blockchain-based traceability.

7. REFERENCES

1. *Bogdan Danov, Daniela Gotseva, "Study of Basic Concepts in the Context of Process Management and Strategies Based on Big Data", 2024 12th International Scientific Conference on Computer Science (COMSCI), pp.1-4, 2024.*
2. *Thesnath A/L Velayudhan, Dinesh Rajassekharan, "Big Data Analytics in Distributed Business Management", Recent Research Reviews Journal, vol.3, no.1, pp.295, 2024.*
3. *Mostafa Abbasi, Rahnuma Islam Nishat, Corey Bond, John Brandon Graham-Knight, Patricia Lasserre, Yves Lucet, Homayoun Najjaran, "A review of AI and machine learning contribution in business process management (process enhancement and process improvement approaches)", Business Process Management Journal, 2024.*
4. *Mehreen Azam, Khurshid Ahmad, "Adoption of big data analytics for sustainability of library services in academic libraries of Pakistan", Library Hi Tech, 2023.*
5. *Srinivasa Rao Gundu, Charanarur Panem, J. Vijaylaxmi, "A Comprehensive Study on Cloud Computing and its Security Protocols and Performance Enhancement Using Artificial Intelligence", Robotic Process Automation, pp.1, 2023.*
6. *Nilesh Kumar Sahu, Manorama Patnaik, Itu Snigdh, "Data Analytics and Its Applications in Brief", Applications of Big Data in Large- and Small-Scale Systems, pp.115, 2021.*
7. *Aparna Kumari, Sudeep Tanwar, "Secure data analytics for smart grid systems in a sustainable smart city: Challenges, solutions, and future directions", Sustainable Computing: Informatics and Systems, vol.28, pp.100427, 2020.*
8. *Riccardo Rialti, Giacomo Marzi, Ambidextrous Organizations in the Big Data Era, pp.69, 2020.*
9. *Seshathiri Dhanasekaran, Baskar Kasi, Vasanthraj Karunkaran, Anburaman Seetharaman, "Big Data for E-Governance", 2018 International Conference on Recent Trends in Advance Computing (ICRTAC), pp.80-84, 2018.*
10. *Venky Shankararaman, Swapna Gottipati, "A Framework for Embedding Analytics in a Business Process", 2015 IEEE 17th Conference on Business Informatics, vol.2, pp.49-54, 2015.*
11. *Vera-Baquero, R. Colomo-Palacios and O. Molloy, "Business Process Analytics Using a Big Data Approach," in IT Professional, vol. 15, no. 6, pp. 29-35, Nov.-Dec. 2013, doi: 10.1109/MITP.2013.60.*