

LARGE-SCALE PROCESSING OF BIG DATA WITH PYTHON IN MODERN BUSINESS ANALYTICS

*T. I. Mshvidobadze, Professor, Gori State University (Georgia),
tinikomshvidobadze@gmail.com, orcid.org/ 0000-0003-3721-9252*

Methods. The paper is based on a comprehensive review of the impact of Big Data Analytics on modern business analytics and presents an innovative solution for efficient processing of this data in high-performance computing (HPC) environments. The study aims to demonstrate how the integration of Big Data technologies can qualitatively change strategic planning, risk management and operational optimization in a business environment. The analysis covers the dynamic landscape of modern business analytics, emphasizing its transformative power in obtaining critical insights from large and diverse data sets. Special attention is paid to the challenges associated with insufficient performance and general versatility of existing big data processing tools in HPC infrastructures. To address these problems, the study discusses PyCOMPS, a task-based programming model in Python, for the first time. The performance and efficiency of PyCOMPS are evaluated by applying it to implement a complex machine learning algorithm, namely Cascade SVM.

Novelty. The novelty of the work lies in the integrated approach, which not only summarizes the multifaceted role of big data analytics (in particular, in customer-oriented initiatives and operational optimization), but also offers a specific, high-performance and flexible solution. PyCOMPS is positioned as an excellent answer to the problem of the lack of productive and universal tools for distributed big data processing in HPC. The results of the Cascade SVM implementation serve as empirical evidence of its advantages.

Results. The study details the transformational path of big data analytics in modern business intelligence, confirming its crucial role in reducing risks and increasing operational efficiency. The main result is a demonstration of the high performance of PyCOMPS for the efficient development and execution of Big Data analytical tasks. The work concludes with practical recommendations for organizations seeking to maximize the potential of big data analytics in the data-driven era.

Practical value. The work has high practical value for IT architects, data engineers and analysts. Discussion of the benefits PyCOMPS provides a direct toolkit for high-performance and efficient development of big data analytics in the environment of modern business intelligence systems. This solution provides better flexibility and performance compared to traditional HPC models, making complex data processing more accessible to developers using Python.

Keywords: *Big Data, Business Intelligence, Data Analytics, Modern Business, PyCOMPS.*

Statement of problem. Big data analytics refers to the systematic analysis of large and complex data sets to uncover hidden patterns, correlations, and valuable insights. In the modern business analytics landscape, the emergence of big data analytics has become a cornerstone of organizational success [4].

This paper explores how advanced analytical techniques, including machine learning and artificial intelligence, are integrated into

big data platforms, enhancing the depth and accuracy of the information obtained.

In today's dynamic business landscape, the intersection of big data analytics and modern business intelligence (BI) is emerging as a transformative force that is changing the way organizations collect, process, and derive insights from vast and diverse data sets. The sheer volume of data generated every day requires scalable solutions for storage and processing.

The speed of data production requires real-time analytics capabilities. While traditional analytical methods are inadequate when it comes to big data [6].

Machine learning (ML) and artificial intelligence (AI) are seamlessly integrated into big data platforms, increasing the depth and accuracy of analysis. Machine learning algorithms, such as clustering and regression, successfully learn from large data sets. Artificial intelligence, with its cognitive capabilities, facilitates understanding, reasoning, and decision-making.

The transformative impact of big data analytics on decision-making is a central theme. Organizations use data-driven analytics for strategic planning and risk management. Predictive analytics enable businesses to anticipate market trends, identify potential risks, and make informed decisions that align with overall objectives. In finance, big data analytics improves fraud detection and risk assessment. In healthcare, it facilitates personalized medicine and improves patient outcomes through predictive analytics.

Real-world applications demonstrate a wide range of impacts. In marketing, campaigns become more targeted and effective through data-driven analytics. In manufacturing, processes optimize efficiency based on recommendations from analytics. Customer-centric initiatives, from personalized recommendations to tailored user experiences, are enhanced by the detailed analysis provided by big data analytics.

The paper discusses real-world applications across various sectors and shows how big data analytics has revolutionized industries such as finance, healthcare, marketing, and manufacturing.

The amount of data that society generates is growing extremely fast. It is expected that by 2025, the amount of useful data will be of more than 18 zettabytes (i.e., GB). [13].

The process of extracting useful information from large amounts of data is also known as big data analytics (BDA) [10]. BDA involves transforming the data using various operations, such as sorting, aggregating, or filtering, as well as using machine learning algorithms to obtain new information and to discover patterns in data.

In this paper, we review and evaluate PyCOMPS [14] as an approach to bridge the gap between BDA and HPC programming models. PyCOMPS is a task-based programming model that can be used to easily build and execute parallel Python applications. PyCOMPS has the advantage of being based on Python, which provides great productivity and is one of the most popular programming languages among data scientists.

The paper concludes by highlighting the changing landscape of big data analytics, reviewing emerging trends such as Edge Analytics and the integration of big data with other transformative technologies.

In essence, this paper reviews the critical role of big data analytics in modern business analytics and the dynamic interplay between analytics and organizational innovation.

Analyses of recent papers. The paper explores the changing landscape of big data analytics, starting with emerging trends such as edge analytics [11]. This approach involves processing data at the edge of the network, closer to the data source. Edge analytics reduces latency, enhances real-time insights, and is particularly relevant in scenarios where immediate responsiveness is critical, such as IoT applications and autonomous systems.

Combining big data analytics with transformative technologies such as the Internet of Things (IoT) and blockchain expands its capabilities [3]. The synergy of these technologies creates ecosystems where data is not only analyzed at scale, but also seamlessly integrated into a variety of applications. From optimizing supply chains to powering smart city initiatives, the convergence of technologies promises new frontiers of innovation. In conclusion, the transformative role of Big Data Analytics in modern business analytics is evident across sectors and industries. Its fundamental principles, integration of advanced analytical techniques, impact on decision-making processes, and changing landscape demonstrate the breadth and depth of its impact.

Despite the challenges, organizations are moving forward, addressing ethical considerations, strengthening cybersecurity measures, and embracing new trends [5]. As we stand at the crossroads of data abundance and technological innovation, the potential of Big Data

Analytics will expand even further. It is not just a data processing tool; it is a catalyst for innovation, a guide to strategic decisions, and a cornerstone for businesses that aspire to thrive in the digital age.

The journey continues, and with each advancement in Big Data Analytics, the future of business analytics becomes increasingly data-driven and limitless.

Materials and methods. The speed of data production requires real-time analytics capabilities. In addition, the diversity of data, including structured and unstructured formats, requires flexible processing methods. Big data platforms, such as Apache Hadoop and Apache Spark, provide the infrastructure for navigating and analyzing these vast data sets.

The fundamental principle of volume processing involves distributed computing frameworks such as Apache Hadoop and Apache Spark [15]. These frameworks break large data sets into small, manageable chunks that are distributed across multiple nodes or clusters. This distributed approach allows organizations to leverage parallel processing power, significantly reducing the time required to analyze data.

By integrating technologies such as Apache Kafka for streaming processing, organizations can analyze data as it is generated, enabling real-time decision-making. This principle of speed is especially important in fields such as finance, where making split-second decisions can have a significant impact on trading outcomes.

The Hadoop Distributed File System (HDFS) is an example of a storage system designed to handle a variety of data formats [16]. It allows organizations to store and process unstructured data alongside structured data. This versatility is crucial in industries such as healthcare, where patient records may include structured data such as demographics and unstructured data such as medical images or doctor's notes.

The integration of analytical technologies, especially machine learning (ML), is making big data analytics a powerful predictive force [8]. Machine learning algorithms learn from historical data patterns, allowing organizations to make accurate predictions and optimize decision-making processes.

Classification algorithms, such as logistic regression, are used in scenarios where data needs to be categorized into predefined classes (Shah et al., 2020). Regression algorithms, on the other hand, are used to predict numerical values, which makes them valuable for financial forecasting or demand forecasting. By integrating machine learning into Big Data Analytics, organizations can uncover hidden patterns in the vast data sets they accumulate.

Artificial Intelligence (AI) goes beyond machine learning and introduces cognitive capabilities that allow machines to understand, reason, and make autonomous decisions [1]. Natural Language Processing (NLP) and Computer Vision are components of Artificial Intelligence integrated into Big Data Analytics, allowing organizations to extract insights from unstructured data sources such as text or images.

The transformative impact of big data analytics on decision making is a central theme. Organizations use data-driven insights for strategic planning and risk management. Predictive analytics enables businesses to anticipate market trends, identify potential risks, and make informed decisions that align with key objectives. In finance, big data analytics improves fraud detection and risk assessment. In healthcare, it facilitates personalized medicine and improves patient outcomes through predictive analytics.

Data mapping is the first process of data analytics, followed by business understanding, data exploration, data preparation, as shown in Figure 1. Data preparation is followed by data modeling, and finally data evaluation.

In marketing, campaigns become more targeted and effective through data-driven analytics [10]. In manufacturing, processes optimize efficiency based on recommendations from analytics.

Customer-centric initiatives, from personalized recommendations to tailored user experiences, are enhanced by the detailed analysis provided by Big Data Analytics.

Big Data Analytics serves as a strategic compass that enables organizations to anticipate market trends through predictive analytics. By analyzing historical data and identifying patterns, businesses can make informed predictions about future market movements. This is

especially important in dynamic industries such as retail and e-commerce, where understanding consumer behavior and predicting market trends can be the foundation for successful strategic planning.

Big data analytics enables sophisticated scenario modeling, allowing organizations to assess and mitigate potential risks before they materialize [7]. By analyzing historical data and simulating different scenarios, businesses can make risk-based decisions, whether it's financial investments, supply chain management, or project planning. This proactive risk management approach is instrumental in mitigating unforeseen challenges.

The Evolving Landscape and Emerging Trends The paper explores the evolving landscape of Big Data Analytics, beginning with

emerging trends such as edge analytics. This approach involves processing data at the edge of the network, closer to the data source. Edge analytics reduces latency, enhances real-time insights, and is particularly relevant in scenarios where immediate responses are crucial, such as in IoT applications and autonomous systems. Federated learning, a decentralized approach to machine learning, is gaining prominence. It allows models to be trained across multiple decentralized devices without centralizing raw data. This privacy-preserving technique is valuable in industries where data security and privacy are paramount, such as healthcare and finance. The fusion of Big Data Analytics with transformative technologies like the Internet of Things (IoT) and blockchain extends its capabilities.

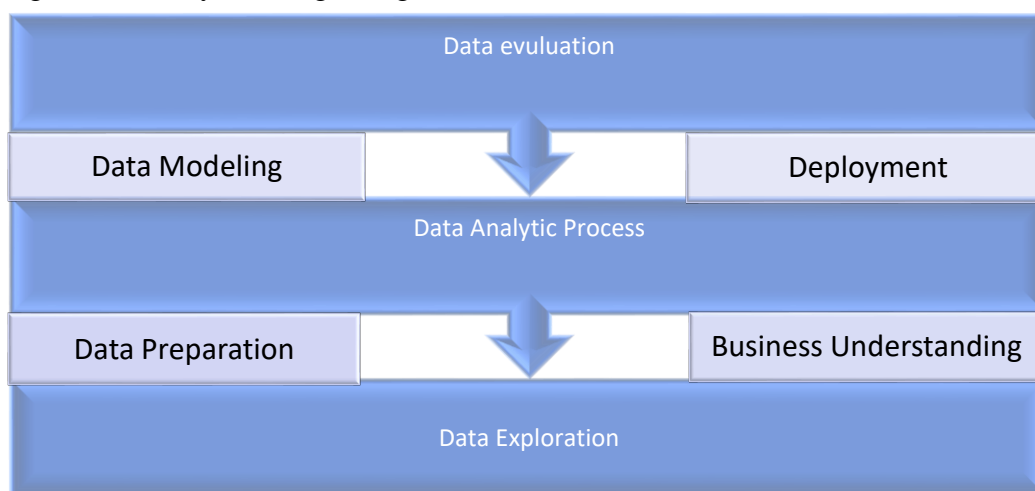


Figure 1. Schematic of the Data Analytic Process

The synergy of these technologies creates ecosystems where data is not only analyzed at scale but also seamlessly integrated into diverse applications. From optimizing supply chains to enhancing smart city initiatives, the convergence of technologies promises new frontiers of innovation. In the rapidly evolving landscape of Big Data Analytics, staying ahead of emerging trends is paramount for organizations seeking to harness the full potential of their data [12].

Programming model – PyCOMPSs. We consider the Python-based software framework PyCOMPS as an excellent solution for distributed big data processing on HPC infrastructure. A machine learning algorithm (Cascade SVM) is implemented using PyCOMPS and its performance is evaluated based on these algorithms.

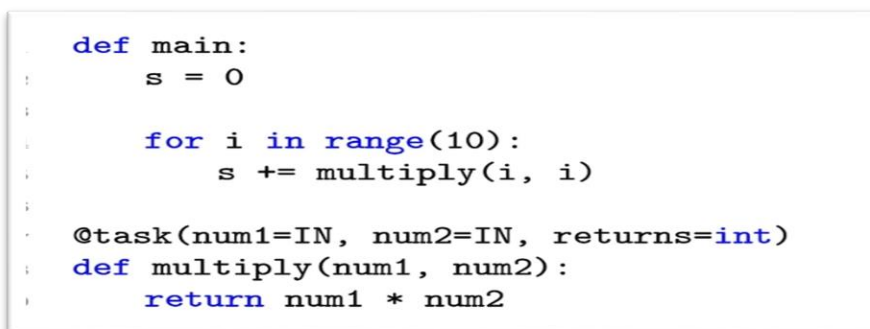
PyCOMPSs is a task-based programming model that makes the development of parallel and distributed Python applications easier. PyCOMPSs consists of two main parts: programming model and runtime. The programming model provides a series of simple annotations that developers can use to define potential parallelism in their applications. The runtime analyzes these annotations at execution time, and distributes the computation automatically among the available resources. The main component of PyCOMPSs' programming model is the task annotation, which defines units of computation that can be executed remotely.

PyCOMPSs applications are regular Python applications with certain annotations that help the runtime to exploit parallelism.

The simplicity of PyCOMPSs' programming model allows for fast development of data

analytics algorithms in a highly productive language that is widely used in the scientific community, and that is surrounded by a large ecosystem of mathematical libraries [9]. Moreover,

any existing Python application can be easily parallelized by just including some annotations in the code.



```
def main:
    s = 0

    for i in range(10):
        s += multiply(i, i)

@task(num1=IN, num2=IN, returns=int)
def multiply(num1, num2):
    return num1 * num2
```

Figure 2. Example PyCOMPSs application

In an experiment conducted by Javier Álvarez Cid-Fuentes et al., the performance of the PyCOMPSs C-SVM implementation was evaluated using 7 publicly available datasets. [2].

In experiments with C-SVM, PyCOMPS imposes communication overhead and achieves higher execution times. This indicates that PyCOMPS generally performs better with large applications and long-running tasks. Therefore, BDA applications that process large amounts of data typically run for more than 4 minutes and have medium to high granularity tasks.

Recommendation. The transformative role of Big Data Analytics in modern business analytics is evident across all sectors and industries. Its fundamental principles, integration of advanced analytical techniques, impact on decision-making processes, and changing landscape demonstrate the breadth and depth of its impact.

Despite the challenges, organizations are moving forward, addressing ethical considerations, strengthening cybersecurity measures, and embracing new trends. As we stand at the crossroads of data abundance and technological innovation, the potential of Big Data Analytics is poised to expand even further. It is not just a data processing tool; it is a catalyst for innovation, a guide to strategic decisions, and a cornerstone for businesses that aspire to thrive in the digital age.

In the ever-expanding landscape of modern business analytics, the transformative role of Big Data Analytics is undeniable. As organizations navigate the complexities of vast data sets and dynamic market environments, har-

nessing the full potential of analytics becomes not only a strategy but also a necessity for sustainable success.

Recommendations for organizations seeking to optimize their use of Big Data Analytics:

- In the digital age, data privacy is paramount, so organizations must prioritize ethical data practices, implement privacy-preserving measures such as anonymization and encryption to strike a balance between collecting valuable information and respecting individual privacy.

- Organizations should strengthen cybersecurity measures,

ensure strong encryption protocols and continuous monitoring to protect data assets.

- Address machine learning privacy issues by preserving the locality of sensitive data.

- Synergy with transformative technologies such as the Internet of Things (IoT) and blockchain, evaluating how blockchain applications can increase trust and security in data processes, especially in scenarios related to transactions and supply chain management.

- It is best to use the artificial intelligence-based software framework PyCOMPS, which is considered an excellent solution for processing distributed big data.

Conclusions. In conclusion, the transformative role of Big Data Analytics in modern business analytics is an ongoing journey characterized by continuous innovation and adaptation. As organizations navigate a data-driven future, these recommendations provide guiding principles for unlocking the full potential of analytics.

The profound impact of Big Data Analytics is being felt across industries, from strategic planning and risk management to operational optimization and customer-centric initiatives.

As we stand at the crossroads of data abundance and technological evolution, organizations that employ ethical data practices, strengthen cybersecurity measures, explore emerging trends, and foster a culture of innovation will thrive in the dynamic landscape of modern business analytics. The journey to a data-driven future is an opportunity to shape a narrative of success and innovation led by the transformative capabilities of Big Data Analytics.

In this paper, we show that PyCOMPS provides a general-purpose programming model with the best compromise between productivity, flexibility, and performance. PyCOMPS allows developers to write BDA algorithms with more functionality and less complexity.

References

1. Abele, D., & D'Onofrio, S. (2020). Artificial intelligence—the big picture. *Cognitive Computing: Theorie, Technik und Praxis*, 31-65.
2. Amela R., Ishii K., Morizawa R., 2020, Efficient development of high performance data analytics in Python, *Future Generation Computer Systems*, Volume 111. Pages 570-581. <https://doi.org/10.1016/j.future.2019.09.051>
3. Alam, T. (2022). Blockchain cities: the futuristic cities driven by Blockchain, big data and internet of things. *GeoJournal*, 87(6), 5383-5412. <https://doi.org/10.1007/s10708-021-10508-0>
4. Bharadiya, J.P. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. *American Journal of Artificial Intelligence*, 7(1), 24. <https://doi.org/10.11648/j.ajai.20230701.14>
5. Chukwu, E., Adu-Baah, A., Niaz, M., Nwagwu, U., & Chukwu, M.U. (2023). Navigating ethical supply chains: the intersection of diplomatic management and theological ethics. *International Journal of Multidisciplinary Sciences and Arts*, 2(1), 127-139.
6. Dekimpe, M.G. (2020). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*, 37(1), 3-14. <https://doi.org/10.1016/j.ijresmar.2019.09.001>
7. Figueira, P.T., Bravo, C.L., & López, J.L.R. (2020). Improving information security risk analysis by including threat-occurrence predictive models. *Computers & Security*, 88, 101609. <https://doi.org/10.1016/j.cose.2019.101609>
8. Nguyen, D.K., Sempinis, G., & Stasinakis, C. (2023). Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology. *European Financial Management*, 29(2), 517-548. <https://doi.org/10.1111/eufm.12365>
9. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A. 2011, *Scikit-learn: machine learning in python*. *Learn. Res.*, pp. 2825-2830.
10. Rosário, A.T., & Dias, J.C. (2023). How has data-driven marketing evolved: Challenges and opportunities with emerging technologies. *International Journal of Information Management Data Insights*, 3(2), 100203.
11. Rawat, K.S., & Sood, S.K. (2021). Emerging trends and global scope of big data analytics: a scientometric analysis. *Quality & Quantity*, 55, 1371-1396.
12. Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augmented Human Research*, 1-16. <https://doi.org/10.1007/s41133-020-00032-0>
13. Turner V. The digital universe of opportunities: rich data and the increasing value of the internet of things, *International Data Corporation* (2014).
14. Tejedor E., Becerra Y., Alomar G., Queralt A., Badia R.M., Torres J., Cortes T., Labarta J. P2017, COMPSs: parallel computational workflows in python. *Int. High Perform. Appl.*, 31 (1), pp. 66-82. <https://doi.org/10.1177/1094342015594678>
15. Xu, Y., Liu, H., & Long, Z. (2020). A distributed computing framework for wind speed big data forecasting on Apache Spark. *Sustainable Energy Technologies and Assessments*, 37, 100582.
16. Zeebaree, S.R., Shukur, H.M., Haji, L.M., Zebari, R.R., Jacksi, K., & Abas, S.M. (2020). Characteristics and analysis of hadoop distributed systems. *Technology Reports of Kansai University*, 62(4), 1555-1564.

МАСШТАБНА ОБРОБКА ВЕЛИКИХ ДАНИХ ЗА ДОПОМОГОЮ PYTHON У СУЧАСНІЙ БІЗНЕС-АНАЛІТИЦІ

Т. І. Мішвідобадзе, професор, Горійський державний університет, Грузія

Методи. Робота базується на всебічному огляді впливу аналітики великих даних (Big Data Analytics) на сучасну бізнес-аналітику та представляє інноваційне рішення для ефективної обробки цих даних у високопродуктивних обчислювальних (HPC) середовищах. Дослідження має на меті продемонструвати, як інтеграція технологій великих даних може якісно змінити стратегічне планування, управління ризиками та оптимізацію операцій у бізнес-середовищі. Аналіз охоплює динамічний ландшафт сучасної бізнес-аналітики, наголошуючи

на її трансформаційній силі в отриманні критичних інсайтів з великих та різноманітних наборів даних. Особлива увага приділяється викликам, пов'язаним із недостатньою продуктивністю та загальною універсальністю існуючих інструментів обробки великих даних у НРС-інфраструктурах. Для вирішення цих проблем, у дослідженні вперше обговорюється PyCOMPS – модель програмування, заснована на завданнях (task-based programming model) на мові Python. Продуктивність та ефективність PyCOMPS оцінюються шляхом його застосування для імплементації складного алгоритму машинного навчання, а саме Cascade SVM.

Новизна. Новизна роботи полягає в інтегральному підході, який не лише підсумовує багатогранну роль аналітики великих даних (зокрема, у клієнтоорієнтованих ініціативах та оперативній оптимізації), але й пропонує конкретне, високопродуктивне та гнучке рішення. PyCOMPS позиціонується як відмінна відповідь на проблему відсутності продуктивних та універсальних інструментів для розподіленої обробки великих даних у НРС. Результати імплементації Cascade SVM слугують емпіричним доказом його переваг.

Результати. Дослідження детально окреслює трансформаційний шлях аналітики великих даних у сучасній бізнес-розвідці, підтверджуючи її вирішальну роль у зниженні ризиків та підвищенні операційної ефективності. Головним результатом є демонстрація високої продуктивності PyCOMPS для ефективного розроблення та виконання аналітичних завдань Big Data. Робота завершується наданням практичних рекомендацій для організацій, які прагнуть максимально використати потенціал аналітики великих даних в еру, керовану даними.

Практичне застосування. Робота має високу практичну цінність для IT-архітекторів, інженерів даних та аналітиків. Обговорення переваг PyCOMPS надає безпосередній інструментарій для високопродуктивного та ефективного розвитку аналітики великих даних у середовищі сучасних бізнес-аналітичних систем. Це рішення забезпечує кращу гнучкість та продуктивність порівняно з традиційними НРС-моделями, роблячи складну обробку даних більш доступною для розробників, які використовують Python.

Ключові слова: великі дані, бізнес-аналітика, аналітика даних, сучасний бізнес, PyCOMPS.

Стаття надійшла до редакції 10.10.25 р.

Прийнята до публікації 25.10.25 р.

Дата публікації 26.12.25 р.