

**Big Data Theory and Practice**

***7 BDIN006W.1***

**Hasan Enes Guray**

**Student ID: 19489124**

**Word Count: 2601**

**Banking and Securities Sector**

**Group Members:  
Folusho Victor Arokoyo  
Saeed Jamshidloo  
Chinyere Unamba  
Oluwabukola Atere   
Hasan Enes Guray  
Drici Mourad  
Kiana Rezaei Amrabadi  
Soheib Kohneposhi**

Table of Contents

[1 INTRODUCTION 3](#_Toc155622054)

[2 CHALLENGES AND DATA LANDSCAPE 4](#_Toc155622055)

[2.1 Data Security and Privacy Concerns 4](#_Toc155622056)

[2.2 Data Quality and Integrity 4](#_Toc155622057)

[2.3 Regulatory Compliance 4](#_Toc155622058)

[2.4 Challenges of Legacy Systems 4](#_Toc155622059)

[2.5 Scalability and Skill Gap 5](#_Toc155622060)

[3 TECHNOLOGY & SOLUTION ANALYSIS 5](#_Toc155622061)

[3.1 Technology 5](#_Toc155622062)

[3.1.1 Hadoop 5](#_Toc155622063)

[3.1.2 Apache Spark 6](#_Toc155622064)

[3.1.3 MongoDB 6](#_Toc155622065)

[3.2 Solutions 7](#_Toc155622066)

[4 OUTCOMES & REFLECTION 8](#_Toc155622067)

[4.1 Outcome 8](#_Toc155622068)

[4.2 Reflection 9](#_Toc155622069)

[5 MEETING MINUTES 10](#_Toc155622070)

[6 CONCLUSION 10](#_Toc155622071)

[7 INDIVIDUAL COMPONENT 11](#_Toc155622072)

[7.1 Group Work Evaluation 11](#_Toc155622073)

[7.2 Individual Contribution Assessment 11](#_Toc155622074)

[8 REFERENCES 13](#_Toc155622075)

[9 APPENDICES 15](#_Toc155622076)

[A Meeting Minutes 1 15](#_Toc155622077)

[B Meeting Minutes 2 16](#_Toc155622078)

[C Meeting Minutes 3 18](#_Toc155622079)

[D Meeting Minutes 4 20](#_Toc155622080)

[E Meeting Minutes 5 21](#_Toc155622081)

[F WhatsApp Message 1 23](#_Toc155622082)

[G WhatsApp Message 2 23](#_Toc155622083)

**ABSTRACT**

This technical report analyses the pivotal role of big data in the banking and securities industry. It focuses on the implementation and challenges of advanced data analytics in this sector, highlighting essential aspects such as data volume, speed, diversity, accuracy, and value. The document discusses the importance of these technologies in refining risk assessment processes, customer segmentation, and enhancing overall operational effectiveness. It addresses significant obstacles, including safeguarding data, adherence to legal standards, and the need for scalable systems while exploring the utilization of platforms. The study looks at the future of advanced data analytics in finance and how it will work with new technologies like blockchain and the Internet of Things. It predicts a big change in how efficient and safe financial services are, which will completely change the industry in light of digital progress.

# INTRODUCTION

With the current digital revolution, the banking sector exemplifies the concrete influence of Big Data, as its worldwide market is anticipated to surge from $138.9 billion in 2020 to $229.4 billion by 2025(*Big Data In Banking: Opportunities, Challenges, And Future Prospects – Avenga.* 2023). This paradigm shift paves the way for big data to have a significant impact on banking by forcing financial institutions to reassess their information operations. This course explores the growth, difficulties, and uses of big data in banking, revealing the Five Vs (Vinaya Keskar, Jyoti Yadav, Ajay H. Kumar, 2020).

When it is talked about the volume of data in the banking and securities sector, the massive amount of information coming from market activities, customer interactions, and transactions should be considered. It's like a flood of data, and to manage it, there's a big push towards using Big Data tech like cloud computing and distributed storage systems. The idea is to have a system that can grow and adapt as the data does, kind of like how your phone storage needs to keep up with all the photos and apps you download.

For the variety of the data, everything from emails and messages to more organized records like logs and transaction data exist. To make sense of this mix, financial experts are turning to fancy analytics tools and algorithms. It's like having a translator who can understand and interpret different languages. Machine learning, especially, is essential for deciphering all these data formats and finding the hidden gems of insight.

Also how fast data comes to people in the financial world is an essential point. Market changes and transactions happen in a flash. To keep up, real-time processing and streaming analytics are key. The systems have to be quick and sharp to make sense of data that's always on the move.

But speed and size are not everything. The data should be reliable. In the banking sector, even a tiny error can throw off big decisions. Mistakes must be meticulously detected. Strong rules and checks are super important to keep the data clean and trustworthy.

Lastly, the value extracted from all this data should not be forgotten. Advanced analytics, including machine learning and predictive modelling, are crucial tools. But it's not just about finding these insights; they need to be relevant and help in making smart business decisions. The whole point is to use this data to improve services, manage risks better, and boost overall performance in the financial world.

# CHALLENGES AND DATA LANDSCAPE

In the banking sector, one of the primary areas benefiting from Big Data, there are distinct challenges closely associated with the nature and size of the data involved. Big Data is both a solution and a challenge in this evolving environment.

## Data Security and Privacy Concerns

The banking industry handles ever-growing volumes of data, encompassing diverse private and personal details, making it more susceptible to cyber threats and unauthorized data intrusions. To maintain the soundness of the financial system and protect their clients' privacy rights, these institutions are compelled to implement rigorous cybersecurity measures. Utilizing complex data analytics and AI tools, banks can improve their surveillance, detect anomalies quickly, and prevent security breaches. Regulations like GDPR have imposed stricter data security standards, making compliance a complex task and only 38% of organizations worldwide are ready to handle these threats, highlighting the criticality of robust cybersecurity measures in banking (Shalimov, 2023).

## Data Quality and Integrity

Faulty or incomplete data can result in poor decision-making, impacting areas from customer relationships to risk management. Achieving smooth integration and accuracy among diverse data sources is challenging. Big Data solutions, with their sophisticated analytics, assist in refining, processing, and consolidating data from multiple sources, ensuring banks base decisions on reliable and accurate data. The task of validating and cleaning big data is rigorous, necessitating constant attention to ensure the data's validity, accuracy, and completeness. This process is complicated further by the need to manage a mix of structured and unstructured data (Wissen Team, 2023).

## Regulatory Compliance

The heavily regulated banking industry faces complexities in complying with data storage, usage, and sharing regulations, compounded by the increasing data volume. While Big Data technologies add to this complexity, they also simplify compliance processes. Banks are empowered through these systems to oversee and administer regulatory processes with efficiency. This adherence to regulatory frameworks does more than merely align with legal requirements; it fosters an environment characterized by transparency and dependability in operations. Compliance becomes increasingly intricate with the growing volume of data. Regulations such as GDPR impose complex rules that govern access to critical client data, making navigation through the regulatory landscape a demanding task (Wissen Team, 2023).

## Challenges of Legacy Systems

Many banks still operate on outdated systems that struggle with the growing volume and complexity of data. Upgrading these systems is expensive and demands a cultural shift within the organization. Despite these hurdles, Big Data presents a way forward with its efficient processing and analysis of large data volumes, encouraging banks to update their legacy systems. Many leading banks still operate on older systems, which struggle to handle the volume and complexity of modern data. Upgrading these systems is not only expensive but also requires an operational and cultural shift within the institution (Wissen Team, 2023).

## Scalability and Skill Gap

As banks grow, so does the volume of data they manage. Ensuring that Big Data solutions are scalable to accommodate this growth presents technical challenges and requires ongoing investment. Additionally, there is a significant demand for skilled professionals in Big Data analytics, which can slow down implementation processes and impact the quality of insights derived from data (Zubenko, 2023).

In the domain of banking, the extensive utilization of Big Data stands as a cornerstone for developing intricate customer profiles and strategizing risk management. This technology breakthrough surpasses conventional approaches by providing a strong foundation for detecting fraud and creating customized services that cater to the specific demands of each consumer. The essence of Big Data analytics rests in its ability to condense extensive and unrelated datasets into clear and actionable insights. These insights are crucial for guiding banks toward making prudent decisions, resulting in improved operational efficiency and a significant increase in client satisfaction levels. An exhaustive scrutiny of consumer behavioural patterns, along with a comprehensive study of transactional histories and developing market dynamics, enables banks to refine their service offers. Adopting a strategic strategy helps identify upcoming threats and ensures that banks maintain a strong position in the constantly changing banking sector (Naveira et al., 2018).

The incorporation of Big Data in the financial industry, albeit posing significant difficulties, offers opportunities for profound transformations. It serves as a catalyst for innovation, improves operational efficiency, and plays a crucial role in the creation of services that prioritize client demands and preferences.

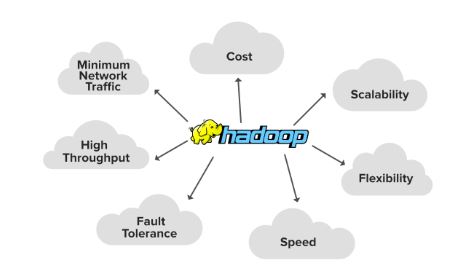
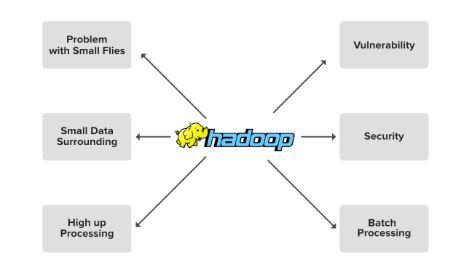
# TECHNOLOGY & SOLUTION ANALYSIS

Big data technologies facilitate the analysis of heterogeneous data sets that beyond the limits of conventional databases. They increase operational efficiency and facilitate quicker and better decision-making. They are essential in the banking field to improve corporate intelligence and handle huge volumes, velocities, and varieties of data (Vyas, 2023). The most frequently used ones in the banking area are Hadoop, Apache Spark, Apache Flink, Apache Kafka, MongoDB, Sisense, and RapidMiner (Hillier, 2023).

## Technology

### Hadoop

Created by Apache, Hadoop is a big data management framework. It provides a distributed storage infrastructure for batch information processing using straightforward programming models, handling hardware faults, and processing huge datasets. Because of its scalability, it can provide highly available service for a single server up to thousands of workstations (Vyas, 2023).

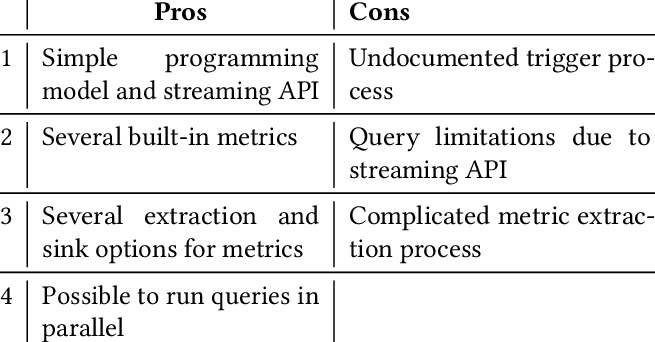


**Figure 2**  *This is a schema, which shows the disadvantages of Hadoop('Hadoop - Pros and Cons', 2020)*

**Figure 1**  *This is a schema, which shows the advantages of Hadoop('Hadoop - Pros and Cons', 2020)*

### Apache Spark

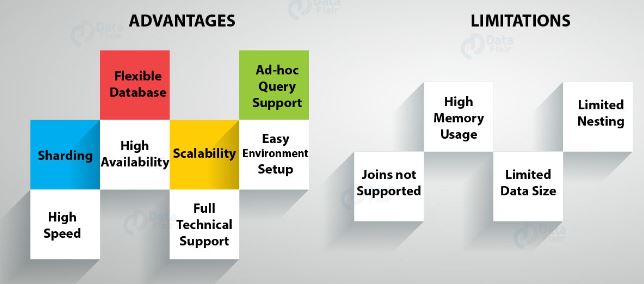
A lot of banking structures use the open-source analytics engine Apache Spark for high-performance data processing and scalable computing, making it a top choice for large-scale analytics. Its sophisticated distributed SQL engine executes queries more quickly than other data warehouses and provides adaptive query execution (Vyas, 2023).



**Figure 3** *This is a table, which shows the pros and cons of Apache Spark (Todor Ivanov,).*

### MongoDB

MongoDB is a NoSQL database that uses AI models to help build more relevant and responsive consumer experiences. In addition to having inherent vector capabilities that make use of large language models to create intelligent applications, it integrates data tiering and federation for optimal storage (Vyas, 2023).



**Figure 4** *This is a table, which shows the advantages and cons of MongoDB (Team, 2018).*

## Solutions

The banking business is now changing a lot because of big data solutions. They help make customers happy, manage risks better, and run operations more efficiently. Banks are focusing on using analytics in areas like sales management, money risk control, and security against hacks. This provides better choices for decisions being made with more knowledge. But it's still hard to put these numbers and data fully into how businesses work. Only a few banks have been able to do this (McKinsey & Company, 2023).

In banking, big data is used to predict what customers might do next. This uses old customer information and intelligent machines to guess their future actions. Using this method, banks can think about what customers need ahead of time and suggest new products or services. This makes it better to look at how much a customer is worth in the long run and predict when they might leave (Hitachi Solutions, 2023). Marketing plans made for certain groups of customers have become more common, making them connect better and boost conversion rates.

Big data helps separate customers into groups. This allows for better marketing towards them and studying their feedback more efficiently. It's essential to stop fraud by watching what people spend money on. It also looks for anything unusual (Avenga, 2023). Big data solutions help make banks safer from hackers and can automate up to 30% of their work. This leads to saving money and fewer mistakes in bank operations.

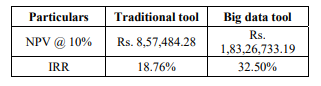
However, old computer systems usually need extra help to take care of more data work, and there are big worries about keeping information safe and private. Extra challenges include sticking to rules, high costs for setting it up, and the need for particular skills. Even though there are problems with big data in banking, the good things it can do, like better risk control and efficient work processes, make a considerable difference (Eastern Peak, 2023; N-iX, 2023).

In the end, big data solutions are changing how banks work. They open up lots of chances for new ideas and better connecting with customers. But banks have to deal with many problems. These include adding analytics for their work, looking after the security of data, and making sure they follow all rules set by lawmakers.

# OUTCOMES & REFLECTION

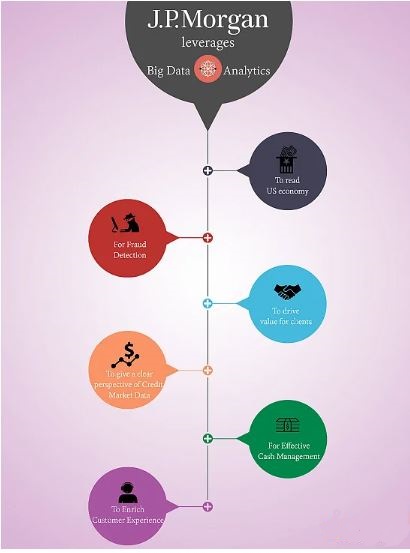
## Outcome

According to the research field (Chandani *et al.,* 2015), banks can achieve significant advantages by implementing big data, including reduced costs, effective customer channel identification, fraud prevention, risk analysis, and customized products. Big data is financially viable even with higher initial costs, as demonstrated by the cost-benefit analysis with NPV and IRR.



**Figure 5**  *This is a table, which shows the general summary of the financial information related to big data.*

Using Apache Hadoop and its 150+ petabytes of data, JPMorgan Chase transformed its operations by utilizing the big data analytics (Projectpro, 2023). Efficient management of large amounts of emails, social media data, and transaction records was made possible by this adoption. The bank demonstrated its dedication to the good of society by improving risk management, optimizing sales of foreclosed properties, and applying analytics to gain economic insights.



**Figure 6** *This is a schema, that illustrates**big data applications in JP Morgan.*

Global data production jumped to over 59 zettabytes in 2020 following the COVID-19 pandemic('Big Data: Its Advantage, Challenges and Relevance', ). The skilful use of big data by HDFC Bank made personalized customer deals possible. Big Data analytics, sometimes known as the "new oil," enables companies like Netflix to save billions of dollars by retaining customers. However, obstacles like regulatory limitations, a lack of qualified workers, and privacy issues make implementation difficult.

## Reflection

Chandani in their paper (Chandani *et al.,* 2015) highlights big data's disruptive potential in the banking industry and highlights how it can lead to competitive advantage and well-informed decision-making. It is accepted that there are hurdles, such as skill development, infrastructure investment, and cultural transition. Big data has created a lot of hype, but research shows that adoption requires a customized strategy that fits company culture and strategic objectives.

A review of JPMorgan Chase's big data integration reveals a revolution in the banking industry (Projectpro, 2023). The deliberate application of Hadoop for economic analysis, fraud detection, and customer-focused services represents a forward-thinking development in the industry. It is very helpful to see how data analytics is changing banking operations to get a sense of how important technology is in forming financial environments.

The article ('Big Data: Its Advantage, Challenges and Relevance', ) highlights the importance of big data on a global scale and the effective use of big data by HDFC Bank. While major international players use data analytics to increase efficiency, Nepal finds it difficult to adopt this technology. Adoption is hampered by resistance from established business models, a lack of technology infrastructure, and inadequate education. The opportunities are in changing attitudes, supporting technology education, and keeping up with international developments—particularly for the next wave of Nepalese entrepreneurs. A cultural shift toward data-driven decision-making is necessary for the successful integration of big data, which will revolutionize Nepal's business ecosystem.

# MEETING MINUTES

Group 3 in GP12, engaged in a series of five meticulously documented meetings. Each meeting, adhering to a structured agenda, encompassed a review of prior minutes, updates on task progress, and the allocation of new research areas, ranging from the analysis of key Big Data applications to the scrutiny of technological tools like Hadoop and noSQL. Central to the meetings was the emphasis on collaborative problem-solving and academic integrity, ensuring each member's contribution aligned with the project's goals. This efficient and systematic approach to the meetings, detailed in the appendix, was pivotal in steering the project to its successful completion.

# CONCLUSION

This essay highlights the crucial significance of Big Data in the banking and securities sector. In the discussion, several challenges were highlighted, among which were issues related to data security and privacy, the maintenance of data quality and integrity, adherence to regulations, and integration with existing systems. Despite their complexity, these issues were recognized as not being insurmountable. The progress of Big Data technologies, such as Hadoop, Apache Spark, and MongoDB, together with the rise of creative methods like AI-driven analytics and real-time processing, were seen as powerful answers to these difficulties.

The application of Big Data in banking has led to major improvements, including better customer profiling, enhanced risk management, and more efficient operational processes. The transformation of customer service, product development, and market strategies in banks using Big Data has been observed, becoming more responsive and adaptable to market shifts. As the industry evolves with digital advancements and increased data creation, Big Data is seen as key in shaping banking's future. Financial institutions are enabled to use data more effectively for decision-making, innovative services, and stronger risk management through Big Data.

From a forward-looking standpoint, the amalgamation of Big Data with developing technologies is entailed by the future of banking. Profound transformations in the banking industry are potentially brought about by the convergence of Big Data with technologies such as blockchain, the Internet of Things, and quantum computing. Unmatched efficiency, security, and customer-centric services are anticipated to be brought about by this integration, with the operations and interactions of banks with their clients being radically transformed.

To summarize, the progression of Big Data in the banking and securities sector is characterized by continuous expansion and advancement. The sector's capacity to adapt, accept, and employ data-driven insights is crucial in determining the success of the business and the transformation of banking in the digital era.

# INDIVIDUAL COMPONENT

## Group Work Evaluation

In our Big Data module's group coursework, strategic organization was a cornerstone of our success. Our group adopted a structured approach, organizing weekly meetings to ensure optimal collaboration and efficiency. These meetings served as platforms for task allocation and deadline setting, fostering a systematic and transparent workflow. I undertook the responsibility of documenting these meetings, ensuring a clear record of decisions and actions. This approach facilitated effective communication and decision-making, with each member's opinion valued and incorporated into a joint decision-making process.

Task allocation was strategically managed, with each member choosing tasks based on their expertise and interests. This methodology not only leveraged the unique strengths of each participant but also ensured a high level of engagement and ownership. The quality of our collective output was significantly enhanced by this tailored approach, as evidenced by the thoroughness and depth of our final report. Each member contributed to their sections, which were then collectively reviewed and integrated by Saeed and myself, ensuring a cohesive and comprehensive end product.

This process was not without its challenges, particularly in balancing the limited time available with other commitments such as our respective MSc Projects. However, this constraint also served as a valuable learning experience in time management and handling complex tasks. The evidence supporting our efficient group dynamics and effective task management can be found in our WhatsApp messages, and meeting minutes, all of which are organized in the appendices for easy reference and validation of our claims.

## Individual Contribution Assessment

Each team member's contributions significantly shaped our project's success. As a leader, I navigated our team through various challenges, ensuring smooth coordination and progress. Kiana and Saeed demonstrated exceptional problem-solving skills, efficiently addressing complex issues and offering innovative solutions. Their analytical abilities were pivotal in overcoming technical hurdles. Drici and Victor infused the project with creativity, bringing fresh perspectives and ideas that enriched our work. Their inventive approach was instrumental in differentiating our project. Soheib, Chinyere, and Oluwabukola’s investigative natures greatly expanded our research scope, enabling us to access diverse and accurate sources, thus enhancing the project's depth and validity.

While recognizing these strengths, it is also essential to identify areas for improvement to foster individual and collective growth. Chinyere, Victor, and Oluwabukola could benefit from further development in presentation skills, both in preparation and delivery, to more effectively communicate our findings. Similarly, Soheib, Drici, and Kiana could enhance their academic writing abilities, a vital aspect of presenting research in a structured and scholarly manner.

Saeed’s consistent and high-quality contributions across all areas stood out, leaving little room for improvement suggestions. His dedication and skill set were instrumental in guiding the project's direction and execution. My contributions, along with Saeed’s, were particularly notable in terms of effort and commitment, playing a crucial role in the project's overall direction and success. This assessment not only highlights each member's strengths and potential areas of growth but also underscores the collective effort that contributed to the project's achievement.

# REFERENCES

Vinaya Keskar, Jyoti Yadav, Ajay H. Kumar (2020) '5V’s of Big Data Attributes and their Relevance  
and Importance across Domains', *International journal of innovative technology and Exploring Engine,* 9(11), pp. 485-488 Available at: <https://doi.org/10.35940/ijitee.I3092.0789S319.>

Kashyap Vyas, (2023) Top 15 Big Data technologies you need to know <https://www.datamation.com/big-data/big-data-technologies/>

Will Hillier (2023) The Top 7 Big Data Tools for Data Analysts. <https://careerfoundry.com/en/blog/data-analytics/big-data-tools/>.

Mounika Narang (2023) Big Data technologies that everyone should know in 2024 <https://www.knowledgehut.com/blog/big-data/big-data-technologies#what-are-big-data%C2%A0technologies?%C2%A0>

Hitachi Solutions. (2023). Modern Data Analytics in Banking: Benefits, Outlook & More. [online] Available at: global.hitachi-solutions.com

N-iX. (2023). From Big Data to big opportunities: Data analytics in banking. [online] Available at: [www.n-ix.com](http://www.n-ix.com)

Projectpro. (2023) 'How JPMorgan uses Hadoop to leverage Big Data Analytics', projectpro.io, . Available at: https://www.projectpro.io/article/how-jpmorgan-uses-hadoop-to-leverage-big-data-analytics/142

'Big Data: Its Advantage, Challenges and Relevance', Proxy Infotech, . Available at: <https://www.proxyinfotech.com/knowledge_hub/big-data-its-advantage-challenges-and-relevance>.

Shalimov, A. (2023) Big Data in the banking industry: The main challenges and use cases, Eastern Peak. Available at: https://easternpeak.com/blog/big-data-in-banking-and-financial-services/ (Accessed: 07 January 2024).

Wissen Team (2023) The critical data analytics challenges in the banking industry, Wissen. Available at: https://www.wissen.com/blog/the-critical-data-analytics-challenges-in-the-banking-industry (Accessed: 07 January 2024).

Dvorski Lacković, Ivana & Kovšca, Vladimir & Lacković Vincek, Zrinka. (2020). A Review of Selected Aspects of Big Data Usage in Banks’ Risk Management. Journal of information and organizational sciences. 44. 317-330. 10.31341/jios.44.2.7.

Bedeley, R. (2014). Big Data Opportunities and Challenges: The Case of Banking Industry. In SAIS 2014 Proceedings (Vol. 2). Retrieved from <https://aisel.aisnet.org/sais2014/2>

Calle, J.P. (2022) Case study: JP Morgan Chase’s financial troubles, Pirani. Available at: https://www.piranirisk.com/blog/case-study-jp-morgan-chases-financial-troubles (Accessed: 07 January 2024).

Downs, E. (2018) Banks Need Non-Relational Data Solutions to Transform and Compete, MarkLogic. Available at: https://www.marklogic.com/blog/banks-need-non-relational-data-solutions-to-transform-and-compete/ (Accessed: 07 January 2024).

Mathur, V. (2022) Big Data In Banking Industry: Benefits, Uses and Challenges, Analytics Steps. Available at: https://www.analyticssteps.com/blogs/big-data-banking-industry-benefits-uses-and-challenges (Accessed: 07 January 2024).

Malik, P. (2013). Governing Big Data: Principles and practices. IBM Journal of Research and Development, [online] 57(3/4), pp.1:1–1:13. doi:https://doi.org/10.1147/jrd.2013.2241359.

Rhahla, M., Allegue, S. and Abdellatif, T. (2021). Guidelines for GDPR compliance in Big Data systems. Journal of Information Security and Applications, 61, p.102896. doi:https://doi.org/10.1016/j.jisa.2021.102896.

Khan, A.W., Khan, M.U., Khan, J.A., Ahmad, A., Khan, K., Zamir, M., Kim, W. and Ijaz, M.F. (2021). Analyzing and Evaluating Critical Challenges and Practices for Software Vendor Organizations to Secure Big Data on Cloud Computing: An AHP-Based Systematic Approach. IEEE Access, 9, pp.107309–107332. doi:https://doi.org/10.1109/access.2021.3100287.

atlan.com. (n.d.). 6 Benefits of Data Lineage: Why Businesses Are Eager to Invest. [online] Available at: https://atlan.com/data-lineage-benefits/ [Accessed 23 Dec. 2023].

McKinsey & Company (2021). Global Banking Practice Building the AI bank of the future.[online] Available at: https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/building%20the%20ai%20bank%20of%20the%20future/building-the-ai-bank-of-the-future.pdf.

Ross, S. (2022) What Are the Main Benchmarks That Track the Banking Sector?, Investopedia. Available at: https://www.investopedia.com/ask/answers/061715/what-are-main-benchmarks-track-banking-sector.asp (Accessed: 07 January 2024).

Taylor, D. (2023) Top 15 big data tools and software (Open source) 2024, Guru99. Available at: https://www.guru99.com/big-data-tools.html (Accessed: 07 January 2024).

'Hadoop - Pros and Cons', (2020) GeeksforGeeks, -06-27T14:58:49+00:00. Available at: https://www.geeksforgeeks.org/hadoop-pros-and-cons/ (Accessed: Jan 7, 2024).

'Big Data: Its Advantage, Challenges and Relevance', Proxy Infotech, . Available at: https://www.proxyinfotech.com/knowledge\_hub/big-data-its-advantage-challenges-and-relevance/ (Accessed: Jan 4, 2024).

Todor Ivanov Table 2 : Structured Streaming Pros and Cons. Available at: https://www.researchgate.net/figure/Structured-Streaming-Pros-and-Cons\_tbl1\_324364953 (Accessed: Jan 7, 2024).

# APPENDICES

Meeting Minutes 1

**Group Name/Number:** *Group 3 in GP12 / Banking and Securities*

**Meeting Date:** *20 Nov 2023*

**Meeting Time:** *2:30pm to 3pm*

**Location:** *Cavendish Campus, The University of Westminster, London*

**Attendees:** *Folusho Victor Arokoyo (FVA), Saeed Jamshidloo (SJ), Chinyere Unamba (CU), Oluwabukola Atere (OA), Hasan Guray (HG), Drici Mourad (DM), Kiana Rezaei Amrabadi (KRA)*

**Absences:** *-*

**Agenda:** Opening of the Meeting, Addressing Challenges and Concerns, Allocation of New Tasks, Setting Next Meeting Date and Time, Closure of Meeting

**Opening of the Meeting**

* *Meeting called to order at 2.30pm by HG.*

**Addressing Challenges and Concerns:**

* *Since there was a difference of opinion on whether the Banking and Securities sector should focus on the entire sector or specific working areas, an email will be sent to Ms. Yerashenia.*

**Allocation of New Tasks:**

* *Detailed Investigation part will be started:*
  1. *“Pinpoint major Big Data applications or projects prevalent in the industry.” under the “Project Landscape” assigned to FVA.*
  2. *“Understand the primary objectives of such ventures and assess their accomplishment*.*” under the “Project Landscape” assigned to CU.*
  3. *“Scrutinise the motivating factors and obstacles that led to these initiatives.” under the “Project Landscape” assigned to OA.*
  4. *“Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.” under the “Technology Adoption” assigned to SJ and KRA.*
  5. *“Decipher the reasons behind the choice of certain technologies over others.” under the “Technology Adoption” assigned to DM.*
  6. *“Combine and organize all research outputs along with the general research and collect them in a single file.” assigned to HG.*
  7. *“Preparing the first Meeting Minutes.” assigned to HG.*
  8. *“Send an email regarding the concern.” assigned to HG.*

**Setting Next Meeting Date and Time:**

* *Next meeting scheduled for 27 November.*

**Closure of the Meeting:**

* *Meeting adjourned at 3pm by HG.*

**Next Meeting:**

* *Date: 27 November 2023*
* *Time: 2.30 pm*
* *Location: Cavendish Campus, The University of Westminster, London*

**Action Items:**

* *“Pinpoint major Big Data applications or projects prevalent in the industry.”: Assigned to FVA, Due by 26 Nov.*
* *“Understand the primary objectives of such ventures and assess their accomplishment*.*”: Assigned to CU, Due by 26 Nov.*
* *“Scrutinise the motivating factors and obstacles that led to these initiatives”: Assigned to OA, Due by 26 Nov.*
* *“Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.”: Assigned to SJ and KRA, Due by 26 Nov.*
* *“Decipher the reasons behind the choice of certain technologies over others.”: Assigned to DM, Due by 26 Nov.*
* *“Combine and organize all research outputs along with the general research and collect them in a single file.”: Assigned to HG, Due by 27 Nov.*
* *“Preparing the first Meeting Minutes.”: Assigned to HG, Due by 21 Nov.*
* *“Send an email regarding the concern”: Assigned to HG, Due by 20 Nov.*

**Prepared by:** *Hasan Guray*

**Date:** *20 Nov 2023*

Meeting Minutes 2

**Group Name/Number:** *Group 3 in GP12 / Banking and Securities*

**Meeting Date:** *27 Nov 2023*

**Meeting Time:** *2:30pm to 3pm*

**Location:** *Cavendish Campus, The University of Westminster, London*

**Attendees:** *Folusho Victor Arokoyo (FVA), Saeed Jamshidloo (SJ), Chinyere Unamba (CU), Oluwabukola Atere (OA), Hasan Guray (HG), Drici Mourad (DM), Kiana Rezaei Amrabadi (KRA)*

**Absences:** *-*

**Agenda:** Opening of the Meeting, Review of Previous Meeting Minutes, Progress Update on Assigned Tasks, Discussion of Current Week’s Topics, Addressing Challenges and Concerns, Allocation of New Tasks, Setting Next Meeting Date and Time, Closure of the Meeting.

**Opening of the Meeting**

* *Meeting called to order at 2.30pm by HG.*

**Review of Previous Meeting Minutes:**

* *Summary of last meeting’s minutes presented by HG.*
* *Minutes approved as read and amended.*

***Progress Update on Assigned Tasks****:*

* *FVA reported on “Pinpoint major Big Data applications or projects prevalent in the industry.” under the “Project Landscape”*
* *CU reported on “Understand the primary objectives of such ventures and assess their accomplishment*.*” under the “Project Landscape”.*
* *OA reported on “Scrutinise the motivating factors and obstacles that led to these initiatives.” under the “Project Landscape”.*
* *SJ and KRA reported on “Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.” under the “Technology Adoption”.*
* *DM reported on “Decipher the reasons behind the choice of certain technologies over others.” under the “Technology Adoption”.*
* *HG reported on “Combine and organize all research outputs along with the general research and collect them in a single file.”.*
* *HG reported on “Send an email regarding the concern.”*

***Discussion of Current Week’s Topics:***

* *Topic 1 to 5: All of research outputs were missing references, it will cause issues for reporting compilation.*
* *Topic 6: All of the research outputs have been combined.*
* *Topic 7: An email sent regarding the issues, but already a detailed explanation for the coursework has been provided in the lecture, and that clarified the issue.*

**Addressing Challenges and Concerns:**

* *Owing to the omission of source citations in the research documentation, the applicability of the findings was significantly compromised, rendering the research indirectly utilizable.*
* *The research conducted necessitates bolstering through empirical case studies, complementing the broader theoretical examination of the sector.*

**Allocation of New Tasks:**

* *Detailed Investigation part will be started:*
  1. *“Pinpoint major Big Data applications or projects prevalent in the industry.” under the “Project Landscape” assigned to FVA.*
  2. *“Understand the primary objectives of such ventures and assess their accomplishment*.*” under the “Project Landscape” assigned to CU.*
  3. *“Scrutinise the motivating factors and obstacles that led to these initiatives.” under the “Project Landscape” assigned to OA.*
  4. *“Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.” under the “Technology Adoption” assigned to SJ and KRA.*
  5. *“Decipher the reasons behind the choice of certain technologies over others.” under the “Technology Adoption” assigned to DM.*
  6. *“Combine and organize all research outputs.” assigned to HG.*
  7. *“Preparing the Meeting Minutes.” assigned to HG.*
  8. *“Finding respective real life cases related to topic.” assigned to everyone.*

**Setting Next Meeting Date and Time:**

* *Next meeting scheduled for 4 December.*

**Closure of the Meeting:**

* *Meeting adjourned at 3pm by HG.*

**Next Meeting:**

* *Date: 4 December 2023*
* *Time: 2.30 pm*
* *Location: Cavendish Campus, The University of Westminster, London*

**Action Items:**

* *“Pinpoint major Big Data applications or projects prevalent in the industry.”: Assigned to FVA, Due by 3 Dec.*
* *“Understand the primary objectives of such ventures and assess their accomplishment*.*”: Assigned to CU, Due by 3 Dec.*
* *“Scrutinise the motivating factors and obstacles that led to these initiatives”: Assigned to OA, Due by 3 Dec.*
* *“Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.”: Assigned to SJ and KRA, Due by 3 Dec.*
* *“Decipher the reasons behind the choice of certain technologies over others.”: Assigned to DM, Due by 3 Dec.*
* *“Combine and organize all research outputs along with the general research and collect them in a single file.”: Assigned to HG, Due by 4 Dec.*
* *“Preparing the first Meeting Minutes.”: Assigned to HG, Due by 28 Nov.*
* *“Finding respective real life cases related to topic.”: Assigned to everyone, Due by 3 Dec.*

**Prepared by:** *Hasan Guray*

**Date:** *27 Nov 2023*

Meeting Minutes 3

**Group Name/Number:** *Group 3 in GP12 / Banking and Securities*

**Meeting Date:** *4 Dec 2023*

**Meeting Time:** *2:30pm to 3pm*

**Location:** *Cavendish Campus, The University of Westminster, London*

**Attendees:** *Folusho Victor Arokoyo (FVA), Saeed Jamshidloo (SJ), Chinyere Unamba (CU), Oluwabukola Atere (OA), Hasan Guray (HG), Drici Mourad (DM), Kiana Rezaei Amrabadi (KRA)*

**Absences:** *-*

**Agenda:** Opening of the Meeting, Review of Previous Meeting Minutes, Progress Update on Assigned Tasks, Discussion of Current Week’s Topics, Addressing Challenges and Concerns, Allocation of New Tasks, Setting Next Meeting Date and Time, Closure of the Meeting.

**Opening of the Meeting**

* *Meeting called to order at 2.30pm by HG.*

**Review of Previous Meeting Minutes:**

* *Summary of last meeting’s minutes presented by HG.*
* *Minutes approved as read and amended.*

***Progress Update on Assigned Tasks****:*

* *FVA reported on “Pinpoint major Big Data applications or projects prevalent in the industry.” under the “Project Landscape”*
* *CU reported on “Understand the primary objectives of such ventures and assess their accomplishment*.*” under the “Project Landscape”.*
* *OA reported on “Scrutinise the motivating factors and obstacles that led to these initiatives.” under the “Project Landscape”.*
* *SJ and KRA reported on “Recognize the key technologies and tools in use, such as Hadoop, noSQL, in-memory data processing, data streaming, etc.” under the “Technology Adoption”.*
* *DM reported on “Decipher the reasons behind the choice of certain technologies over others.” under the “Technology Adoption”.*
* *HG reported on “Combine and organize all research outputs along with the general research and collect them in a single file.”.*
* *HG reported on “Preparing the Meeting Minutes.”.*
* *Everyone reported on “Finding respective real life cases related to topic.”.*

***Discussion of Current Week’s Topics:***

* *Topic 1 to 5: All of research outputs were with references and very comprehensive.*
* *Topic 6: All of the research outputs have been combined.*
* *Topic 7: Meeting Minutes has been successfully prepared.*
* *Topic 8: So many real life cases have been explored.*

**Addressing Challenges and Concerns:**

* *There is so limited time to prepare the presentation, so the main focus should be presentation this week.*

**Allocation of New Tasks:**

* *Presentation will be prepared:*
  + *“Project Overview” slide will be prepared and presented by CU and OA.*
  + *“Technology Adoption” slide will be prepared and presented by DM and FVA.*
  + *“Analysis of Impact and Solutions” slide will be prepared and presented by KRA and SJ.*
  + *“Data Governance and ROI” slide will be prepared and presented by HG.*
  + *“Challenges Faced During the Research” slide will be prepared and presented by SJ.*
  + *“References” slide and cover page will be prepared by HG.*
* *Research will be continued.*
  + *“Impact Analysis” will be researched by CU, OA, and FVA.*
  + *“Solution Analysis” will be researched by KRA; SJ, DM, and HG.*
  + *All of the outputs will be consolidated by HG.*
* *“Meeting Minutes” will be prepared by HG.*

**Setting Next Meeting Date and Time:**

* *Next meeting scheduled for 8 December.*

**Closure of the Meeting:**

* *Meeting adjourned at 3pm by HG.*

**Next Meeting:**

* *Date: 8 December 2023*
* *Time: 9.30 am*
* *Location: Online Zoom Meeting*

**Action Items:**

* *“Presentation will be prepared.”: Assigned to everyone, Due by 7 Dec.*
* *“Research will be continued.”: Assigned to everyone, Due by 7 Dec.*
* *“Meeting Minutes will be prepared”: Assigned to HG, Due by 4 Dec.*

**Prepared by:** *Hasan Guray*

**Date:** *4 Dec 2023*

Meeting Minutes 4

**Group Name/Number:** *Group 3 in GP12 / Banking and Securities*

**Meeting Date:** *8 Dec 2023*

**Meeting Time:** *9:30am to 10:30am*

**Location:** *Online Zoom Meeting*

**Attendees:** *Folusho Victor Arokoyo (FVA), Saeed Jamshidloo (SJ), Chinyere Unamba (CU), Oluwabukola Atere (OA), Hasan Guray (HG), Drici Mourad (DM), Kiana Rezaei Amrabadi (KRA), Soheib Kohneposhi (SK).*

**Absences:** *-*

**Agenda:** Opening of the Meeting, Review of Previous Meeting Minutes, Progress Update on Assigned Tasks, Discussion of Current Week’s Topics, Addressing Challenges and Concerns, Allocation of New Tasks, Setting Next Meeting Date and Time, Closure of the Meeting.

**Opening of the Meeting**

* *Meeting called to order at 9.30am by HG.*

**Review of Previous Meeting Minutes:**

* *Summary of last meeting’s minutes presented by HG.*
* *Minutes approved as read and amended.*

***Progress Update on Assigned Tasks****:*

* *Presentation has been presented and recorded, also uploaded.*
* *All of the research outputs have been combined.*
* *Meeting Minutes have been prepared.*

***Discussion of Current Week’s Topics:***

* *Topic 1: The first 30 minutes have been spent for presentation. SJ recorded and uploaded it.*
* *Topic 2: The research was so comprehensive and proper for the coursework.*
* *Topic 3: Meeting Minutes has been successfully prepared.*

**Addressing Challenges and Concerns:**

* *Everyone needs to upload presentation link and file to Student Hub before 1pm.*

**Allocation of New Tasks:**

* *Research will be continued.*
  + *“Data Governance & ROI” will be researched by CU, OA, and FVA.*
  + *All of the outputs will be consolidated by HG.*
* *“Meeting Minutes” will be prepared by HG.*

**Setting Next Meeting Date and Time:**

* *Next meeting scheduled for 23 December.*

**Closure of the Meeting:**

* *Meeting adjourned at 10:30am by HG.*

**Next Meeting:**

* *Date: 23 December 2023*
* *Time: 10 am*
* *Location: Online Zoom Meeting*

**Action Items:**

* *“Research will be continued.”: Assigned to everyone, Due by 22 Dec.*
* *“Meeting Minutes will be prepared”: Assigned to HG, Due by 8 Dec.*

**Prepared by:** *Hasan Guray*

**Date:** *8 Dec 2023*

Meeting Minutes 5

**Group Name/Number:** *Group 3 in GP12 / Banking and Securities*

**Meeting Date:** *23 Dec 2023*

**Meeting Time:** *10am to 11am*

**Location:** *Online Zoom Meeting*

**Attendees:** *Folusho Victor Arokoyo (FVA), Saeed Jamshidloo (SJ), Chinyere Unamba (CU), Oluwabukola Atere (OA), Hasan Guray (HG), Drici Mourad (DM), Kiana Rezaei Amrabadi (KRA), Soheib Kohneposhi (SK).*

**Absences:** *-*

**Agenda:** Opening of the Meeting, Review of Previous Meeting Minutes, Progress Update on Assigned Tasks, Discussion of Current Week’s Topics, Addressing Challenges and Concerns, Allocation of New Tasks, Closure of the Meeting.

**Opening of the Meeting**

* *Meeting called to order at 10am by HG.*

**Review of Previous Meeting Minutes:**

* *Summary of last meeting’s minutes presented by HG.*
* *Minutes approved as read and amended.*

***Progress Update on Assigned Tasks****:*

* *All of the research outputs have been combined.*
* *Meeting Minutes have been prepared.*

***Discussion of Current Week’s Topics:***

* *Topic 1: The research was so comprehensive and proper for the coursework.*
* *Topic 2: Meeting Minutes has been successfully prepared.*

**Addressing Challenges and Concerns:**

* *We have so limited time to submit the coursework. Everyone should catch the deadline.*
* *Harvard referencing style should be used for in-text citations.*
* *Everyone needs to check the academic report and share the feedbacks.*

**Allocation of New Tasks:**

* *“Academic Report Compilation” will be started.*
  + *“Introduction” part will be prepared by FVA. (The relationship between 5Vs of Big Data and Banking sector will be investigated.)*
  + *“Challenges & Data Landscape” part will be prepared by CU and OA.*
  + *“Technology & Solution Analysis” part will be prepared by DM and SK.*
  + *“Outcomes & Reflection” part will be prepared by KRA and SJ.*
  + *“Meeting Minutes” will be consolidated by HG.*
  + *“Conclusion” part will be prepared by HG and SJ.*
* *“Meeting Minutes” will be prepared by HG.*

**Closure of the Meeting:**

* *Meeting adjourned at 11am by HG.*

**Action Items:**

* *“Academic Report Compilation will be started”: Assigned to everyone, Due by 3rd Jan.*
* *“Meeting Minutes will be prepared”: Assigned to HG, Due by 23 Dec.*

**Prepared by:** *Hasan Guray*

**Date:** *23 Dec 2023*

WhatsApp Message 1

A screenshot of a chat

Description automatically generated

WhatsApp Message 2

A screenshot of a chat

Description automatically generated