DOI: https://doi.org/10.37075/EA.2021.4.03 Economic Alternatives, 2021, Issue 4, pp. 534-546

Financial Data Processing in Big Data Platforms

Received: 22.07.2021
Available online: 28.12.2021

Luben K. Boyanov*

Abstract

Today's digital society generates more and more data on a daily basis in all areas of human activities, especially in the financial sector. Such data can be collected, stored, processed, and analyzed, providing serious analytical opportunities for the end users. A lot of such systems are implemented and work using cloud technologies, which have a number of advantages, but they use a pay-per-use model and thus are not very suitable for medium and small organizations, non-profit and academic institutions. In this paper, a system, capable of fetching, storing, and processing big data is proposed and tested with financial data. It uses an open-source component-based approach and can be custom-built and implemented in national universities or centers of competence/excellence. That can present unique opportunities to researchers and developers to use and work with Big data on economic and financial problems, to investigate dependencies, use large simulation and forecast models and analyze results, using the new technologies and Big data provided by them.

Keywords: Big data, IT platforms, financial data, big data processing.

JEL: L86, O33, O3, O32

1. Introduction

The financial sector has always been closely linked with data processing. It has been one of the first that has utilized computing devices and has also been in the front line of digital computer applications benefiting from the advantages of machine data processing. The last few decades have further accelerated and enhanced the relations Finance – Information Technology (IT). With the mass employment and utilization of Internet and cell phones, with almost any human activity being projected in the IT world, the global digitalization has presented unthinkable so far opportunities to collect, store and process financial information from almost any spot in the world in real time. The transformation from using some and many digital data to having Big data has appeared. Big data is about enormous volumes of data, arriving in a big variety and at high velocity. Big data impact is changing the world landscape, as people have never been able to collect, store, process and analyze such amount of information, making the entire transformation sustainable (Pappas et al. 2018). The financial sector, being so dynamic is a big player in the field – Big data is already part of corporate finance, financial accounting, auditing, compliance, reporting, federal reserve, capital markets, brokerage and trading, banking,

* Department of Information Technologies and Communications, University of National and World Economy

insurance, real estate, money management, etc. (Oberst 2015). All those areas generate enormous amount of data. It can be structured (managed in traditional databases), semi-structured (does not have a formal data model but has tags and/or markers that can separate data elements) and unstructured (varying significantly in organization and not suitable for processing in relational databases). Structured data is typically stored and controlled within an organization and is used for operational decision making. Unstructured data (video, audio, emails, social media content, web sites, and others) holds the biggest volumes and varieties and offers unthinkable so far opportunities. It also offers significant opportunities for analytics, unknown so far. As one author puts it: “Big data in finance refers to the petabytes of structured and unstructured data that can be used to anticipate customer behaviors and create strategies for banks and financial institutions.” (Big Data Finance, 2019.) Data is becoming more valuable than ever and calling it “the new gold” is no news. Even so, big financial players continue to attract the attention to it (Mark Cuban: ‘Data Is the New Gold’ 2021). People using IT applications can start exploring the values, hidden in Big data. This paradigm holds also significant potential for new areas of analysis. Predictions in financial markets can lead to big profits and reduce risks. Fraud detection and prevention can minimize losses. Using big data, provided by the customer himself improves sales, and enhances the customer’s performance (Hallikainen, Savimäki, and Laukkanen 2020). Banks like Deutsche bank have published White papers on their use of Big data (Ermetic Reports 2020). No one questions whether to use IT and Big data but how to use it and what tools and platforms are the most suitable for that.

2. Big data applications and approaches of using it

2.1. Area of big data applications in finance

Any business relies on data and its monitoring, but few can be so dependent on the efficiency of data processing as the field of finance because big money is at stake. Big data effect and analytics is already influencing sales (Hallikainen, Savimäki, and Laukkanen 2020), banking (Hale and Lopez 2019), financial markets (Shen and Chen 2018), predictive econometric modeling (Diebold et al. 2019), stock market insights (Mallon 2021), risk management (Dicuonzo et al. 2019), fraud detection and prevention (Hasan, Popp, and Oláh 2020; Daliri 2020) and many more. And all this leads to benefits and makes the companies using big data perform much better (Duan and Xiong 2015). Profits and performance are not the only advantages – big data usage also creates transparency, analyzes better the risk, leverages consumer data and transforms culture (Razin n.d.). Big data is the contemporary paradigm to gain competitive advantage. The mentioned areas/segments of use of Big data in finance can improve acquisition efficiency, raise the effectiveness of marketing campaigns targeting the right set of customers and personalizing marketing messages, monitor customers’ opinion from social media, predict services and products in which customers are interested, optimize pricing, generate cross-selling opportunities (notifications in specific cases), identify valued and profitable customers, enhance loyalty, etc. One thing that has to be taken into account when using financial big data is the GDPR directive because this kind of data is very sensitive and requires proper policies, handling and an appropriate level of responsibility (Tsaneva 2019). Other novel topics and approaches affecting work with stock markets are cyber-physical systems (Xue et al. 2017) and blockchain (Wang Qian 2021).

2.2. On-premises and cloud systems

A typical on-premises system may not be well suited for gathering and processing Big data, so many big users turn to cloud computing. The cloud paradigm has become very popular, widely deployed for a number of reasons. It offers a lot of benefits – scalability (probably one of the most important), good management and security, pay-per-use, cuts costs of on-site hardware and software, etc. Financial institutions transform well their digital business with cloud computing (Burke 2019). On-premises systems require complex equipment and well-trained IT experts, and this may be a very costly approach. Clouds can offer IaaS (Infrastructure as a Service) – this includes servers, storage, middleware, virtualizations, networking, etc. Clouds can also provide PaaS (Platform as a Service) – an environment for developing, testing, managing software applications, etc. Another class of cloud service is SaaS (Software as a Service) – a software application or a stack of applications, managed by the cloud provider. Clouds can also deliver CaaS (Container as a Service) – container-based virtualization with a complete set of containers, applications, and clusters. Another benefit of the cloud paradigm is that it can respond to the rapid and unexpected demands of the customers. The use of cloud computing is rising, as it is also in the financial sector (EBF-Cloud-Banking-Forum 2020).

It should be noted that there is still a number of drawbacks when using Cloud computing. Probably the major one is security – even though cloud services offer a high level of security, the identity of users, the security of data, and its confidentiality may be at risk, which is a well-known significant concern. A recent report announces that nearly 80% of companies have experienced a cloud data breach in 2019-2020 (Ermetic Reports 2020). In addition, there can be a problem with the adaptation of fluctuations in clients' needs (in case no cloud can offer such new service/infrastructure); possible delays of the communication network; vulnerability in case the cloud is attacked, as well as data mobility in the cloud can all be a problem. The pay-per-use principle may be problematic for academic institutions and small and medium enterprises, which are still not sure how and in which aspect to use possible Big data benefits.

2.3 Big data tools

Nowadays developers can employ numerous tools (and their number is growing each year), used for work with Big data. They are used to fetch data and information from data sets, web sites and other data sources, after which data has to be stored in a scalable and reliable manner and be processed. The large data volumes require control and management of large and scalable data storage. The big variety needs flexible and adaptable tools and approaches. The Big data velocity points out the need of fast processing, which in turn requires fast and scalable hardware and new models and frameworks for processing such kind of data. All these are the challenges, which new tools and hardware have to confront, as traditional frameworks and models are unable to cope with them.

When speaking of tools for dealing with Big data, one must start with Hadoop – the best known big data tool, which is used by more than half the Fortune 50 businesses (CEO 2020). Hadoop's software library allows big data sets to be processed in a distributed way across clusters of computers using simple programming models ("Apache Hadoop" n.d.). It is designed to scale up from single servers to thousands of machines, providing parallel processing and storage of data. Hadoop is a project of the world's largest open source foundation – Apache, which has more than 300 top-level projects with about 2 Petabytes source code downloads ("Apache Projects List" 2021).

Once data is generated from a data source (can be a Web site, sensor, database, etc.), it must be retrieved through different protocols and communication channels, transformed into suitable formats for further processing and loaded. This process is referred to as Extract, Transform, Load (ETL) and it is the interface between the "outside" world and the system. The well known solutions for such kind of tools are Apache NiFi, Eclipse Kura, Streamsets, Azure Data Factory and others.

One of the main characteristics of all Big data systems is the volume and velocity of data- i.e., the necessity to process fast large volumes of data. This requires a component that can take a heavy load of data and events and distribute them to the respective components that will store, process, and analyse them. A well-known open-source tool for distributed event streaming by Apache is Kafka. It is used by more than 80% of all Fortune 100 companies and the majority of banks and insurance companies (Apache Kafka n.d.). This tool delivers messages in high speed on clusters of machines and is typically used as a "broker" in Big data processing systems. Other commonly used solutions for this purpose are Amazon Kinesis, RAITMQ, ActiveMQ and others.

Once the data has been generated, extracted, cleaned, loaded, and stored, the time comes to process the data. Data processing components are large in variety and have different applications depending on the data one works with in the specific implementation. The most commonly used tools for providing solutions to this are Apache Hive, Apache Impala, Apache Spark, Amazon Redshift, Azure Synapse Analytics and others.

The final goal of each data extraction, storage and processing is to provide the information needed for specific decisions. This is possible by employing visualization components for the extracted dependencies. At this point, the users are interested in reports obtained from the processed data. The tools used for this functionality of a Big data architecture are Microsoft Power BI, Tableau, Qlik Sense, Google Data Studio, and others.

Some companies like Cloudera (Lambiente, 2019) provide end-to-end solutions and architectures that use a number of the abovementioned tools. Cloudera uses also gateways like Eclipse Kura, Integration modules like Eclipse Kapua, etc. The important point here is that using open-source solutions/tools with an end-to-end solution eases the activities of the developers. Other well-known end-to-end solutions are offered by Amazon (AWS), Microsoft (Azure), IBM (Watson), Oracle (Oracle Big Data) and others.

Other Big data tools are HPCC (free single platform implemented in a unified architecture with a common programming language for data processing), Qubole (an open-source tool which is self-managed, self-optimizing and allows the data team to focus on business outcomes), HPE (MapR Technologies with HPE Ezmeral Data Fabric – a Big data platform, though somehow costly) etc.

3. Building an on-premises platform for big data

3.1 Basic functional blocks

There are several architectural and functional blocks, needed for a big data processing platform to be useful. Before we look at the proposed organization, first there has to be a Data source block. This is the source of data that will feed the platform with data, that has to be stored, processed, and analyzed. The Data source block can send data directly to the platform or can use Internet or a mobile network. This block can also be a Web site which is the source of some IoT, financial or any type of data. At the Data source, one might have some preprocessing (for example some microcontrollers and microprocessors), or some edge and fog device. The concept of edge and fog computing is about the initial processing of data, which improves the efficiency of the data sent to the place for processing and reduces the requirements for the communication bandwidth (Fog and Edge Computing: Principles and Paradigms 2019).

Upon sending data over Internet, the mobile network or communication channels, data arrives at the premise where the big data platform resides – Figure 1.

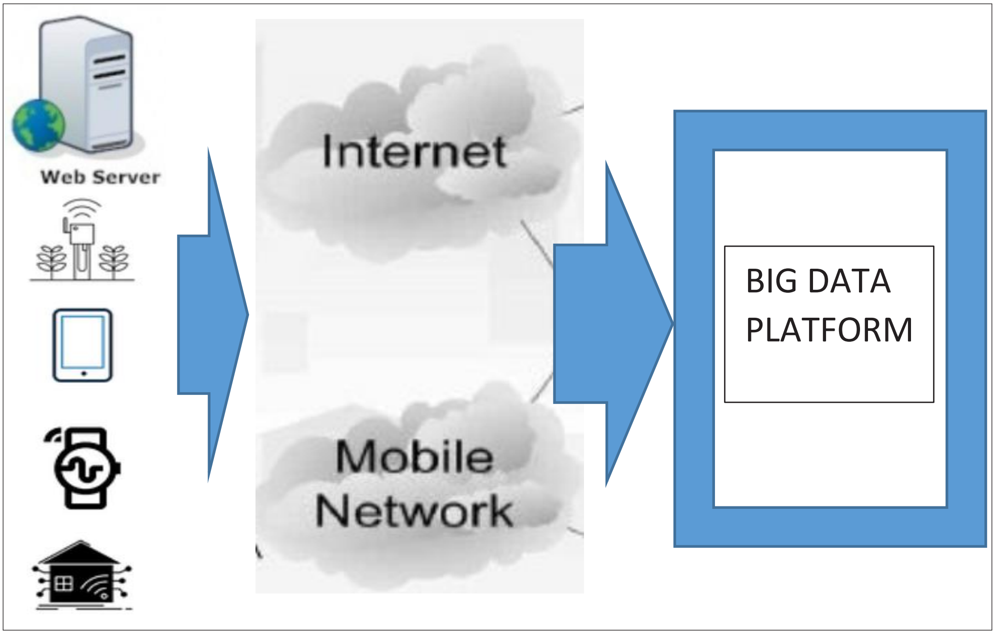

Figure 1. Data source block – communication networks – big data platform.

The main tasks for the Big data platform are to receive/extract, transform, load, store, process, and provide analytics for the end user. This is the chain of functional steps, needed for the procedure of taking data and using it for analytical purposes.

3.2 Requirements for a Big data platform

What are the requirements that we set for our Big data platform? It has to be modular, scalable, open source and non-complex because it has to be implemented, developed, and sustained at a place that has no capacity of hiring a large and highly trained IT staff. The modular structure will allow alternative or improved tools to be employed if needed. The scalability is necessary for the Big data model. Open source will overcome the need of dependence of one user and the availability of solutions and advice of the developers' community. Being non-complex is important as this platform is intended to be run on premises like universities, centers of competence, academic research institutions, associations of SMEs, etc. which have some experts, researchers, and students to work on the model.

The first module of our proposed Big data platform is the ETL one. Its functionalities are to integrate several processes. Such are receiving data from multiple data sources, following by managing them (transforming) into a single and coherent data store. The final step is saving (loading) that data into a repository (data warehouse or similar system). We have chosen this module to be implemented by Apache NiFi. Its fundamental design concepts are closely related to flow-based programming. One of the important concepts here is the FlowFile Processor. A NiFi Flow Processor combines data routing, data transformation and a certain level of integration between systems. Processors have access to attributes of a given FlowFile and its content stream. A very important feature for the developers is the visual creation and management of directed graphs of the processors, an example of which will be given bellow.

The second module of our platform is the data-broker. Its functionalities are to collect information about the places where data will be dispatched, according to different, predefined criteria. Such criteria can be set on the basis of profiles of the tasks to be completed. We have chosen this module to be implemented by Apache Kafka.

The third module is the Data storage. It must save the data, arriving from "things" or sites (or any other sources) after being extracted and transformed or after being dispatched by the broker. For this module we have chosen Hadoop and its HDFS.

The next module is the Data processing module. Its functionalities include responses to queries, data management, data access and any other data processing, which can be found in a database data processing module. For this module we have chosen to use two products – Apache Hive and Apache Impala. They have the same functionalities but differ in speed of processing and precision.

The last module is the analytical one. It has the task to provide analytical capabilities to help the end users with insight in the processed data, providing data visualization through dashboards, interactive reports that brings insights within the Big data. For this module we have chosen Microsoft's Power BI. Compared to the alternatives (Tableau, Qlik Sense, Google Data Studio, and others), Microsoft Power BI is easy to install and integrate with other components. Microsoft Power BI is a powerful tool available to academia and relies on Microsoft for support and development. The platform with its modules is presented in Figure 2.

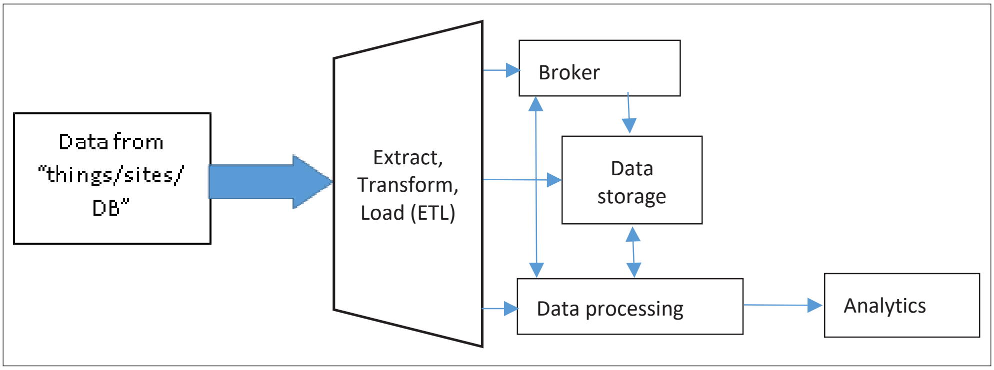

Figure 2. The modules and their interconnection in the Big data platform.

4. Data use and sources

There are various sources of financial data for big data use but many of them are with subscription and a paid option. Such are Nasdaq's Quandl ("Quandl", 2021). Still one look at places like the World Bank Open Data, IMF Economic Data, the UN Comtrade Database, the Global Financial Data, Google Finance, Google Pubic Data Explorer, the US Bureau of Economic Analysis, Financial Data Finder at OSU, the National Bureau of Economic Research, the US Securities and Exchange Commission, the Visualizing Economics site, the Financial Times, etc. (70 Amazing Free Data Sources…, 2017). Even in this list there are some details. For example, Google Cloud Public Datasets are exempt of charges with a Google account and for small queries (defined as up to 1 TB/month in queries) but for queries over that volume, there is a price for the queries. Another condition is that using those datasets, the customer pays only for the resources which the cloud is using to analyze the data. These are resources like computing or storage for the required customer applications.

For the purpose of verification of our platform, using an open source approach, we have chosen an open source financial data. The FinnHub platform (Finnhub.io, 2021) is used. It offers a wealth of data on various financial instruments such as company actions, currencies (including cryptocurrencies) and more. The data has access to two levels – for developers (developer) and for companies (enterprise). The data is in real time and encompases a period of the last 20 years (enterprise level).

For our testing purposes, we have selected IoT data on stock market trading (Kenton 2021) of three large companies, namely Cloudera, Google and Microsoft. Their financial data on trading in shares are loaded every 5 minutes under the following conditions:

- sending a request to the FinnHub platform for financial data;
- receiving financial data from the FinnHub platform;
- processing and storage of financial data in Apache Hadoop (HDFS);
- complete the previous three steps in less than 5 minutes;
- repeat the operations for collecting the financial data in 5 minutes following the end of the previous cycle.

The implementation of this cycle has been performed using the proposed Big data platform/architecture. As listed above the following components have been employed:

- Apache NiFi data stream (the ETL tool);
- Apache Kafka themes (the broker tool);
- Apache Hadoop HDFS file system (the storage tool);
- Apache Hive (Apache Impala) using external tables for storing financial data (the data processing tool(s));
- Power BI (the analytics and visualization tool).

The creation of the necessary directory structure in the Apache Hadoop HDFS is done by executing appropriate commands in the Linux kernel. The creation of external tables with data on the trading of shares of Cloudera, Google and Microsoft from the FinnHub platform is done using DDL SQL query executed in the Hue editor (or in Hive beeline or in Impala's impala-shell).

Update of financial data. After creating the directory structure in Apache HDFS and the external tables in Apache Hive (Apache Impala), we can proceed to loading the financial INO data from the FinnHub platform. This is done using the components Apache NiFi, Apache Kafka, and Apache Hive (Apache Impala) as in the created Big data platform.

The different threads (flow of data) are built using the following Apache NiFi modules (processors):

- **Connect via HTTPS** (**InvokeHTTP)** – a NiFi created processor that connects to the FinnHub platform via HTTPS (Hypertext Transfer Protocol Secure). This processor displays the current financial data for trading in shares of Cloudera, Google and Microsoft. The data is fed to the next processor – EvaluateJsonPath;
- **Retrieve values** (**EvaluateJsonPath)** – a NiFi processor that processes the obtained response from the FinHub platform. The financial data are received in JSON format (JSON payload) and from this JSON, the actual data values of the shares of one of the companies (Cloudera, Google, and Microsoft) are fetched. The EvaluateJsonPath processors for the three companies are similar. The control is then passed to the next processor in the stream – ReplaceText
- **Clean values** (**ReplaceText)** – a NiFi processor that replaces the received from FinnHub JSON data with the values of the shares and the present time in ***unixtime*** format. Data is passed to the next processor – PublishKafka;
- **Put data into proper topics** (**PublishKafka)** – a NiFi processor that loads the data from the ReplaceText processor and publishes them in a separate Apache Kafka topic for current data on the company's activities;
- **Extract data from topics** (**ConsumeKafka)** – a NiFi processor that extracts the data from the Apache Kafka topic for current values of the company's activities and passes them for processing to the next processor – UpdateAttribute.
- **Append a file name** (**UpdateAttribute)** – a NiFi processor that appends a file name to the flow attributes in the Apache Hadoop HDFS, where the current activity data of the respective company from FinnHub's InO platform will be recorded. When this operation is successful, the control is passed to the next processor – PutHDFS;
- **Save data** (**PutHDFS)** – a NiFi processor that records the data for the current values of the shares of the company. This is done in a file in Apache Hadoop HDFS

that is linked to the corresponding external table of Apache Hive.

The flow of financial data in the Big data platform with the NiFi processors is presented in Figure 3.

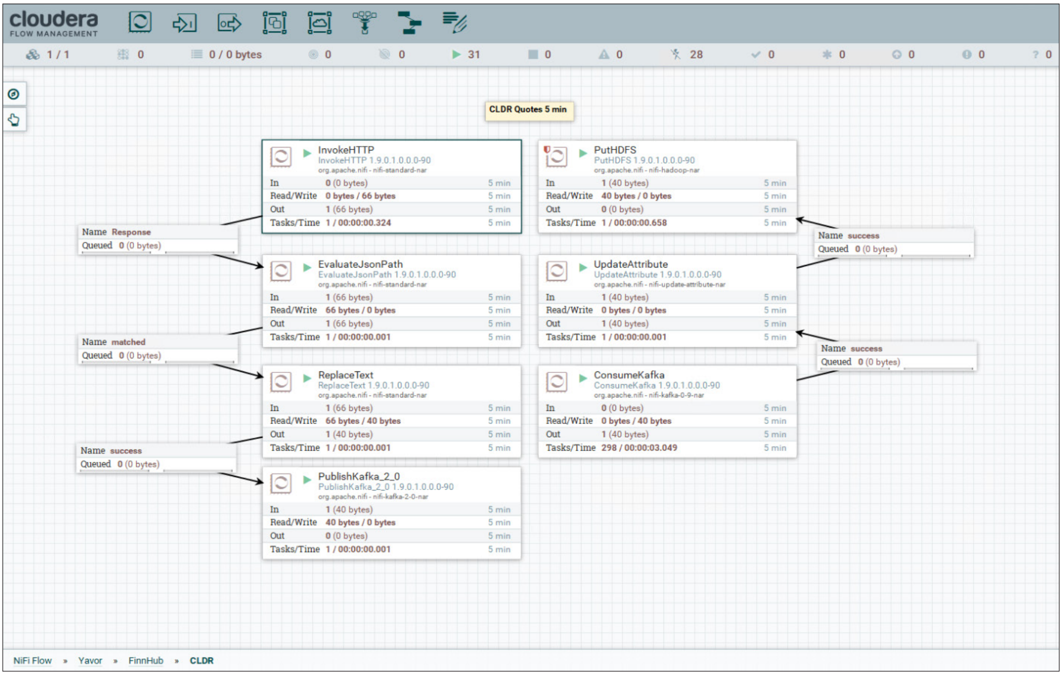

Figure 3. Flow of financial data in the proposed Big data platform.

The presented above approach demonstrates how the external Apache Hive tables are populated with a new record with the new data for the respective company (Cloudera, Google, and Microsoft in this case).

The cycle is repeated every 5 minutes. In case of failure of any of the NiFi processors, data is not submitted to Apache Kafka and in turn – they are not passed and recorded in Hadoop HDFS and Apache Hive (Apache Impala). This will happen in the next iterations, if data for the current values for the shares of the respective company is obtained successfully from the FinHub platform.

5. Testing the platform with financial data and description of the results

The validation of the proposed platform has been done, experimenting with data from stock market trading of the shares of Cloudera (CLDR), Google (GOOG), and Microsoft (MSFT). The data used covers a period of three months (November 2020 – February 2021) with more than 25,000 records for each company.

Based on the results of an experiment, where we have compared the speed of loading data in Microsoft Power BI using both Apache Impala and Apache Hive, a decision has been made to load the data from the stock market only with Apache Impala. Figure 4 shows the chart of Microsoft's stock price variation. The four colors show the current (light blue), high (dark blue), low (red), open (dark red) and the previous closed (pink) prices. The charts of the shares of Google and Cloudera look similar.

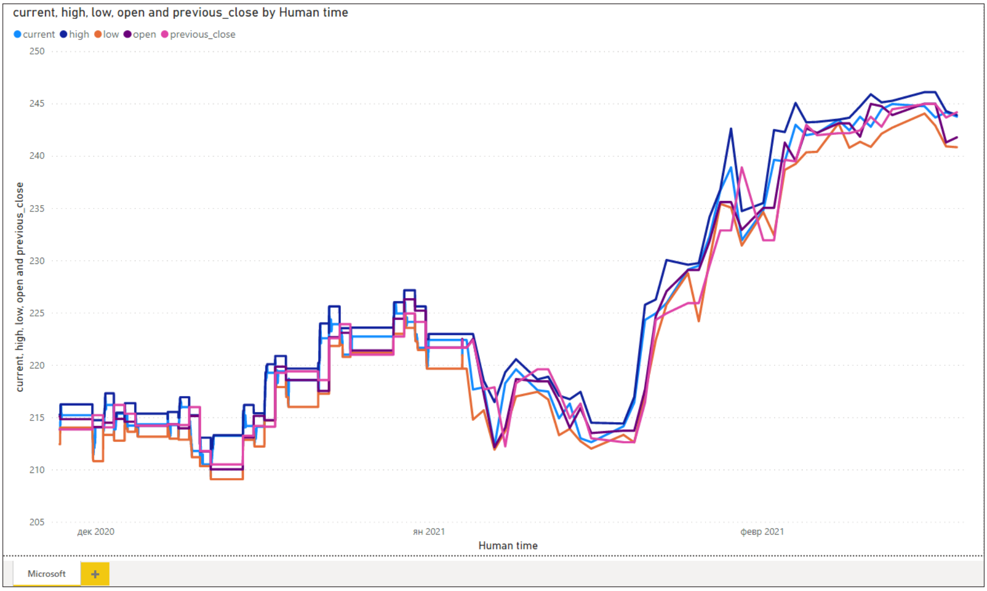

Figure 4. Microsoft stock price variation for the period November 2020 – February 2021.

The chart clearly shows a weekly cycle of the variations of Microsoft's shares until the beginning of January 2021. After Janurary, 4th 2021, the FinnHub platform changed the conditions under which developers can use data from its platform. After this date, the shares are given with daily values (unlike the weekly such until that point). This can be observed in the chart above – until the start of January 2021, there are flat, weekly periods, which are absent after that. The result also demonstrates another functionality of the proposed Big data platform – namely that it can adapt itself to changes in the conditions of the fetched data without need for additional intervention. The visualization of the results and its parameters prove experimentally the adequacy of the applied Big data platform for processing and analysis of financial data. The experiments have been carried out at the data cluster at the University of National and World Economy. The cluster is built in the Centre of Competence on "Digitalization of Economy in an Environment of Big Data" and provides services for digitalization and digital transformation in various economic sectors. The system consists of 40 data nodes, each with 128GB RAM memory and a total storage of 4,5PB. With this (and similar systems), researchers and developers can run their projects on the necessary hardware and software and handle Big data without using expensive public cloud services, and without needing to build their own infrastructure. Comparing our system to other presented approaches and architectures shows the straightforwardness of our methodology. For example in (Project Pro 2021) one can see 3 sources for data indigestion and 7 sources for data management, while in (Zhang, Shi, and Khan 2017) there are three layers with 3 to 5 blocks in each layer.

6. Conclusion

This study demonstrated an approach of creating and implementing a system for Big data processing in the area of finance. It is an example of a methodology on how to design and build such a system that can be useful and affordable for SMEs, non-profit, and academic organizations. Processing big data is the next level of knowledge and success, and in such sense – the next level of organizing business, finance and improving their efficiency.

Following the principle of using well known and widely spread tools for Big data processing (open-source instruments, most of which belong to the Apache project and are used in almost all Big data applications), we have demonstrated that the creation of such a platform is feasible for academic institutions, research organizations and centers of excellence. There is an even greater prospect in this approach – it allows adding and using modules for Machine Learning and Artificial Intelligence. There are a number of such tools, and we intend to experiment with KNIME, a free analytics open-source software that creates visual workflows with an intuitive, drag-and-drop style graphical interface ("Open for Innovation" n.d.).

The proposed approach can be of great help to academics, research and even some government (or local government) authorities as it can be used to investigate and solve a variety of issues, not only related to financial data, as demonstrated in the present work. The promotion of platforms and efficient solutions using Big data is a unique opportunity for accelerating the digital transformation of our society.

Acknowledgements

This work has been supported by the project NoBG05M2OP001-1.002-0002 "Digital Transformation of Economy in Big Data Environment", funded by the Operational Program "Science and Education for Smart Growth" 2014–2020, Procedure BG05M2OP001-1.002 "Establishment and development of competence centers".

References

"70 Amazing Free Data Sources You Should Know." n.d. *KDnuggets* (blog). Accessed May 15, 2021. https://www.kdnuggets.com/70-amazing-free-data-sources-you-should-know.html/.

"Apache Hadoop." n.d. Accessed May 6, 2021. https://hadoop.apache.org/.

"Apache Kafka." n.d. Apache Kafka. Accessed May 6, 2021. http://kafka.apache.org/.

"Apache Projects List." 2021. February 25, 2021. https://projects.apache.org/projects.html?category#big-data.

Big Data Finance. 2019. "Big Data in Finance: Use Cases, Examples, Challenges, and Getting Started." 2019. https://bigdatafinance.tw/index.php/finance/1214-big-data-in-finance-use-cases-examples-challenges-and-getting-started.

Burke, Brian. 2019. *How Cloud Computing Is Transforming and Benefiting Financial Institutions*. https://doi.org/10.13140/RG.2.2.17262.46404.

CEO. 2020. "Top 5 Best Big Data Tools." *The CEO Views* (blog). August 10, 2020. https://theceoviews.com/top-5-best-big-data-tools/.

Daliri, Sajjad. 2020. "Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System." *Computational Intelligence and Neuroscience* 2020 (February): e6503459. https://doi.org/10.1155/2020/6503459.

Dicuonzo, Grazia, Graziana Galeone, Erika Zappimbulso, and Vittorio Dell'Atti. 2019. "Risk Management 4.0: The Role of Big Data Analytics in the Bank Sector." *International Journal of Economics and Financial Issues* 9 (October): 40–47. https://doi.org/10.32479/ijefi.8556.

Diebold, Francis X., Eric Ghysels, Per Mykland, and Lan Zhang. 2019. "Big Data in Dynamic Predictive Econometric Modeling." *Journal of Econometrics*, Big Data in Dynamic Predictive Econometric Modeling, 212 (1): 1–3. https://doi.org/10.1016/j.jeconom.2019.04.017.

Duan, Lian, and Ye Xiong. 2015. "Big Data Analytics and Business Analytics." *Journal of Management Analytics* 2 (1): 1–21. https://doi.org/10.1080/23270012.2015.1020891.

"EBF-Cloud-Banking-Forum_The-Use-of-Cloud-Computing-by-Financial-Institutions.Pdf." n.d. Accessed May 3, 2021. https://www.ebf.eu/wp-content/uploads/2020/06/EBF-Cloud-Banking-Forum_The-use-of-cloud-computing-by-financial-institutions.pdf.

"Ermetic Reports Nearly 80% of Companies Experienced a Cloud Data Breach in Past 18 Months." 2020. June 3, 2020. https://www.businesswire.com/news/home/20200603005175/en/Ermetic-Reports-Nearly-80-of-Companies-Experienced-a-Cloud-Data-Breach-in-Past-18-Months.

Finnhub.io. 2021. "Finnhub – Free Realtime APIs for Stock, Forex and Cryptocurrency." February 19, 2021. https://finnhub.io/.

"GTB_Big_Data_Whitepaper_(DB0324)_v2.Pdf." n.d. Accessed May 3, 2021. https://cib.db.com/docs_new/GTB_Big_Data_Whitepaper_(DB0324)_v2.pdf.

Hale, Galina, and Jose A. Lopez. 2019. "Monitoring Banking System Connectedness with Big Data." *Journal of Econometrics*, Big Data in Dynamic Predictive Econometric Modeling, 212 (1): 203–20. https://doi.org/10.1016/j.jeconom.2019.04.027.

Hallikainen, Heli, Emma Savimäki, and Tommi Laukkanen. 2020. „Fostering B2B Sales with Customer Big Data Analytics." *Industrial Marketing Management* 86 (April): 90–98. https://doi.org/10.1016/j.indmarman.2019.12.005.

Hasan, Md. Morshadul, József Popp, and Judit Oláh. 2020. „Current Landscape and Influence of Big Data on Finance." *Journal of Big Data* 7 (1): 21. https://doi.org/10.1186/s40537-020-00291-z.

Kenton, Will. 2021. "What Is a Stock Quote?" Investopedia. February 17, 2021. https://www.investopedia.com/terms/s/stockquote.asp.

Lambiente, Filippo. 2019. "Cloudera End-To-Eed IOT Open Architecture." 2019. https://www.cloudera.com/content/dam/www/marketing/emea/pdfs/cldr-deloitte-2018/D1T2_IOT_PRESENTATION.pdf.

Mallon, Sean. 2021. "The Massive Impact of Big Data in the Stock Market." *SmartData Collective* (blog). February 3, 2021. https://www.smartdatacollective.com/big-data-analytics-has-potential-to-massively-disrupt-stock-market/.

Oberst, Thomas P. 2015. "Applications in Finance for Big Data," 18.

"Open for Innovation." n.d. KNIME. Accessed May 27, 2021. https://www.knime.com/open-for-innovation-0.

Pappas, Ilias O., Patrick Mikalef, Michail N. Giannakos, John Krogstie, and George Lekakos. 2018. "Big Data and Business Analytics Ecosystems: Paving the Way towards Digital Transformation and Sustainable Societies." *Information Systems and E-Business Management* 16 (3): 479–91. https://doi.org/10.1007/s10257-018-0377-z.

Project Pro. 2021. "Hadoop in Financial Sector." ProjectPro. 2021. https://www.dezyre.com/article/hadoop-in-financial-sector/80.

"Quandl." n.d. Accessed May 15, 2021. https://www.quandl.com.

Rajkumar Buyya, Satish Narayana Srirama, ed. 2019. *Fog and Edge Computing: Principles and Paradigms | Wiley*. Wiley.

Razin, Ely. n.d. "Big Buzz About Big Data: 5 Ways Big Data Is Changing Finance." Forbes. Accessed April 30, 2021. https://www.forbes.com/sites/elyrazin/2015/12/03/big-buzz-about-big-data-5-ways-big-data-is-changing-finance/.

Shen, Dehua, and Shu-heng Chen. 2018. "Big Data Finance and Financial Markets." In , 235–48. https://doi.org/10.1007/978-3-319-95465-3_12.

Tsaneva, Monika. 2019. "Implementation of GDPR Requirements in the Information Systems of Consumer Financing Companies." *Serdica Journal of Computing*, no. 3–4: 155–70. http://serdica-comp.math.bas.bg/index.php/serdicajcomputing/article/view/346.

Wang Qian. 2021. "Is Virtual Currency a Standard Financial Bubble? Blockchain Industry Research." In *2020 International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems*, 603–9. International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems. Springer.

Xue, Yunlan, Lingyu Xu, Jie Yu, and Gaowei Zhang. 2017. "A Hot Event Influence Scope Assessment Method in Cyber-Physical Space For Big Data Application." *Intelligent Automation & Soft Computing* 24 (April): 1–9. https://doi.org/10.1080/10798587.2016.1267247.

Zhang, Peng, Xiang Shi, and Samee Khan. 2017. *Can Quantitative Finance Benefit from IoT?* https://doi.org/10.1145/3132479.3132491.