CONF-IRM 2021 Proceedings

International Conference on Information Resources Management (CONF-IRM)

Summer 2021

Covid-19 Emergency Aid: How the Brazilian Government used Social Big Data Analytics to give economic support and protect vulnerable citizens

Rafael Ronqui

Alexandre Silva

Rodrigo Silva

Tiago Lara

Thiago Oliveira

See next page for additional authors

Authors
Rafael Ronqui, Alexandre Silva, Rodrigo Silva, Tiago Lara, Thiago Oliveira, Carlos Brandão, Willia Helouani, and Eduardo Francisco

Covid-19 Emergency Aid: How the Brazilian Government used Social Big Data Analytics to give economic support and protect vulnerable citizens

Rafael Ronqui
FGV EAESP
rafaelmronqui@gmail.com

Thiago Oliveira
FGV EAESP
thisco@gmail.com

Alexandre Silva
FGV EAESP
alexandre.lbs@gmail.com

Carlos Brandão
FGV EAESP
brandao@intelliway.com.br

Rodrigo Silva
FGV EAESP
rodrigorubens@newroad.com.br

William Helouani
FGV EAESP
williamboscardini@gmail.com

Tiago Lara
FGV EAESP
tiago.a.lara@gmail.com

Eduardo Francisco
FGV EAESP
eduardo.francisco@fgv.br

Abstract

The digital transformation has been accountable for major socio-cultural and economic changes, requiring different management solutions from governments and corporations. Data have a fundamental role due to their contribution to the decision-making process. An unexpected accelerator of these changes was a virus that paralyzed Brazil and the world, generating social isolation and freezing the economy while all the attention turned on how to contain and mitigate it. The social and economic impacts, especially on the low-income Brazilian population, were immediate. In May 2020, Dataprev (Social Security Technology and Information Company)'s core capabilities were used to put in place in 14 days the Emergency Aid program, fundamental to prevent the impacts from being even greater for the Brazilian population. In this study, we present the role and relevance of using technology and big data analytics (BDA) as the basis for the implementation of the largest aid program ever developed in the country, in which the Government invested R$ 265 billion (50 billion USD) and benefited more than 65 million people.

Keywords: Big Data Analytics, Data Lake, Social Impact, Brazil, Covid-19

1. Introduction

As new technologies emerge and Internet access becomes widely available, the amount of generated data grows at an exponential ratio. In such a connected world, handling a high volume of data efficiently and turn it into a value is a major challenge for companies and governments. The Global StorageSphere report by International Data Corporation (IDC) appoints that the global data storage capacity has been doubling every four years and suggests that 2020 landing will be 6.8 zettabytes, 16.6% growth over 2019 (Reinsel et al., 2017).

While the availability of these data represents an unprecedented opportunity for businesses, processing and interpreting them is a computational challenge considering that the widely assimilated relational database management systems (DBMS) were not designed for such concurrent record reading and writing competition. The term Big Data emerges to describe these huge data sets that, in contrast to traditional data sets, include masses of unstructured data and real-time analysis that generate opportunities for discoveries (Chen et al., 2014).

In the last decade, several systems were created to enable Big Data, providing horizontal scalability for simple reading and writing activities in databases distributed by many servers. These new-generation data storage systems are known by the acronym NoSQL, with no consensus between the definitions "Not Only SQL" and "Not Relational" (Cattell, 2010). A key feature that makes NoSQL systems stand out is the shared-nothing horizontal scale, which replicates and fragments data across multiple servers, allowing several read and write operations simultaneously.

The use of NoSQL and BDA databases was fundamental for the Brazilian Government through the state company Dataprev to manage and integrate a wide range of data for the distribution of the so-called "Emergency Aid", a financial help provided during the Covid-19 pandemic targeting the population without income sources or whose income was impacted due to the sanitary restrictions. This initiative allowed the Brazilian Government to technically specify the business rules, integrate databases, develop a mobile app, and release the payment for the affected population in 14 days, benefiting more than 65 million people. This initiative was the largest operation of resources transferring for the socially vulnerable population in Brazilian history (Cidadania, 2020).

2. Context

Due to the new Covid-19 pandemic, global poverty is expected to have the highest growth since 1990 (Sumner et al., 2020). According to World Bank Group, it is estimated that millions of people migrated to extreme poverty living on less than U$$ 1.90 a day due to the pandemic crisis (World Bank Group, 2020). Brazil was one of the most impacted countries by Covid-19, causing the death of 194.949 people by the end of 2020 (Cidadania, 2020).

With an estimated population of 211 million people, Brazil is the largest country in Latin America and the fifth most populous country in the world with an average per capita income of US$ 5.386,04 per year (IBGE, 2021). In addition to the humanitarian issue, the pandemic has accentuated other country issues such as political, health, social, and economic ones, paralyzing several important sectors such as tourism and manufacturing industries. The unemployment rate grew dramatically in the country during the beginning of the pandemic, reaching 14,6% of the active population (IBGE, 2021).

As the pandemic numbers started growing in the country, most states adopted lockdown initiatives to preserve the population's life by avoiding a bottleneck in the hospitals. As a collateral effect, the domestic GDP fell 5% in a month (BCB, 2020) pushing the Brazilian Government to adopt measures to support the population in filing their most basic needs. One of the main actions the Brazilian Government took to minimize the situation of social vulnerability and protect the purchasing power of the most affected population was through law No. 13,982 enacted on April 2, 2020, which changed the benefit of continued provision (BPC) by including the Emergency Aid (EA) modality. The main objective of the EA was to offer the population financial support to face the pandemic crisis, including specifically the unemployed people, the informal population, individual microentrepreneurs (MEI), and the self-employed citizens. The amount offered by the Brazilian Government was R$ 600.00 (113 USD) per person in the target group or R$ 1,200.00 (226 USD) in the case of families where the woman was the main and only responsible for the expenses (República, 2020). Benefited citizens received 5 full installments of aid, followed by 4 more installments of half the initial amount.

After the Ministry of Citizenship defined the public policy around the Emergency Aid, the distribution of benefits was left to two state companies: Dataprev (Social Security Technology and Information Company), in charge of processing the requests and check if the citizen is allowed to receive the benefit; and Caixa Econômica Federal, accountable for paying the approved amount.

Dataprev is a Brazilian public company, founded in 1974, and is responsible for providing technical information and communication solutions for the implementation of public policies in the Brazilian State. The state-owned company is a benchmark for quality in the processing and treatment of large volumes of data, storing and managing the National Social Information Register (CNIS), which enables the granting of social rights to the Brazilian population automatically, including pensions and maternity wages. As a result, the company processes around 8% of the Brazilian GDP per year (Dataprev, 2020).

Caixa Econômica Federal (CEF) is a Brazilian banking institution founded in 1861 and has around 105 million customers, the largest public bank in Latin America. It is currently responsible for the administration and payment of the social and labor benefits granted by the Brazilian Government. CEF also maintains several cultural institutions across the country and is responsible for the realization and administration of the federal lottery (CEF, 2021).

2.1 Technological Challenge

After law No. 13,982 was enacted, the Brazilian government faced the challenge of developing a database and system architecture, processing the emergency aid mobile application for millions of Brazilians in the shortest time possible. This technology should be designed to consider data processing on a large scale, allowing consultation with multiple databases from different Brazilian institutions, covering billions of records and with low processing time. As requirements, the system to be developed should be highly scalable, available, resilient, and offer high performance, in addition to implementing mechanisms that minimize the possibilities of fraud to grant benefits to the population.

In this context, the design of the technological solution included a complex analysis of scenarios, the definition of rules, and the adoption of appropriate technologies for big data and analytics. The system architecture team defined a database unification approach in an internal server environment, due to the solid infrastructure of the DataPrev data centers.

To validate the resulting data from the performed processing and analysis, the approval process was carried out by the Ministry of Citizenship, which used statistical techniques to validate the samples and analyze compliance. Such measures allowed the system to be progressively improved, correcting deficiencies and benefiting the transparency of the system before the effective data transmission process to CEF.

Another technologically challenging scenario was to enable tracking of the benefit request by the citizen, displaying details of the request, as well as the reason for granting or denying the aid according to the rules established by the Ministry of Economy. The targeted platform had to support millions of simultaneous accesses of citizens who were looking for information about their aid on specific dates. For this purpose, a web platform for consultation was developed, including a result contestation tool for citizen's manifestation when he disagrees with the denial decision.

3. Literature Review

3.1 Data Structuring Models

Data and information are essential parts of a collection of facts and figures stored in a specific format. Data can be organized in several different ways, such as logical or mathematical models that are termed as a data structure. These data can be structured, semi-structured, and unstructured.

Structured data can be understood as static data that can be easily organized (Agrawal & Patel, 2016) in a relational format (Amin et al., 2018), easy to organize, and in a predefined format, which can be managed by technologies that allow consulting, analyzing and managing data and understanding their relationships (Weglarz, 2004).

Semi-structured data do not need a definition schemewhile having a flexible structure (Agrawal & Patel, 2016). Semi-structured are characterized by not having a clear or implicit pattern definition and can be composed of heterogeneous elements, resulting in an irregular structure without any predefined data model (Lin et al., 2018).

Unstructured data refers to the fact that no identifiable structure within this type of data is available and cannot be classified, researched, viewed, or analyzed in the same way as structured data (Sint et al., 2009). Unstructured data is chaotic raw material generated by companies, social networks, IoT devices, geolocation, text, among others.

3.2 NoSQL Databases

As technological development and the massive progress of the Internet take place, an enormous new amount of information from many different resources and services are made available to humanity. This extraordinary mass of data is produced by and about people, things, and their interactions. Thus, this large volume of information needs to be addressed to become useful (Fahad et al., 2014).

The relational database was the first computational way of organizing data proposed by IBM researchers Donal Chamberlin and Raymondy, originally called SEQUEL (Structured English QUEry Language) and later renamed SQL for legal reasons. Given the limitations of SQL databases for processing semi-structured and unstructured data, new solutions were developed based on the non-relational database approach, better known as NoSQL (Not Only SQL or Not Relational SQL). Relevant IT companies, such as Google, Amazon, Facebook, Alibaba, and IBM, are dealing with a huge amount of data and using NoSQL databases as a solution, instead of traditional relational database systems, which can't efficiently handle big data (Ahmed et al., 2018).

It is noteworthy that such systems do not apply exclusively to a specific technology, but several different ones that apply to specific big data situations (Čerešňák & Kvet, 2019). Unlike relational databases, data repetition is acceptable (Vyawahare et al., 2017). NoSQL databases can handle large, complex, varied, and constantly changing data sets, being appropriate to Big Data applications (Gupta et al., 2018).

Features	Relational Database SQL	No relational Database NoSQL	Source
Data Structure	Structured – RDBMS	No Structured	Ishwarappa & Anuradha (2015).
Data Storage	Tables	Key values, Documents, Columns, and charts	Padhy et al. (2011), , Gudivada et al. (2014).
Data Normalization	Tables Should be normalized	No normalization is required	Khasawneh et al., (2020).
Data Organization	Predefined layouts and scheme	Dynamics schemes	Gyorödi et al. (2015), Saxena & Sachdeva (2018).
Physical memory dependency	The amount of data stored depends mainly on the physical memory of the system	The amount of data stored does not depend on the Physical system memory. It can be scaled horizontally according to the requirement	Sharma et al. (2015), Amin et al. (2018), Fahd et al. (2019).
Processing Time	Processing time depends on server machine configuration	Processing time depends on the number of machines in the cluster	Padhy et al. (2011), Fahd et al. (2019).

Query Language	SQL	Proprietary	Sharma et al. (2015), Gupta et al. (2018), Khasawneh et al. (2020).
Processing	May effectively control millions and billions of records	May effectively control millions and billions of records	Gudivada et al. (2014), Khasawneh et al. (2020).
Scalability	Vertical	Horizontal	Gyorödi et al. (2015), Čerešňák & Kvet (2019).
Standard	ACID - Atomicity, Consistency, Isolation, and Durability	CAP - Consistency, Availability and Partition tolerance	Ahmed et al. (2018), Khasawneh et al. (2020).

Table 1: SQL and NoSQL comparison
Source: As shown in the last column of the table

3.3 Relationship Among NoSQL, Big Data, and Big Data Analytics

In the so-called data age, the term Big Data has gained importance in the academic and corporate world. Big Data is used to describe the amount of data beyond the capacity of traditional data processing technology and applications to efficiently process, store, manage, distribute and process them (Francisco, 2017) .

Big Data can be recognized through five dimensions known as Big Data 5Vs (McAfee & Brynjolfsson, 2012). It includes Volume, which refers to the large amount of data generated from the Internet, social media, IoT equipment, among others; Variety, associated with different data sources and types, including structured, semi-structured, and unstructured data; Velocity, referring to the speed at which data is generated and with the ability to process asynchronously and in real-time (Zikopoulos & Eaton, 2011); Veracity, in terms of trust in the information collected and processed that can be used for decision making (Gandomi & Haider, 2015); and Value, referring to the value of the information that will be explored and derived in favor of organizations.

The potential value of Big Data is unlocked when leveraged to drive decision making, using processes and technologies to prepare and retrieve data for analysis. Analytics refers to techniques used to analyze and acquire intelligence from big data. Thus, BDA can be viewed as a sub-process in the overall process of ‘insight extraction’ from Big Data (Gandomi & Haider, 2015). Big Data and BDA have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from the sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization techniques (H. Chen et al., 2012).

Non-relational databases like NoSQL are appropriate to handle Big Data applications due to their ACID (Atomicity, Consistency, Isolation, Durability) properties (Ahmed et al., 2018). NoSQL technologies must also offer Consistency, Availability, and Partition Tolerance (CAP) to work effectively with Big Data (Brewer, 2000).

Based on this context and their capabilities, BDA has been used by governments extensively for public policymaking and reviewing purposes. Also, it has been used by companies and organizations to process social media data, analyze trends of massive diverse static and dynamic transactional data sets, address the clutter in data, perform statistical analysis, and store or search within both structured and unstructured data (Grover & Kar, 2017).

3.4 Data Lake

Data Lake is a highly scalable data repository containing data from one or more sources, stored natively, without manipulation, and in its raw state (Miloslavskaya & Tolstoy, 2016). Data Lakes reduce initial implementation efforts, as they ingest data in any format without requiring an initial scheme. They also facilitate the acquisition of new data, in addition to access to its original raw format (Llave, 2018). Data Lakes are typically built to handle large volumes of unstructured data that are generated at high speed, in contrast to Data Warehouses, which supports highly structured data. Data Lakes use dynamic analytical applications to generate insights, making processed data accessible as soon as they are created (Miloslavskaya & Tolstoy, 2016).

Khine & Wang (2018) highlight the importance of conducting case studies on Data Lake technologies and their applications in addition to business use cases, while there is still a large academic study field for greater and better use and exploration of that architecture.

3.5 Redis Database

Redis is a key-value NoSQL database that operates with in-memory mode, where a large part of the data is stored and processed in the memory, rather than in a file system and physical disks (Chopade & Pachghare, 2019). In Redis, as in other NoSQL databases, prior schema definition is not necessary. Designed as an open-source system, Redis implements a storage system through data structures that can be used as a database, cache, or message broker. Redis incorporates important features such as abstractions of many structures, replication, transaction support, persistence, automatic partitioning, atomic operations, asynchronous replication, and automatic failover (Redis.io, 2020).

Ingesting streaming data in real-time is a common requirement for many BDA use cases. In areas such as IoT, e-commerce, security, communications, entertainment, finance, and retail, which are highly dependent on timely and accurate decision-making processes, real-time data collection and analysis is indeed essential for business. Redis has become a popular choice for these scenarios, as it consists of a lightweight in-memory database platform, reaching millions of operations per second with latencies less than milliseconds while using minimal resources (Kumar, 2017).

4. Methodology

To analyze in-depth the impact of information technology in a social environment strongly impacted by the Coronavirus pandemic, this research opted for the qualitative method of participant observation. The data collection used in this modality had the participation of a researcher as an insider, which contributed to obtaining a detailed and sequential view of the observed events. Participant observation aims to produce a dense description of the research context in natural environments (Luiz Marietto & Sanches, 2013). For this matter, this research counted on the active participation of a member of the IT department from Dataprev, involved in the planning, implementation, and analysis of the Emergency Aid project.

In his seminal article, Gold (1958) determines the typology of the Participating Observer based on the intention and degree of involvement that the observer adopts during the research. This research adopts the degree of Full Participant, when the researcher, a routine member of the group has his scientific interest unknown to the group and keeps his interest in disguise. It should be noted that the publication of the research results on this article is approved by the institution's board of directors.

5. Emergency Aid and Big Data Analytics

The worldwide pandemic crisis demanded from governments the implementation of systems with the capacity to handle BDA in a very short timeframe, aiming at storing, processing, and consulting huge volumes of unstructured data coming from different sources at great speed, to help the decision-making of several departments and sectors. Data Lakes address such requirements, as they offer developers and data scientists repositories that enable broader access to raw data. Data Lakes can be used as a preparation environment for Data Warehouses, as experimentation platforms for scientists and data analysts, and finally, as a direct source for Business Intelligence in the self-service modality (Llave, 2018).

The Emergency Aid data lake combined 23 different databases from public administration that Dataprev already had access to. Those data included the National Register of Social Information (CNIS) and other sources from the Federal Public Administration, States, Municipalities, and Judiciary. A complete overview of the designed solution is displayed on the model from Figure 1. The figure displays the complete data source list used on the Emergency Aid data lake. After the initial processing, the analytical rules defined by the Ministry of Citizenship were applied and validated, with the results sent for the aid payment directly in the citizens' digital accounts.

The results of the processing needed to be presented to citizens transparently, informing the reason for granting or denying the payment. If a citizen believes that the result is not correct, a contestation form could be used to inform any mistaken applied business rule, to update their personal information, or to identify any fraud in the declaration process. Figure 1 summarizes all the data flow from the Emergency Aid data lake, including qualification, integration, and data security processes.

A multidisciplinary team was formed to deliver the project, composed of architects, developers, DBAS, data administrators, testers, leaders, and business analysts. The Emergency Aid application architecture was designed to enable simultaneous data consumption at both relational and No-SQL databases. This feature grants the solution data integrity, better response time, and flexibility.

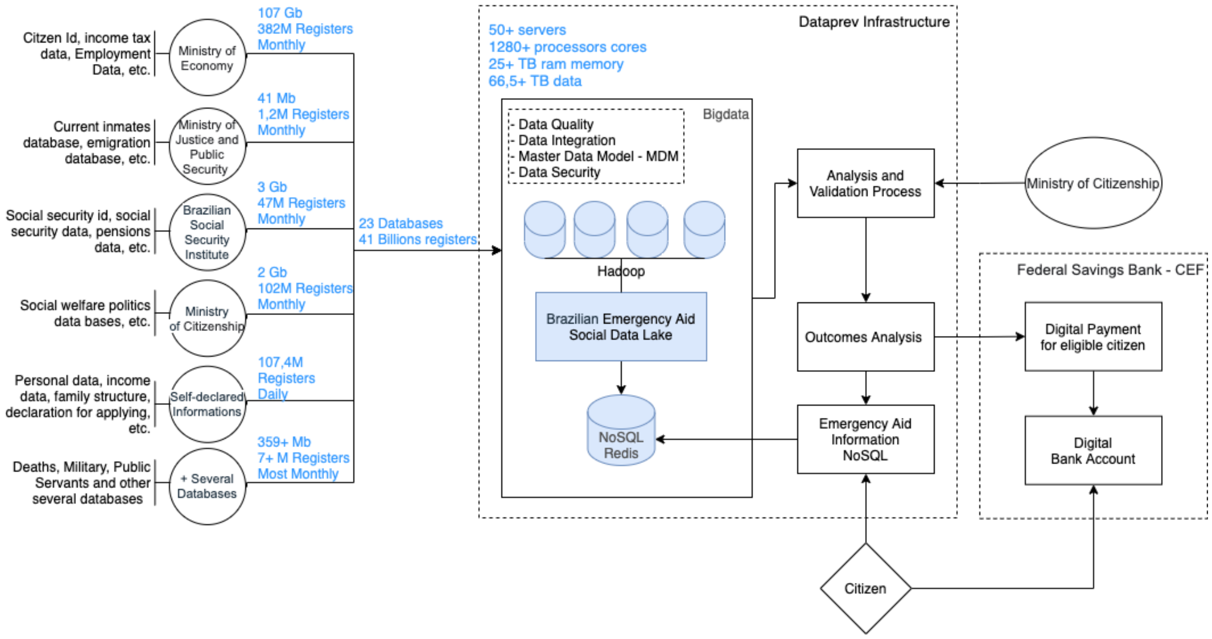

Figure 1: Conceptual Model of Dataprev's Emergency Aid solution
(Source: the authors)

As the requirements for Emergency Aid included the access and processing in almost real-time of an enormous data flow generated in high volume and speed, a non-relational database with a key-value type such as the Redis database revealed to be the most appropriate option for Dataprev's solution architects. Its features and architecture enable the creation of scalable databases, ranging from a simple prototype to systems with hundreds of gigabytes of data and millions of requests per second (Carlson, 2020).

To guarantee the number of simultaneous accesses to the Emergency Aid consultation application, Dataprev architects structured a responsive frontend based on React that serves for new and future data handling applications based on Redis technology with NoSql memory database. Additionally, the solution backend relied on Java Rest using Oracle Weblogic application servers.
Loads to the Redis database were made as soon as the update of the main relational database was changed by the benefit analysis processes. The application launch on the night of May 3rd, 2020 achieved 500,000

simultaneous accesses on the platform without any media advertisements. In less than 5 days, 21 million hits had been made with response times of less than 1 second. The overall infrastructure had more than 50 virtual servers, with its capacity tested in a pre-production environment with more than 1,000 views per second (Govbr, 2020). During the application go-live, a monitoring room was set up with panels based on Elasticsearch and Prometheus technologies, which enabled the application's telemetry around quantitative data access, including response time and errors.

The alternative of building and operating a Data Lake in a Redis database environment to support the Emergency Aid system was a highly coherent and feasible choice, as it would meet the requirements for implementing a high performance and capacity system, to be implemented in a very short space of time, unifying several databases (new and existing), dealing with a gigantic mass of structured and unstructured data, submitted to thousands of simultaneous requests for access and coming from the population of a country with continental dimensions.

6. Outcomes

Emergency Aid had a massive impact on protecting the most vulnerable citizens during the Covid-19 pandemic in 2020. According to the National Household Sample Survey (PNAD), the benefit was requested by 125 mi Brazilian citizens and approved for 67,8 mi, representing 32,13% of the country's population reached (Casa Civil, 2020). This reach was higher than other countries in Latin America, such as Argentina which covered 20% of its population with social protection (Blofield et al., 2021).

The total investment from the Brazilian Government on Emergency Aid exceeded 54,4 billion USD (Casa Civil, 2020). Poverty and inequality in the country have been reduced due to the impact of the EA, but after a reduction in the public policy payments, the average income is returning to the same levels before the pandemic (Duque, 2020). According to a study carried out by the Brazilian government, the income transfer programs managed to contain extreme poverty by 80% in 2020. If there is currently an estimated 2.1% of the population in a situation of extreme poverty, without the programs that index would be 12.4% of the population (Govbr, 2020).

7. Final Remarks

This study reveals how the evolution and use of NoSQL and BDA technologies were essential to meet the public measures taken by the Brazilian government to mitigate the social harms caused by the Covid-19 pandemic. Developed in only 14 days and first applied in May 2020, the Emergency Aid platform prevented more than 20 million Brazilians from returning to extreme poverty during 2020. Such application enabled the largest financial resources transferring for the most vulnerable population in Brazilian history.

The quick solution presented by Dataprev using Big Data technologies highlights the benefits of an effective and neutral IT Governance in the public sphere for the execution of public policies at the appropriate time. Moreover, the Emergency Aid case revealed the value that the NoSQL and BDA can provide to social programs. The Brazilian government's response to the social crisis suggests a trend of incorporating technology to foster and enable social policies around the world.

Although the Emergency Aid initiative was successful, it revealed some deficiencies regarding the data quality, resulting in fraud occurrences by people who should not receive the benefit according to interdisciplinary bodies such as the Federal Police and the Brazilian Justice. Moreover, a large amount of data in decentralized databases hindered the speed of updating and validating the request for certain groups of citizens.

The Brazilian Emergency Aid case proves that assimilated Big Data technology enables efficiency and agility for governments by solving complex needs. Data availability is a valuable asset for implementing public policies by identifying patterns, optimizing processes, and enabling decision-making.

Several questions around Big Data Analytics and public policies are yet to be answered by further research, including how to measure the effectiveness of income distribution; how can governments use NoSQL and BDA capabilities and technologies to prevent frauds and corruption in social services; how BDA can enable the creation of new smart public services; and how BDA can help governments to forecast the impact of new public policies.

References

Agrawal, S., & Patel, A. (2016). A Study on Graph Storage Database of NOSQL. International Journal on Soft Computing, Artificial Intelligence and Applications, 5(1). https://doi.org/10.5121/ijscai.2016.5104

Ahmed, M. R., Khatun, M. A., Ali, M. A., & Sundaraj, K. (2018). A literature review on NoSQL database for big data processing. International Journal of Engineering and Technology(UAE), 7(2), 902–906. https://doi.org/10.14419/ijet.v7i2.12113

Amin, A., Muttoo, M. A., & Bhatia, K. (2018). NOSQL: A Very Dynamic Approach for Managing Big Data. International Journal of Trend in Scientific Research and Development, Volume-2(Issue-4), 71–73. https://doi.org/10.31142/ijtsrd12962

BCB. (2020). Banco Central do Brasil. Retrieved 04 january de 2021, from - https://www.bcb.gov.br/publicacoes/focus.

Blofield, M., Lustig, N., & Trasberg, M. (2021). Social Protection During the Pandemic: Argentina, Brazil, Colombia, and Mexico. https://www.cgdev.org/sites/default/files/Note-Blofield_Lustig_Trasberg-Revised_0.pdf

Brewer, E. A. (2000). Towards robust distributed systems (abstract). https://doi.org/10.1145/343477.343502

Carlson, J. (2020). Redis In Action e-book. Manning.

Casa Civil. (2020). Auxílio Emergencial chega a 29 milhões de domicílios em outubro. https://www.gov.br/casacivil/pt-br/assuntos/noticias/2020/dezembro/auxilio-emergencial-chega-a-29-milhoes-de-domicilios-em-outubro-segundo-pnad-covid-19

Cattell, R. (2010). Scalable SQL and NoSQL data stores. SIGMOD Record, 39(4), 12–27. https://doi.org/10.1145/1978915.1978919

CEF. (2021). Caixa Econômica Federal. Retrieved 04 january 2021, from https://www.caixa.gov.br/Downloads/caixa-demonstrativo-financeiro/Apresentacao_de_Resultados_3T20.pdf.

Čerešňák, R., & Kvet, M. (2019). Comparison of query performance in relational a non-relation databases. Transportation Research Procedia, 40, 170–177. https://doi.org/10.1016/j.trpro.2019.07.027

Chen, H., Chiang, R. H. L., Storey, V. C., & Robinson, J. M. (2012). Special Issue: Business Intelligence Research Business Intelligence and Analytics: From Big Data To Big Impact. MIS Quarterly , 36(4), 1165–1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chopade, R., & Pachghare, V. K. (2019). Ten years of critical review on database forensics research. Digital Investigation, 29, 180–197. https://doi.org/10.1016/j.diin.2019.04.001

Cidadania, M. da. (2020). Ministério da Cidadania do Brasil. Retrieved 04 january 2021, from https://www.gov.br/cidadania/pt-br.

Dataprev. (2020). Empresa de Tecnologia e Informações da Previdência Social. Retrieved 04 january 2021, from https://portal2.dataprev.gov.br/conheca-dataprev-quem-somos/empresa.

Duque, D. (2020). Pobreza e Desigualdade aumentam de novo PNAD Covid mostrou impacto do Auxilio Emergencial nos indicadores sociais. Instituto Brasileiro de Economia. https://blogdoibre.fgv.br/posts/pobreza-e-desigualdade-aumentam-de-novo-pnad-covid-mostrou-impacto-do-auxilio-emergencial-nos

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Foufou, S., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE Transactions on Emerging Topics in Computing, 2(3), 267–279. https://doi.org/10.1109/TETC.2014.2330519

Fahd, K., Venkatraman, S., & Khan Hammeed, F. (2019). A Comparative Study of NOSQL System Vulnerabilities with Big Data. International Journal of Managing Information Technology, 11(4), 1–19. https://doi.org/10.5121/ijmit.2019.11401

Francisco, E. R. (2017). Indicações Bibliográficas - Big data analytics e ciência de dados: Pesquisa e tomada de decisão. Revista de Administração de Empresas.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gold, R. L. (1958). Roles in sociological field observations. Social Forces. https://doi.org/10.2307/2573808

Govbr. (2020). Governo do Brasil - Dataprev e Ministério da Cidadania lançam portal de consulta à situação do auxílio emergencial. Retrieved 12 january 2021, from https://www.gov.br/cidadania/pt-br/noticias-e-conteudos/desenvolvimento-social/noticias-desenvolvimento-soc.

Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. Global Journal of Flexible Systems Management, 18(3), 203–229. https://doi.org/10.1007/s40171-017-0159-3

Gudivada, V. N., Rao, D., & Raghavan, V. V. (2014). NoSQL Systems for Big Data Management. 190–197. https://doi.org/10.1109/services.2014.42

Gupta, A., Tyagi, S., Panwar, N., Sachdeva, S., & Saxena, U. (2018). NoSQL databases: Critical analysis and comparison. 2017 International Conference on Computing and Communication Technologies for Smart Nation, IC3TSN 2017, 2017-Octob(October 2017), 293–299. https://doi.org/10.1109/IC3TSN.2017.8284494

Gyorödi, C., Gyorödi, R., & Sotoc, R. (2015). A Comparative Study of Relational and Non-Relational Database Models in a Web- Based Application. International Journal of Advanced Computer Science and Applications, 6(11), 78–83. https://doi.org/10.14569/ijacsa.2015.061111

IBGE. (2021). Instituto Brasileiro de Geográfia e Estática (2021). Retrieved 04 january 2021, from https://www.ibge.gov.br/indicadores.

Ishwarappa, & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and hadoop technology. Procedia Computer Science, 48(C), 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Khasawneh, T. N., Al-Sahlee, M. H., & Safia, A. A. (2020). SQL, NewSQL, and NOSQL Databases: A Comparative Survey. 2020 11th International Conference on Information and Communication Systems, ICICS 2020, April, 13–21. https://doi.org/10.1109/ICICS49469.2020.239513

Khine, P. P., & Wang, Z. S. (2018). Data lake: a new ideology in big data era. ITM Web of Conferences, 17, 03025. https://doi.org/10.1051/itmconf/20181703025

Kumar, R. (2017). How to use Redis for real-time stream processing Handling fast data ingest in Redis. InfoWorld.

Lin, Y., Jun, Z., Hongyan, M., Zhongwei, Z., & Zhanfang, F. (2018). A method of extracting the semi-structured data implication rules. Procedia Computer Science, 131, 706–716. https://doi.org/10.1016/j.procs.2018.04.315

Llave, M. R. (2018). Data lakes in business intelligence: Reporting from the trenches. Procedia Computer Science, 138, 516–524. https://doi.org/10.1016/j.procs.2018.10.071

Luiz Marietto, M., & Sanches, C. (2013). Strategy as Practice: A study of the practices of strategic action in the SMEs store cluster. INTERNATIONAL JOURNAL OF MANAGEMENT & INFORMATION TECHNOLOGY. https://doi.org/10.24297/ijmit.v4i1.4633

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90(10), 4.

Miloslavskaya, N., & Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. Procedia Computer Science, 88, 300–305. https://doi.org/10.1016/j.procs.2016.07.439

Padhy, R. P., Patra, M. R., & Satapathy, S. C. (2011). RDBMS to NoSQL: Reviewing Some Next-Generation Non-Relational Database's. International Journal of Advanced Engineering Sciences and Technologies, 11(11), 15–30.

Redis.io. (2020). Introduction to Redis. Internet, 2020.

Reinsel, D., Gantz, J., & Rydning, J. (2017). Data Age 2025 : Don 't Focus on Big Data; Focus on the Data That's Big. IDC White Paper.

República, P. da. (2020). Planalto da República - Retrieved 04 january 2021, from http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l13982.htm.

Saxena, U., & Sachdeva, S. (2018). An insightful view on security and performance of NoSQL databases. Communications in Computer and Information Science, 799(March 2018), 643–653. https://doi.org/10.1007/978-981-10-8527-7_54

Sharma, S., Tim, U. S., Gadia, S., Wong, J., Shandilya, R., & Peddoju, S. K. (2015). Classification and comparison of NoSQL big data models. International Journal of Big Data Intelligence, 2(3), 201. https://doi.org/10.1504/ijbdi.2015.070602

Sint, R., Schaffert, S., Stroka, S., & Ferstl, R. (2009). Combining unstructured, fully structured and semi-structured information in semantic wikis. CEUR Workshop Proceedings, 464, 73–87.

Sumner, A., Hoy, C., & Ortiz-Juarez, E. (2020). Estimates of the impact of COVID-19 on global poverty. Unuwider, April, 1–9.

Weglarz, G. (2004). Two Worlds of Data Unstructured and Structured. DM Review.

World Bank Group. (2019). World Bank Country and Lending Groups. In World Bank Group.

Vyawahare, H. R., Karde, P. . D., & Thakare, V. . D. (2017). Brief Review on SQL and NoSQL. International Journal of Trend in Scientific Research and Development (IJSTRD), 2(5), 748–751.

Zikopoulos, P., & Eaton, C. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Education.