Data Quality in a Decentralized Environment

Alexander Bogdanov[1], Alexander Degtyarev[1,3], Nadezhda Shchegoleva[1(✉)], and Valery Khvatov[2]

[1] Saint Petersburg State University, 7–9, Universitetskaya Emb., St. Petersburg 199034, Russia
{a.v.bogdanov,n.shchegoleva}@spbu.ru, deg@csa.ru
[2] DGT Technologies AG, 15 Wertheim Ct, Richmond Hill, ON L4B 3H7, Canada
valery.khvatov@gmail.com
[3] Plekhanov Russian University of Economics, 36, Stremyanny per, Moscow 117997, Russia

Abstract. One of the most important aspects of applying distributed ledger technologies in the field of big data is ensuring the necessary data quality and calculating the corresponding metrics. The paper proposes a conceptual framework for working with Master Data in a decentralized environment. The greatest effect of this framework methods is increasing a real-time integrity for the data segments that have the direct impact on overall data quality. The proposed approach provides the result thanks to a special platform architecture similar to the blockchain and built-in artificial intelligence agents - oracles.

Keywords: Data quality · Distributed ledger technologies · Master data management · F-BFT consensus

1 Introduction

1.1 Motivation

Modern organizations grow their business by forming groups of companies, alliances and by building relationships with subcontractors and partners through the processes of horizontal and vertical integration. Integration allows for the creation of a unified information space within one organization or even along the entire value chain. The effectiveness of integration and business continuity depends on the alignment of core directories, as well as the matching of configurations and metainformation between connected organizations.

Imprecise data is one of the quintessential problems for organizing the consolidation and interaction between various data sources. Ensuring high data precision is especially crucial for the digital economy, since data correctness is the key to making the right and timely business decisions.

A study done by the Gartner Group in 2011 [1] shows that on average organizations lose at least $9.7M per year due to bad data quality, operational ineffectiveness and lost opportunities. According to another study [2], only 12% of organizations use their data's potential to attain strategic advantages.

O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12251, pp. 58–71, 2020.
https://doi.org/10.1007/978-3-030-58808-3_6

The situation is made worse by the influence of Big Data and by the largely decentralized nature of modern business. However, modern technologies can help overcome the problem of data quality. According to [3], the use of decentralized registry and artificial intelligence technologies will elevate data quality by 50% by 2023. It is true that the inherent qualities of blockchains render them capable of controlling consistency and integrity, which are some of the most important data quality attributes.

However, selecting such solutions in an arbitrary way may lower data accessibility and present issues wish scalability, particularly as a result of inherent limits in organizing storage. The ideas presented below provide a conceptual framework for managing data quality through the integration of master-data using the DGT platform [4], which combines distributed ledger technologies powered by the F-BFT consensus with Machine Learning advancements.

1.2 Data Quality Criteria

Data Quality (DQ) is a characteristic that shows the degree to which data is useful. DQ is an integral characteristic of data, which is presented through its attributes that cover different aspects of data usage. According to the ISO 9000:2015 standards, the key criteria for data quality are: (1) Relevance, (2) Accuracy, (3) Integrity, (4) Completeness, (5) Consistency and (6) Timeliness. In practical cases, these parameters carry different weights for different organizations, so the end model of data quality is heavily dependent on the domain (Fig. 1).

Fig. 1. The main DQ Attributes by ISO 9000:2015.

Quality attributes are not limited to these core criteria and could feature additional components depending on the specifics of the processing system. Source [5] presents valid quality criteria and their widespread synonyms. The total quality evaluation may be attained through the average of attribute scores within the accepted quality model for a given system.

In terms of decentralized systems, Data Integrity is the most crucial attribute, which we will examine in detail. It is the attribute that is most sensitive to diversions in the object model, terminology and differences in variability of separate measurements.

1.3 Quality Challenges for Big Data

Data of any nature and size is subject to distortions, many of which are connected with missing values, duplicates, contradictions, anomalous values and lack of inclusivity. These problems may be seen from different points of view: violations of structure, format, random or intentional substitution of values, loss of associations and etc.

Big data significantly increases the probability of the occurrence of such errors due to the influence of its characteristics (size, speed, diversity, accuracy and cost), which reflect its distributed nature:

- Unified information frequently includes several data types (structured, half-structured, unstructured), which makes reconciliations difficult even among data that carries the same meanings;
- Semantic differences in definitions may lead to the same positions being filled differently. For example, the term "revenue" may mean different things for various organizations, which means that the tabulated report will lose meaning and lead to incorrect conclusions;
- Differences in formats and syntaxes (for example, time marks) will also lead to internal conflicts and contradictions between subsets of values.

Managing data quality in such conditions makes it impossible to use one metric across the entire set. Moreover, the well-known Brewer's CAP theorem [6] presents an additional limitation: the impossibly of ensuring simultaneous integrity of the complete dataset for a number of distributed systems (in particular for some NoSQL databases, which support only eventual consistency).

Two crucial trends map the direction in which the big data quality framework is growing: decentralization and data virtualization. The first trend illustrates the necessity of adapting distributed ledger technologies to control data quality, while the second shows the necessity of abandoning verification according to a given data structure (since it may vary).

A constructive approach to calculated data quality metrics is the selection of stable information objects and applying validation rules according to them in real-time. In other words, it is proposed to use a two-phase data processing approach:

- Pre-processing of incoming data with the identification of main information objects and validation of their attributes;
- Processing of quality attributes across the entirety of available data, taking into account the discrepancies in versions of transactional information.

Such an approach will prevent "data depletion" as a result of the upload and will adapt a standard mechanism of calculating quality in the decentralized conditions of data processing.

This approach is practically equivalent to building a reliable system for managing master-data in the Big Data paradigm and storing fast transactional information in the maximum number of versions possible.

1.4 Master Data and Identification

The work of any IT-system or a combination of IT-systems comes down to processing, storing and transferring data. Data represents repeatedly interpreted formalized information. Depending on the specifics of the objects reflected in data, it is customary to identify data classes that will indicate the data structure and the objects belonging to a certain type of information.

In terms of any corporate information system that participates in such an informational exchange, we can identify the following information types:

- Transactional (or operational) data – a rapid stream that describes the changes in statuses of information objects, such as money transfers, product shipments, sensor indicators;
- Analytic data – slices of operational data prepared for decision-making;
- Master Data – which is necessary for identifying information objects; these are sets of data with a relatively slow rate of change, including normative-directive information, metadata, parameters and configuration of informational objects.

According to the classic definition stated by Gartner [8], Master Data is defined as "… the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts."

Therefore, Master Data is a conditionally constant set of data that defines the composition of the domain being automated and the basis for describing the business logic of the application. Master Data can have a flat, hierarchal or network structure depending on the existing business processes.

Master Data conditionally includes such subgroups as directors, metadata and configurations, depending on the existing models of information management and object life cycles.

Let's take a look at Master Data in a decentralized context. Suppose, several organizations with a common list of counterparties are cooperating with one another. Such a list (directory) is the Master Data and within the framework of a single organization it would be recorded following the approach commonly known as the "golden record". However, synchronizing this list of counterparties for several organizations may be difficult:

- Names of counterparties often come as part of complex unstructured information (such as an invoice) and needs to be selected out;
- Such names may have geographic, language and stylistic differences, which complicate the selection of similar objects from previous entries.

Because of that, the process of evaluating quality (in particular, Integrity) requires an identification procedure that will:

- use probabilistic and statistical methods (in particular, NLP – Natural Language Processing methods);
- rely on a large spectrum of external sources, such as a KYC (Know Your Customer) system or an Open Technical Dictionary (OTD);
- include the process of distributed harmonization and standardization – distributing the information about the decisions made in regards to creating a new object or a link made to an existing object.

Therefore, Master Data has a direct influence on the quality of information and in the context of distributed decentralized systems, it requires a process of agreement/consensus between nodes/hosts that would represent the different sides of the informational exchange.

2 Conceptual DGT Quality Framework

2.1 The Approach

The main strategy is based on the separation of Master Data processing into a separate type of data processing for a distributed environment. Master Data is one of the most important information assets that a modern organization has. With the continuous digitization of processes, the creation of digital twins, and the advent of the fourth industrial revolution, the importance of master data and its management will only grow.

In essence, Master Data refers to all of the static information that is used to identify critical elements of an organization and its business processes. Assigning incoming operational information to objects requires identification, as does the formation of consistent datasets for analytics.

Thus, Master Data, transactional data, and analytic data are interdependent and part of the same context. Errors and discrepancies in Master Data can cause the same or even greater damage as the discrepancies in transaction data. For example, an error in identifying a client during billing may have a much larger impact than an erroneous calculation of the invoice amount for a correct customer.

Master Data maintain the consistency of a common information array between different information systems, divisions and organizations. The most important characteristic of Master Data is the slow rate of change in the informational exchange between several participants. When working with Master Data, the following management styles can be identified:

- Transaction-based. In this style, Master Data is a part of the transmitted transactions; the data is selected directly from the transactions and the management of such Master Data is not a task shared by all of the participants in the information exchange;
- Centralized Master Data. This style features a single system for all participants, which keeps master-data in a common storage. Other systems use it as a reference. Due to a large number of reasons, such a solution is often limited in use by small systems and a small number of participants in the information exchange;

- Shared Master Data. In this style, the management of Master Data is separated in its own distinct stream that is composed either of a series of ETL procedures, which transfer Master Data from one participant to the next, or from an information exchange over distributed ledger technologies.

In a decentralized environment, big data is composed of a large number of independent data sources, additional data streams generated by digital objects, significant flexibility in settings and an overlap of various life cycles. Therefore, the information exchange has additional properties that need to be taken into account:

- The limitations of centralized solutions. Due to the presence of a single point of failure, centralized architectures have a low degree of adaptivity and are subject to risks in managing change processes. They are susceptible to the influence of subjective decisions and cannot support the increasing level of complexity that requires distributed data processing;
- Access to data in real time. Supporting continuity of cooperation between different systems requires an asynchronous access to data – a capability possessed by decentralized registries;
- Smart data processing. Data processes require not only the right integration, but the use of machine learning – artificial intelligence (AI) functions for comparison of complex datasets. New technologies allow for the calculation of the degree of similarity for different datasets, as well as data quality metrics, and other metrics that allow for the improvement of information models;
- Storage of change history. In times of very dense information steams, it is necessary to track not only the changes to parameters, but also connections between these objects. To serve these objectives, there need to be storage systems of a particular type, which would consider hierarchy, correlation, and other specifics (graph databases).

In the framework of the approach being discussed, these problems are solved by utilizing innovational technologies that support great speed of decision-making and reduce losses due to data mismatch:

- The integration layer of the system is built on a high-performance DGT core, which ensures the formation of a unified Master Data registry and its distribution between the participants of an information exchange with a large degree of horizontal scalability and the ability to track the entire history;
- Smart modules (oracles) that track data in real-time and participate in building reconciled datasets while simultaneously measuring quality metrics;
- Developed API that can plug into not only the different corporate systems and analytic instruments, but also to a variety of instruments of data management and profiling.

2.2 Distributed Ledger Technologies Layer

Distributed ledger technologies store shared data in a shared database (registry, ledger) replicated multiple times between several nodes. The rules of inserting data, changing information, and the specifics of registry replication are governed by an algorithm that is commonly called a consensus.

In the DLT approach, the distribution and support of the registry (Master Data registry) is conducted using a network of nodes that are the agents of the distributed MDM structure. A consensus determines which agents may insert data and under which rules. The sum of these rules constitutes the process of validation that prevents entry duplication, while the signature and cryptographic security requirements additionally guarantee the immutability of the registry (its integrity).

Not every consensus will allow to synchronize distributed Master Data. For example, the well-known PoW (Proof-of-Work) that operates in the Bitcoin environment through excessive calculations is not well-suited for validating big data.

The approach below describes the model for working with big data based on the F-BFT consensus [10]. The model's features include:

- Data processing is done in a hybrid consortium-based network built on a federative principle: nodes are grouped in clusters with changing leaders and network access is limited by a set of conditions;
- Registry entry is done as a result of "voting" in a cluster and the subsequent "approvals" of an arbitrator node. Both "voting" and "approval" are a series of checks-validations in the form of calculations with binary results;
- Each network node receives information and identifies informational objects as one of the Master Data classes;
- If an object is new, then there is an attempt to initiate a specialized transaction to insert data into the corresponding registry through a voting mechanism of intermediary nodes (validation process). If the new object is approved by other nodes, the object is added to the registry and the information is disseminated in the network. If a new object is denied, the initializing node receives a link to an existing object;
- The distributed data storage system (registry) takes the form of a graph database (DAG, Directed Acyclic Graph) that allows for coexistence of several transaction families for different object classes, while maintaining the network's horizontal scalability.

This approach allows for the separation of fast streaming data from batch processing of the MDM distributed ledger. Along the way, the quality attributes are calculated in terms of information integrity, as a measure of conflict for a given object. This technology is advantageous in cleaning data in real-time without limiting data availability.

2.3 The Artificial Intelligence Layer

The difficult problem with data quality control is the low efficiency of manual checks with increasing volume and variability of data. In such cases, machine learning modules can help assess quality early in processing, diagnose data absence problems, availability of unforeseen data types, non-standard parameter values, contradictions between different sets, etc.

The use of artificial intelligence (AI) allows for the resolution of several important tasks:

- Clearing text data using Natural Language Processing (NLP) technologies and extract MD from loosely structured texts. NLP modules can determine the degree of correspondence between objects based on context;
- Ensuring compliance against set standards and Master Data management practices; conversion of MD into standard form;
- High-speed comparison of datasets (Entity Resolution) based on closeness metrics (most relevant for configurations);
- Measuring data quality directly based on support vector machine (SVM) algorithm.

AI is based on one of several neural networks that are educated through a labeled model – a prepared dataset. A more advanced approach could be used as well, that being a Generative Adversarial Network (GAN) where one network creates learning templates and the other builds recognition.

Within this article, we will take note of the most in-demand techniques that directly influence the quality of big data and the measurement of quality attributes:

- Advanced technique for information objects discovery and identification;
- Data pattern recognition;
- Prediction analysis;
- Anomaly detection.

In big data, some of the information comes in the form of unstructured information, containing references to people, businesses, geographical locations. ML modules can extract and store such information automatically by fixing the objects themselves and the connections between them. Another problem that is solved automatically is the tracking of duplicate records. Tracking such records can be done using a random forest algorithm. This algorithm not only simplifies frequency analysis, but also allows you to build predictive models, forming the expectations of incoming data flow.

Even a small data entry error can have a significant impact on analytical tools, on the usefulness of the data. Finding anomalies based on machine learning will improve the quality of data, automatically correct found errors, eliminate conflicts of formats, process the inclusion of foreign data.

Algorithms like SVM (support vector machine) allow to categorize texts, images, complex sets of scientific data, search by faces, voices, graphic data.

The AI mechanism is embedded into existing verification mechanisms and works in a distributed approach by using smart agents (oracles) that are directly involved in data validation.

2.4 Data Quality Calculation

The approach formulated above allows for the quantification of data quality in real time with the following assumptions:

- The general quality assessment is conducted on the basis of a weighted indicator, estimated as the number of operations necessary to correct the identified error;

- Some errors may be identified at the data validation stage (ex. differences in object naming that can be fixed, incorrect spelling, etc.), while others only during subsequent analysis. Therefore, the quality attribute for a given dataset is constantly recalculated;
- The attribute's weight depends on the current value of source reliability (in the framework being described that also impacts the number of checks – "votes"), as well as the seriousness and priority of error according to the relevant validation rule;
- Object identification based on fuzzy logic and neural network results;
- Identification of anomalies and correlation with earlier data. When comparing information mechanisms, rules are cartridges (smart-contracts) that are inseparable from the registry and may calculated in accordance with Lowenstein distance, as an example.
- Quality attributes that are incalculable or impossible to evaluate are noted for subsequent analysis.

Algorithm 1. Data Quality Evaluation for Quality Attribute (Integrity)

```
procedure DQE_Attribute
  Input try_Object ← Array ();
  for try_Object in Objects () do
        prove_Object ← IdentificationProcedure(try_Object)
        DQ_Assement ← Initialize ()
        for each Rule in Attribute_Validation_Rules do
            DQ_Assement.Add GetValidResult (Rule, prove_Object)
        end for
        DQ_Assesments.Add DQ_Assement
  end for
  return DQ_Assesments
end procedure
```

The connection with the validation rules is in full accordance with the consensus model – each selected object and its attributes go through checks in the validation rule stack, gradually receiving all of the necessary evaluations. Even the data that is fully rejected may be saved in a separate registry and participate in later evaluations, thus impacting the total quality value.

The identification and validation of quality attributes must be done in accordance with time characteristics – the period given for validation. Checking and rechecking attributes in accordance to time periods is called the alignment of quality processes in the given framework. Such an approach allows for a connection with the life cycle of an information object and for tracking its changes. The general approach to align processes is presented in the Fig. 2 below.

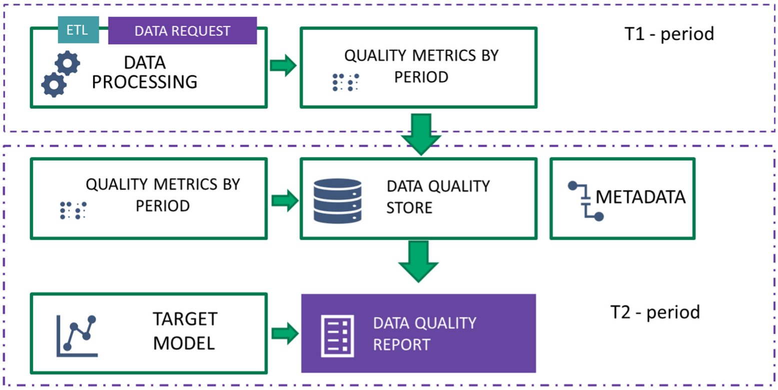

Fig. 2. Quality process aligning.

The total quality ratio can be calculated as weighted average by the following indicators:

- Number of unidentified (unidentified) objects that have been recovered in the future;
- Data inaccessibility statistics based on frequency of requests;
- Processing data-gathering conflicts, including anomalies and going beyond data validation ranges;
- Distance between the initial and final data vectors;
- Coincidence with results from other sources;
- Timeline lengths and data latency;
- Estimates of cleaning time relative to the overall download cycle

Checks are based on historical data directly in validators' rules when "voting" nodes for data insertion.

2.5 Implementation

We used the DGT platform to implement this MDM/DQ platform. Specially adapted transaction families provided MDM information for the sharing ledger on multiple nodes in a cluster of related organizations. The basic architecture of the framework is shown in the Fig. 3 below.

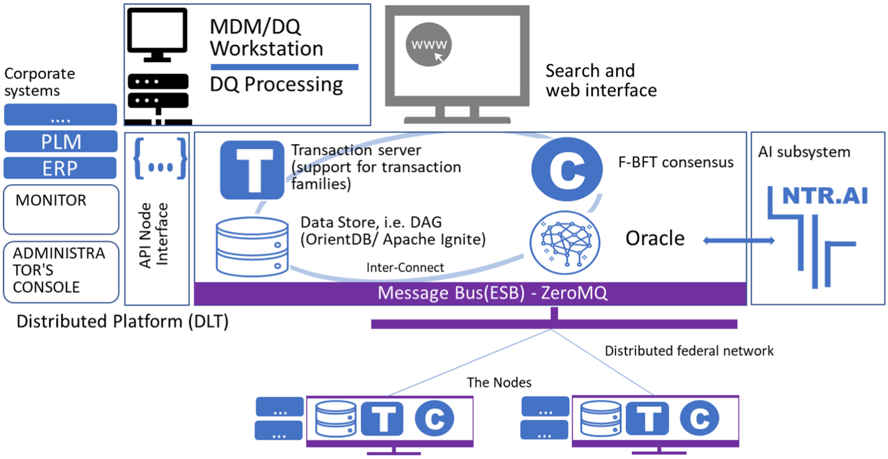

Fig. 3. Framework implementation.

3 Practical Cases

Applying this framework gives special advantages for vertical and horizontal integration of independent economic objects. Below, we describe relevant use scenarios where distributed technologies coupled with artificial intelligence provides the maximum benefit.

3.1 Case A. Data Quality in Aviation

An exchange of data in transport and logistic companies was never a simple issue. In particular, airline companies are subjected to significant expenses from flight cancellations, delays, ground handling overhead, and the search for balance between safety and profitability.

Complex data processing streams combine with the necessity to make decisions in real time, as required by strict industry regulation and by an extensive ecosystem of air travel. This ecosystem includes the airline itself, clients (individual/corporate/freight), airports, partner airlines global distribution systems (GDS), industry associations (IATA), government bodies, suppliers and service suppliers, as well as other participants of the value chain (hotels, car rental companies, tour agents, ad networks).

Even small discrepancies in Master Data may lead to millions of dollars in losses for airlines. Special circumstances affecting the data synchronization methodology is the presence of significant contradictions between the participants of the ecosystem, as well as a large number of groups and alliances between the larger ecosystem, which forms a multi-level federative network where the consensus may take form of intersecting clusters. This will allow for the construction of multi-level consensuses and branched MD registries available in real time.

Perhaps the most pressing problem that unites the interests of the largest players is optimizing ground operations and ensuring timely departures. Directory data and metainformation that coordinates the data streams of on-board crews and support staff (cleaning services, baggage taggers, other ground services) involved in the stages of the flight will optimize expenses and present necessary data for analysis, identify any source of a potential delay, and enable quality decision-making.

A recent study [11] shows that there are currently no universal instruments of managing data exchange in airlines, while existing instruments (ERP, SRM, TMS, WMS) do not allow for full transparency between various participants. Difficulties arise when checking the correctness of paperwork, labeling, and authenticity of information about the origin of transported goods.

The use of the suggested framework will allow to save time on checking paperwork, customs control, and phytosanitary control, as well as enable tracking of goods, simplify the identification of counterfeit products, and improve the quality control of logistics operations.

3.2 Case B. IoT

The Industry 4.0 and greater use of Internet of Things promise significant improvements for managing the environment and industry objects, formulating predictive analytics, improving the quality of equipment and safety levels of industrial facilities.

The large sum of data generated by various sensors, combined with the low cost of sensors, have led to an entire class of IoT platforms that focus on collecting data on the direct physical level. This partially alleviates the issue of heterogeneity of the IoT environment and narrows interactions to a few common platforms/hubs [12].

However, in addition to ensuring data accessibility, there is the issue of overall data integrity between various sub-systems, organizations, and entire industries. Examples of such cluster interactions include:

- Interaction of medical systems, especially concerning sensitive data between different organizations, hospitals, device manufacturers;
- An exchange of data between energy producers, distributive systems, and users;
- Data received in food supply chains, with the necessary integration of the manufacturer's devices with those of logistics companies and food selling organizations.

Significant research, such as [13], has been devoted to the necessity of forming multi-level architecture for distributed devices, where the IoT layer would be supplemented by a DLT layer. A federative network that is capable of processing several transaction families would be highly useful, considering the broad class of devices and their geographic dispersion. In this case, different devices, protocols, and other configurating parameters would be synchronized in the total MDM stream, ensuring the necessary level of data quality (Fig. 4).

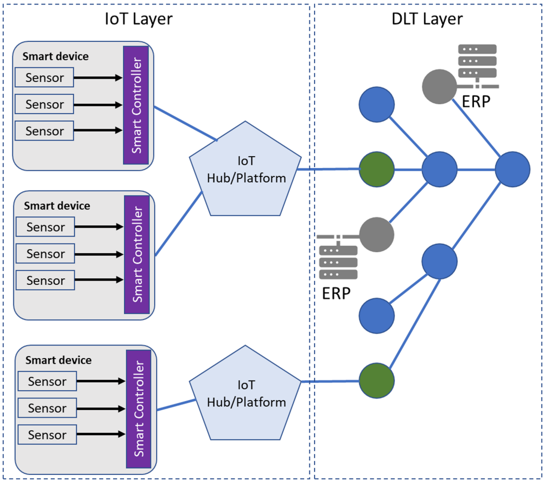

Fig. 4. Two-tier IoT/DLT network.

3.3 Case C. Public Health Care Data

The latest COVID-19 events have shown significant dependence of measures taken by governments to mitigate critical situations on high-quality datasets. The Blockchain Research Institute (BRI) has published an extensive report [14] where it identified five key directions for this technology to be used in fighting the pandemic and its consequences. One of such directions is forming a registry of combined data and statistics for:

- Storing data according to categories, illness statuses;
- Tracking parameters of the measures taken, available hospital beds, ventilators, masks;
- Tracking product deficits and etc.

As everyone witnessed, despite the best efforts of Johns Hopkins University the death rate statistics were quite malleable due to differences in counting methodologies, the consumer market displayed a low level of adaptivity across several goods, while the medical industry of many countries was unprepared for a pandemic. As was shown in the report, countries with good access to data (China, Singapore, Korea) were able to take effective measures in fighting the disease. Countries with limited data capabilities (such as Italy, Spain, the US) experienced significant difficulties.

The presence of a DLT framework as a key element of facilitating data exchange could have saved many lives. We should also consider the following parameters of this necessary solution:

- The presence of horizontal scalability as a mandatory element for working with critical data;
- Elevated security and cryptographic protection according to medical standards;
- Parallel processing of several transaction families to reflect the necessary data variability.

4 Conclusion and Outlook

The framework presented in this study allows us to formulate the requirements for systems necessary to maintain the quality of big data in terms of its integrity. The problems described above allow us to conclude that the distributed ledger and artificial intelligence technologies are in demand for monitoring and assuring quality of big data systems and in a broader sense of the multitude of dynamic systems used in the modern digital world.

Applying distributed ledger technologies to the task of maintaining master data between organizations will provide a united information space for groups of companies integrated horizontally or vertically, allow real-time quality indexes to be calculated and effective information exchange in operational data, improve the quality of analytical data, and ultimately make the decision-making process itself a quality.

References

1. Measuring the Business Value of Data Quality. Gartner Inc. (2011)
2. Assessing Your Customer Intelligence Quotient. Forrester. http://www.forrester.com/Assessing+Your+Customer+Intelligence+Quotient/fulltext/-/E-RES53622?docid=53622
3. Predicts 2020: Data and Analytics Strategies — Invest, Influence and Impact, Gartner Report (2019)
4. DGT Platform Overview. http://dgt.world/docs/DGT_About.pdf
5. List of Conformed Dimensions of Data Quality. CDDQ Open Standard. http://dimensionsofdataquality.com/alldimensions
6. Brewer, E.A.: Towards robust distributed systems. In: (Invited Talk) Principles of Distributed Computing, Portland, Oregon, July 2000
7. ISO/IEC/IEEE 24765-2010: Systems and software engineering – Vocabulary
8. Gartner: Master Data Management (MDM). https://www.gartner.com/en/information-technology/glossary/master-data-management-mdm
9. ECCMA Open Technical Dictionary. https://www.eotd.org/
10. Bogdanov, A., Uteshev, A., Khvatov, V.: Error detection in the decentralized voting protocol. In: Misra, S., Gervasi, O., Murgante, B., Stankova, E., Korkhov, V., Torre, C., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O., Tarantino, E. (eds.) ICCSA 2019. LNCS, vol. 11620, pp. 485–494. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24296-1_38
11. IATA: Blockchain in aviation. Exploring the fundamentals, use cases, and industry initiatives. White paper|October 2018. https://www.iata.org/contentassets/2d997082f3c84c7cba001f506edd2c2e/blockchain-in-aviation-white-paper.pdf
12. Fernández-Caramés, T.M., Fraga-Lamas, P.: A review on the use of blockchain for the Internet of Things. IEEE Access **6**, 32979–33001 (2018)
13. Casado-Vara, R., de la Prieta, F., Prieto, J., Corchado, J.M.: Blockchain framework for IoT data quality via edge computing. In: The 1st Workshop on Blockchain-enabled Networked Sensor Systems (BlockSys'18), November 4, 2018, Shenzhen, China, p. 6. ACM, New York, NY, USA, Article 39 (2018). https://doi.org/10.1145/3282278.3282282
14. Tapscott, D., Tapscott, A.: Blockchain Solution in Pandemic, BRI, April 2020