

УДК 004.63

DOI: 10.25559/SITITO.17.202104.889-895

Научная статья

Оценка временной эффективности форматов хранения больших данных в динамике роста объема данных

В. А. Белов, Е. В. Никульчев*

ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва, Российская Федерация

Адрес: 119454, Российская Федерация, г. Москва, Проспект Вернадского, д. 78

* nikulchev@mail.ru

Аннотация

При разработке озера данных на таких платформах, как Apache Hadoop, важным вопросом становится выбор формата хранения данных. Этот выбор должен опираться на ряд различных критериев, одним из которых являются временные затраты при запуске различных запросов к этим данным. Однако любая система обработки данных предполагает постоянный рост объема этих данных. В связи с этим возникает необходимость изучения эффективности форматов в динамике роста объема данных, хранящихся в системе. В данной статье предлагается методика оценки эффективности форматов хранения данных в озерах данных, построенных на платформе Apache Hadoop, в динамике роста объема данных. Предложен эксперимент, представляющий из себя ряд запусков запросов различной сложности к данным, хранящихся в форматах JSON, Apache Avro, ORC, Apache Parquet. Для запуска запросов использовался фреймворк Apache Spark.

Ключевые слова: большие данные, форматы хранения данных, Apache Hadoop, Apache Spark, оценка эффективности

Авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Белов В. А., Никульчев Е. В. Оценка временной эффективности форматов хранения больших данных в динамике роста объема данных // Современные информационные технологии и ИТ-образование. 2021. Т. 17, № 4. С. 889-895. doi: <https://doi.org/10.25559/SITITO.17.202104.889-895>

© Белов В. А., Никульчев Е. В., 2021



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Evaluation of the Temporal Efficiency of Big Data Storage Formats in the Dynamics of Data Growth

V. A. Belov, E. V. Nikulchev*

MIREA – Russian Technological University, Moscow, Russian Federation

Address: 78 Vernadsky Ave., Moscow 119454, Russian Federation

* nikulchev@mail.ru

Abstract

When developing a data lake on platforms such as Apache Hadoop, the choice of data storage format becomes an important issue. This choice should be based on a number of different criteria, one of which is the time it takes to run different queries on this data. However, any data processing system assumes a constant growth in the volume of this data. In this regard, it becomes necessary to study the effectiveness of formats in the dynamics of growth in the amount of data stored in the system. This article proposes a methodology for assessing the effectiveness of data storage formats in data lakes built on the Apache Hadoop platform in the dynamics of data growth. An experiment is proposed, which is a series of queries of varying complexity to data stored in JSON, Apache Avro, ORC, Apache Parquet formats. The Apache Spark framework was used to run queries.

Keywords: big data, data storage formats, Apache Hadoop, Apache Spark, efficiency evaluating

The authors declare no conflict of interest.

For citation: Belov V.A., Nikulchev E.V. Evaluation of the Temporal Efficiency of Big Data Storage Formats in the Dynamics of Data Growth. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2021; 17(4):889-895. doi: <https://doi.org/10.25559/SITITO.17.202104.889-895>



Введение

При разработке крупных система обработки слабоструктурированных данных, таких как озера данных, часто используют готовые платформы, обладающие минимально необходимым набором инструментов работы с данными. Одной из таких систем является платформа Apache Hadoop [1]. Данная платформа представляет из себя кластерное решение, состоящие из распределенной файловой системы HDFS [2] и набора утилит, предлагающих различный функционал для обработки данных. Хранение данных в данной системе предполагается не в реляционных хранилищах, а в файлах различных форматов. Наиболее известными и популярными форматами для хранения слабоструктурированных данных являются такие форматы, как JSON¹, Apache Avro², ORC³, Apache Parquet⁴. Данные форматы предлагают разные способы и внутреннюю структуру хранения данных.

Выбор формата при разработке озера данных становится важной задачей, которую необходимо решать комплексно, анализируя различные характеристики представленных форматов. В работах [7-11] представлены различные подходы к анализу и экспериментальной оценке форматов хранения данные как в озерах данных, так и в системах, не использующих готовые решения для разработки собственных систем обработки данных. В работе [12] представлен алгоритм выбора формата хранения данных, состоящий из экспериментальной оценки каждого формата на основе запусков запросов к данным с использованием фреймворка Apache Spark [13], и математической модели выбора формата на основе решения задачи многокритериальной оптимизации. Однако данные исследования не затрагивают вопрос изучения изменения эффективности формата при росте объема данных. Поскольку любая система обработки и хранения данных предполагает постоянный рост объема этих данных, данная характеристика имеет важное значение, поэтому следует рассматривать ее как один из критериев выбора формата.

В данном исследовании предлагается алгоритм построения вектора рейтинга альтернатив форматов хранения больших данных в озера данных, построенных на основе платформы Apache Hadoop, по критерию оценки динамики роста объема данных. В качестве альтернатив выступают форматы JSON, Apache Avro, ORC, Apache Parquet [3-6].

Структура статьи представлена следующим образом. В разделе «Методы» предлагается описание алгоритма построения рейтинга альтернатив. В данном разделе описаны данные, используемые для исследования, требования к проведению эксперимента, а также математические методы вычисления рейтинга альтернатив. В разделе «Эксперимент» представлено описание проведенного исследования. Здесь представлены данные об используемых инструментах для проведения

исследования, схема проведенного эксперимента, используемые запросы к данным и результаты замеров времени обработки данных. В разделе «Результаты» представлен результат вычисления рейтинга альтернатив, а также приведен анализ проведенного исследования.

Методы

Для оценки эффективности форматов хранения данных в динамике роста их объема была разработана следующая экспериментальная методика.

Для проведения эксперимента следует использовать несколько наборов данных разного объема. Необходимо подготовить наборы таким образом, чтобы разница между минимальным и максимальным объемом данных превышала десятикратный размер. В противном случае результаты эксперимента могут давать некорректные данные о ресурсной эффективности форматов хранения данных.

После подготовки данных необходимо подготовить ряд запросов к этим данным. При этом рекомендуется использовать запросы разного уровня сложности для получения более достоверных результатов эксперимента. Следует также проанализировать потребности пользователей разрабатываемой платформы для выяснения наиболее часто используемых запросов к данным.

Сам эксперимент заключается в запуске подготовленных запросов к данным разного объема с замером времени обработки этих данных. Следует учесть, что одного запуска недостаточно для точной оценки времени обработки данных. Следует повторить запуск несколько раз, чтобы избежать влияния различных сетевых задержек, оптимизаций в java-приложении и других факторов. В качестве результата следует взять среднее значение всех замеров для одного эксперимента.

Каждое из полученных значений временных характеристик следует использовать для расчета скорости изменения времени обработки файлов разных форматов. Скорость рассчитывается по следующей формуле:

$$rate_i = \frac{duration_i}{duration_1},$$

где $duration_i$ – результаты экспериментальной характеристики для i^{th} набора.

Скорость роста показывает, как изменяется временная эффективность того или иного формата при изменении объема данных. Критерием эффективности формата в данном случае будет выступать значение углового коэффициента регрессии для формата в рамках эксперимента с использованием одного запроса на разных объемах данных.

В результате проведения эксперимента будет получено несколько угловых коэффициентов регрессии для каждого формата. Предлагается использовать усредненное значение этих

¹ Введение в JSON [Электронный ресурс] // ECMA-40 The JSON Data Interchange Standard. Ecma International, 2021. URL: <https://www.json.org/json-ru.html> (дата обращения: 27.08.2021).

² Apache Avro [Электронный ресурс] // Apache Avro™ 1.11.0 Documentation. Apache Avro, 2021. URL: <https://avro.apache.org/docs/current> (дата обращения: 27.08.2021).

³ ORC File Format [Электронный ресурс] // Language Manual ORC. Atlassian, 2021. URL: <https://cwiki.apache.org/confluence/display/hive/languagemanual+orc> (дата обращения: 27.08.2021).

⁴ Apache Parquet [Электронный ресурс] // Apache Software Foundation, 2021. URL: <https://parquet.apache.org/documentation/latest> (дата обращения: 27.08.2021).



коэффициентов по всем экспериментальным замерам для каждого формата:

$$increase_i = \frac{\sum_{j=1}^n regression\ coefficient_{ij}}{n},$$

где $regression_coefficient_{ij}$ – угловой коэффициент регрессии $i^{ю}$ набора на $j^{ю}$ эксперименте, n – количество экспериментов.

Результатом расчета становится вектор значений усредненно-го углового коэффициента регрессии по всем экспериментальным запускам по всем форматам. Формат хранения данных будет считаться более эффективным, чем меньшее значение принимает вычисленный коэффициент.

Эксперимент

Для эксперимента были выбраны следующие форматы хранения данных: JSON, Apache Avro, ORC, Apache Parquet.

JavaScript Object Notation [3] – простой текстовый формат. Данный формат основан на базе языка JavaScript и представляет собой объект, состоящих из пар ключ-значение.

Apache Avro [4] – это линейно-ориентированным форматом хранения данных. Структура файла состоит из заголовка и блоков с данными. Кроме того, внутри файл данного формата содержит JSON-схему содержащихся данных, что облегчает интерпретацию данных при их обработке.

Optimized Row Columnar [5] – колоночный формат, специально разработанный для платформы Apache Hadoop. В отличие от других форматов ORC имеет внутренние индексы для быстрого доступа к данным.

Apache Parquet [21] – это бинарный колоночный формат. Parquet использует архитектуру, основанную на «уровнях определения» (definition levels) и «уровнях повторения» (repetition levels), что позволяет довольно эффективно кодировать данные, а информация о схеме выносится в отдельные метаданные.

Для проведения испытаний был подготовлен кластер Apache Hadoop с установленным Apache Spark. В таблице 1 представлена конфигурация созданного стенда.

Т а б л и ц а 1. Конфигурация экспериментального стенда

Table 1. Experimental stand configuration

Элемент конфигурации	Характеристика
Операционная система	Ubuntu 18.04 LTS
Дистрибутив Hadoop	Cloudera 5.7.0
Количество узлов Hadoop	3
Процессор	Intel Core i7 1.8 GHz 4 ядра
Оперативная память	8 GB
Количество узлов Spark	3
Платформа	Java Virtual Machine
Язык разработки	Java v. 1.8
Версия фреймворка	Apache Spark v. 2.4
Процессор	Intel Core i3 1.2 GHz 2 ядра
Оперативная память	4 GB

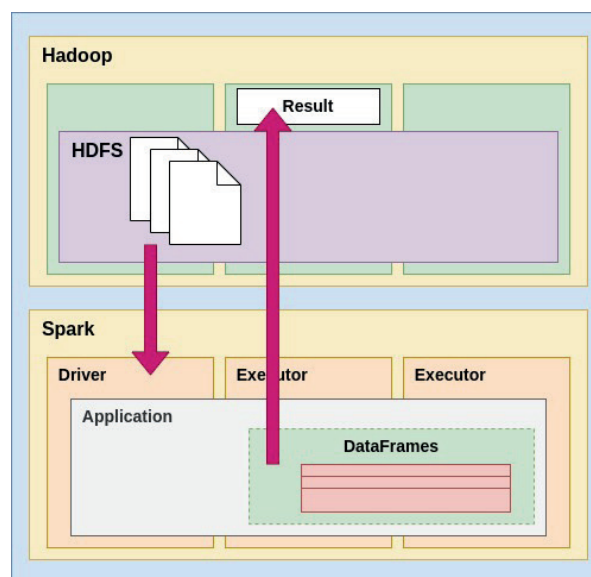
Для проведения исследования было сформировано 5 наборов данных, состоящих из 10^6 записей. Описание сформированных данных представлено в таблице 2.

Т а б л и ц а 2. Описание сгенерированных данных

Table 2. Description of generated data

Название поля	Тип данных
name	string
surname	string
age	integer
country	string
balance	long
card number	string
currency	string
account open date	calendar

На рисунке 1 представлена схема проведенного эксперимента. В HDFS кластера в разных папках располагаются сгенерированные наборы данных. После запуска задачи spark-submit разработанное Spark-приложение считывает переданные ему набор данных определенного формата. В качестве библиотеки для осуществления запросов к данным использовалась библиотека Spark DataFrames. Результат замера сохраняется в файлы логирования на файловой системе Apache Hadoop.



Р и с. 1. Схема эксперимента

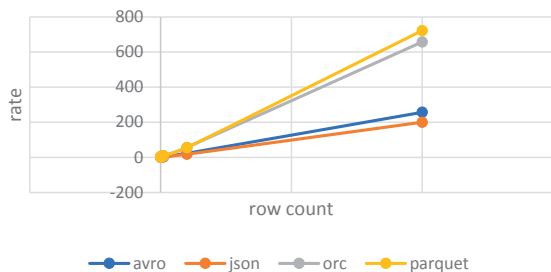
Fig. 1. Experiment scheme

Были подготовлены следующие запросы к данным (в скобках представлен аналогичный SQL-запрос):

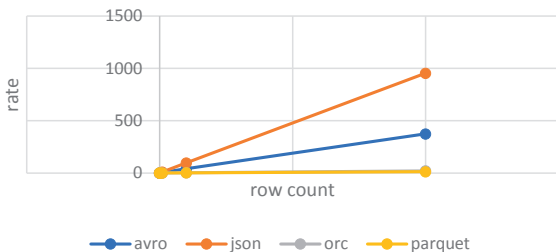
1. Поиск уникальных значений по всему набору данных;
2. Фильтрация данных по заданному условию;
3. Применение функции к каждой записи и возвращение новых значений;
4. Создание витрины данных (в запросе к данным использовались операции фильтрации и агрегации данных).



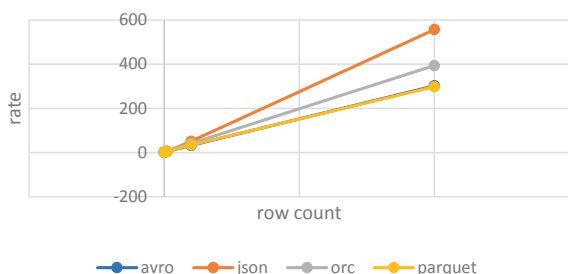
Для каждого сформированного набора данных были проведены замеры времени по каждому запросу к данным. После проведения экспериментальных запусков была вычислена скорость изменения ресурсной эффективности времени обработки данных. На рисунках 2-5 представлены графики роста времени обработки разных наборов данных.



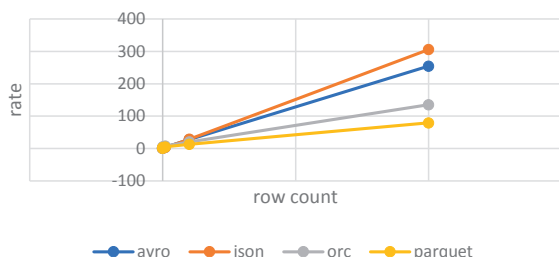
Р и с. 2. Рост времени для операции поиска уникальных значений
F i g. 2. Increase in the time for unique value search operation



Р и с. 3. Рост времени для операции фильтрации данных по условию
F i g. 3. Increase in the time for filter data by condition



Р и с. 4. Рост времени для операции применения функции к данным
F i g. 4 Increase in the time for the operation of applying a function to data



Р и с. 5. Рост времени для операции создания витрины данных
F i g. 5. Increase in the time for data mart creation operation

Результат

После сбора всех необходимых данных были вычислены усредненные угловые коэффициенты регрессии по всем запускам. В таблице 3 представлен результаты вычислений.

Т а б л и ц а 3. Усредненный угловой коэффициент по всем экспериментам

Table 3. Average angular coefficient for all experiments

Формат	Значение коэффициента
Avro	$19 \cdot 10^{-4}$
JSON	$46 \cdot 10^{-4}$
ORC	$13 \cdot 10^{-4}$
Parquet	$11 \cdot 10^{-4}$

Согласно полученным результатам, рейтинг форматов выглядит следующим образом:

$$parquet \leq orc < avro < json$$

Результат проведенного исследования не является универсальным для других задач, а лишь отражает целесообразность использования форматов в рассматриваемых условиях. Следует отметить, что одним из факторов, влияющих на результат эксперимента, являются типы данных, используемые в записях. Такие форматы, как Apache Parquet и ORC, обладающие метаданными, легче работают с числовыми значениями. При использовании строковых значений результаты исследования будут отличаться.

Данное исследование изучает время выборки и обработки данных в форматах, используемых для постоянного хранения данных. Представленное исследование не затрагивает вопросы потоковой передачи данных.

В рамках данного исследования не проводилось изучение затрат ресурсов процессора и оперативной памяти из-за особенностей работы фреймворка Apache Spark, связанных с тем, что данный фреймворк потребляет все выданные ему ресурсы. Увеличение ресурсов при увеличении объема данных не дает точной оценки форматов хранения данных. Исследование направлено на изучение изменения времени работы с данными в разных форматах в динамике роста при статичности выдаваемых ресурсов [14-25].

Заключение

В статье представлена методика экспериментальной оценки временной эффективности формата хранения данных в озерах данных, построенных на платформе Apache Hadoop, в динамике роста объема этих данных. Методика основана на ряде экспериментальных запусков запросов к данным с последующим вычислением вектора рейтинга альтернатив изучаемых форматов. В качестве критерия оценки используется усредненный угловой коэффициент регрессии, полученный по всем запросам к разным объемам данных. После вычисления данной характеристики можно оценить, время обработки какого формата изменяется меньше всего. Формат считается более эффективным, чем меньше значение принимает вычисленный усредненный угловой коэффициент регрессии.



Для проведения экспериментальной части исследования были использованы форматы JSON, Apache Avro, ORC и Apache Parquet, поддерживаемые системой Apache Hadoop и фреймворком Apache Spark, как одним из самых популярных фреймворков для обработки больших данных.

Результаты проведенного эксперимента не отражают преимущества одних форматов перед другими, а отображает целесообразность выбора формата в заданных условиях. При изменении различных факторов, таких как используемые типы данных, результаты исследования будут отличаться.

Разработанная методика оценки может быть полезна для

практического применения. При разработке больших систем обработки данных на основе платформы Apache Hadoop возникает вопрос выбора формата хранения данных. Важным критерием выбора является изменение затрат ресурсов при увеличении объемов данных. Данное исследование предлагает методику оценки этих изменений.

Кроме того, важно обратить внимание на то, что данное исследование не имело целью изучение функциональных особенностей форматов хранения данных. Основной целью исследования стало формирование алгоритма оценки эффективности форматов хранения данных в динамике роста объема данных.

References

- [1] Mavridis I., Karatza H. Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark. *Journal of Systems and Software*. 2017; 125(C):133-151. (In Eng.) doi: <https://doi.org/10.1016/j.jss.2016.11.037>
- [2] Lee S., Jo J.-Y., Kim Y. Survey of Data Locality in Apache Hadoop. *2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)*. IEEE Press, Honolulu, HI, USA; 2019. p. 46-53. (In Eng.) doi: <https://doi.org/10.1109/BCD.2019.8885148>
- [3] Bourhis P., Reutter J.L., Suárez F., Vrgoč D. JSON: Data model, Query languages and Schema specification. *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS '17)*. Association for Computing Machinery, New York, NY, USA; 2017. p. 123-135. (In Eng.) doi: <https://doi.org/10.1145/3034786.3056120>
- [4] Boufeua A., Finkers R., van Kaauwen M., Kramer M., Athanasiadis I.N. Managing Variant Calling Files the Big Data Way: Using HDFS and Apache Parquet. *Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT '17)*. Association for Computing Machinery, New York, NY, USA; 2017. p. 219-226. (In Eng.) doi: <https://doi.org/10.1145/3148055.3148060>
- [5] Gohil A., Shroff A., Garg A., Kumar S. A Compendious Research on Big Data File Formats. *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE Press, Madurai, India; 2022. p. 905-913. (In Eng.) doi: <https://doi.org/10.1109/ICICCS53718.2022.9788141>
- [6] Durner D., Leis V., Neumann T. JSON Tiles: Fast Analytics on Semi-Structured Data. *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*. Association for Computing Machinery, New York, NY, USA; 2021. p. 445-458. (In Eng.) doi: <https://doi.org/10.1145/3448016.3452809>
- [7] Ramírez A., Parejo J.A., Romero J.R., Segura S., Ruiz-Cortés A. Evolutionary composition of QoS-aware web services: A many-objective perspective. *Expert Systems with Applications*. 2017; 72:357-370. (In Eng.) doi: <https://doi.org/10.1016/j.eswa.2016.10.047>
- [8] Gholamshahi S., Hasheminejad S.M.H. Software component identification and selection: A research review. *Software: Practice and Experience*. 2019; 49(1):40-69. (In Eng.) doi: <https://doi.org/10.1002/spe.2656>
- [9] Munir R.F., Abelló A., Romero O., Thiele M., Lehner W. A cost-based storage format selector for materialized results in big data frameworks. *Distributed and Parallel Databases*. 2020; 38(2):335-364. (In Eng.) doi: <https://doi.org/10.1007/s10619-019-07271-0>
- [10] Wang X., Xie Z. The Case For Alternative Web Archival Formats To Expedite The Data-To-Insight Cycle. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20)*. Association for Computing Machinery, New York, NY, USA; 2020. p. 177-186. (In Eng.) doi: <https://doi.org/10.1145/3383583.3398542>
- [11] He D., Wu D., Huang R., Marchionini G., Hansen P., Cunningham S.J. ACM/IEEE Joint Conference on Digital Libraries 2020 in Wuhan virtually. *ACM SIGWEB Newsletter*. 2020; (1):1-7. (In Eng.) doi: <https://doi.org/10.1145/3427478.3427479>
- [12] Belov V.A., Nikulchev E.V. Experimental evaluation of the temporal efficiency of big data processing for specified storage formats. *International Journal of Open Information Technologies*. 2021; 9(9):95-102. Available at: <https://www.elibrary.ru/item.asp?id=46515796> (accessed 27.08.2021). (In Russ., abstract in Eng.)
- [13] Salloum S., Dautov R., Chen X., Peng P.X., Huang J.Z. Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*. 2016; 1(3):145-164. (In Eng.) doi: <https://doi.org/10.1007/s41060-016-0027-9>
- [14] Chong D., Shi H. Big data analytics: a literature review. *Journal of Management Analytics*. 2015; 2(3):175-201. (In Eng.) doi: <https://doi.org/10.1080/23270012.2015.1082449>
- [15] Moro Visconti R., Morea D. Big Data for the Sustainability of Healthcare Project Financing. *Sustainability*. 2019; 11(13):3748. (In Eng.) doi: <https://doi.org/10.3390/su11133748>
- [16] Cappa F., Oriani R., Peruffo E., McCarthy I.P. Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety and Veracity on Firm Performance. *Journal of Product Innovation Management*. 2021; 38(1):49-67. (In Eng.) doi: <https://doi.org/10.1111/jpim.12545>
- [17] Nazari E., Shahriari M.H., Tabesh H. Big Data Analysis in Healthcare: Apache Hadoop, Apache spark and Apache Flink. *Frontiers in Health Informatics*. 2019; 8(1):e14. (In Eng.) doi: <http://dx.doi.org/10.30699/fhi.v8i1.180>



- [18] Gusev A., Ilin D., Nikulchev E. The Dataset of the Experimental Evaluation of Software Components for Application Design Selection Directed by the Artificial Bee Colony Algorithm. *Data*. 2020; 5(3):59. (In Eng.) doi: <https://doi.org/10.3390/data5030059>
- [19] Belov V., Tatarintsev A., Nikulchev E. Choosing a Data Storage Format in the Apache Hadoop System Based on Experimental Evaluation Using Apache Spark. *Symmetry*. 2021; 13(2):195. (In Eng.) doi: <https://doi.org/10.3390/sym13020195>
- [20] Shahzad A., Usman Ali M., Ferzund J., Sarwar M.A., Rehman A., Mehmood A. Modern Data Formats for Big Bioinformatics Data Analytics. *International Journal of Advanced Computer Science and Applications*. 2017; 8(4):366-377. (In Eng.) doi: <http://dx.doi.org/10.14569/IJACSA.2017.080450>
- [21] Plase D., Niedrite L., Taranovs R. A comparison of HDFS compact data formats: Avro versus Parquet. *Mokslas – Lietuvos Ateitis = Science – Future of Lithuania*. 2017; 9(3):267-276. (In Eng.) doi: <https://doi.org/10.3846/mla.2017.1033>
- [22] Sakr S., Liu A., Fayoumi A.G. The family of MapReduce and large-scale data processing systems. *ACM Computing Surveys*. 2013; 46(1):11. (In Eng.) doi: <https://doi.org/10.1145/2522968.2522979>
- [23] Ene A., Im S., Moseley B. Fast clustering using MapReduce. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11)*. Association for Computing Machinery, New York, NY, USA; 2011. p. 681-689. (In Eng.) doi: <https://doi.org/10.1145/2020408.2020515>
- [24] Chang F., Dean J., Ghemawat S., Hsieh W.C., Wallach D.A., Burrows M., Chandra T., Fikes A., Gruber R.E. Bigtable: A Distributed Storage System for Structured Data. *ACM Transactions on Computer Systems*. 2008; 26(2):4. (In Eng.) doi: <https://doi.org/10.1145/1365815.1365816>
- [25] Assunção M.D., Calheiros R.N., Bianchi S., Netto M.A.S., Buyya R. Big Data computing and clouds. *Journal of Parallel and Distributed Computing*. 2015; 79(C):3-15. (In Eng.) doi: <https://doi.org/10.1016/j.jpdc.2014.08.003>

*Поступила 27.08.2021; одобрена после рецензирования 13.10.2021; принята к публикации 28.11.2021.
Submitted 27.08.2021; approved after reviewing 13.10.2021; accepted for publication 28.11.2021.*

Об авторах:

Белов Владимир Александрович, аспирант, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Российская Федерация, г. Москва, Проспект Вернадского, д. 78), ORCID: <https://orcid.org/0000-0001-8769-2529>, belov_va@mail.ru
Никутьчев Евгений Витальевич, профессор кафедры интеллектуальных систем информационной безопасности, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Российская Федерация, г. Москва, Проспект Вернадского, д. 78), доктор технических наук, профессор, ORCID: <https://orcid.org/0000-0003-1254-9132>, nikulchev@mail.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Vladimir A. Belov, Postgraduate Student, MIREA – Russian Technological University (78 Vernadsky Ave., Moscow 119454, Russian Federation), ORCID: <https://orcid.org/0000-0001-8769-2529>, belov_va@mail.ru
Evgeny V. Nikulchev, Professor of the Intelligent Cyber-Security System Department, MIREA – Russian Technological University (78 Vernadsky Ave., Moscow 119454, Russian Federation), Dr.Sci. (Tech.), Professor, ORCID: <https://orcid.org/0000-0003-1254-9132>, nikulchev@mail.ru

All authors have read and approved the final manuscript.

