BIG DATA TECHNOLOGIES: CONCEPT, PURPOSE, AND APPLICATIONS

A. O. Utepbergenov
Sain-Petersburg state university, department of Human resource management

Abstract: The purpose of the article is to identify and analyze opportunities to optimize human resource management in the selected companies using digital technologies, in particular to retain valuable employees and develop professional and other competencies, as well as to develop recommendations to similar companies to implement and develop BIG DATA technologies in talent management in today's economy.

Key words: digital age, high-impact work, MapReduce, Hadoop, HRM theorists

In the digital age, organizations need to embrace modern technology to remain successful. The application of big data using digital tools in the context of the need for constant renewal allows companies to be more efficient.

Today's challenges of globalization, the transition from industrial to networked digital economy, new speeds in business, mobility and transparency [1]. To ensure success in the face of rapid change in the digital age, business rules are changing - management models are competing [2] to ensure productivity, innovation, flexibility and adaptability. Benchmarking is becoming a prerequisite for competitiveness: by 2020, 75% of companies will be digital [3]; there is already an active cluster of companies doing business only online; scientists predict changes in the workplace over the next 20 years similar to those resulting from the new technological revolution; business and HR strategies are undergoing a digital transformation, the consequences of which are determined only by 20% of paper-based document management; companies that create products and services for the digital world are gaining momentum.

Big data technologies are not magical algorithms that will automatically transform HR departments. Deep dive into the subject area, analysis of the context, correct task definition are the main components of success.

The theoretical base on big data is small and the number of foreign sources prevails over Russian ones. The most up-to-date information on the issue in question can be found in articles and on websites. Of the works that deal with big data in the field of personnel management, it is possible to distinguish such domestic authors as D.Y. Zhukova and M.I. Pogrebnyak N.I. Nagibina A. Borisova V.G. Konovalova, R.A. Dolzhenko. Of the foreign authors we can distinguish: D. Laney, K. Lynch, G. Press, T. Tse, M. Esposito, M. West, M and K Edwards. Because of the lack of theoretical basis on this topic, a gap is created in this part. It is necessary to rely on current news in business, mainly Internet

The study will use both theoretical methods: the study of scientific literature - as well as case study methods: trend analysis in human resource management, comparative analysis, visualization approach for hiring and selecting employees and talent management.

HRM theorists have considered sets of HRM practices either individually or as systems (Combs, Liu, Hall, & Ketchen, 2006)[4], divided into two sections. The first, "HR Research and

Practice," brings together articles that discuss in general terms the use of BD in HRM. The second includes high-impact work practice systems (Huselid, 1995) [5] in which BD approaches have been applied. These are selection and hiring; evaluation and development; information, training, and knowledge; and strategy, efficiency, and effectiveness.

Recommendations on how to improve their use and how to collect the missing data will be presented, and the overall effectiveness of the implemented technologies will be calculated.

With the development of the digital environment, the role of information has changed significantly. If before the development of computer technology, the speed of information creation and distribution allowed us to designate it as a source of necessary information and knowledge, then in today's realities we can safely say that information is gradually acquiring the status of an independent resource, which has its own value [6].

Initially, the set of approaches and technologies included tools for mass-parallel processing of indefinitely structured data, such as NoSQL DBMS, MapReduce algorithms and Hadoop project tools.

MapReduce is a model of distributed parallel computing in computer clusters introduced by Google. According to this model the application is divided into a large number of identical elementary tasks performed on nodes of the cluster and then naturally summarized in the final result.

NoSQL (NotOnly SQL) is a generic term for various non-relational databases and repositories, it does not refer to a particular technology or product.

Hadoop is a freely distributed set of utilities, libraries and frameworks for developing and executing distributed programs running on clusters of hundreds or thousands of nodes. It is considered one of the fundamental technologies of big data.

After analyzing the performance of different characteristics (Table 1), we can conclude that even in a small organization Protected Employee Time Recording (PERS) is necessary for processing various types of communication of structural units, for the security of confidential personal data or, for example, financial statements. The concept of Big Data successfully combines the qualities of program management systems, adaptive management systems and intelligent management systems [7].

Table 1

Comparison table for Big Data (compiled by the author)

Characteristics	SQL	NoSQL	BDA	data (comparison characteristics)
Big Data analytics object-oriented	Yes	Different options are possible	No	Practical similarities characteristics of the studied programming languages without significant differences
Multimedia types Data	Yes	Yes	Not limited	Practical similarities characteristics of the studied programming languages

				without significant differences
Max table size	Yes Specify the size by the number	Yes Specify the size by the number	Practical not limited	Practical similarities characteristics of the studied programming languages without significant differences
Max number of tables per	Not limited	Not limited	Not limited	There are minor differences
Max number of tables in the database each connection	Not limited	Not limited	Not limited	There are minor differences
Max number of users DATABASE	Not limited	Not limited	Not limited	There are minor differences
Recommended capacity per user	Changeable value	Changeable value	Changeable value	Есть отличия но незначительные

Usually BIG DATA is described using the following characteristics[7]:

1. Volume - The amount of data generated and stored. The size of the data determines the significance and potential of the data, and whether it can be considered as BIG DATA.

2. Variety - The type of data. Big Data can consist of text, images, audio, video. Big Data can complement missing data when matched against each other.

3. Velocity - speed. Here we mean the speed at which data is generated and processed. Very often Big Data is used in real time.

4. Variability - The inconsistency of the data sets can hinder the processing and management of the data.

5. Veracity - the quality of data directly affects the accuracy of data analysis.

National standard "Information technology. Big Data Reference Architecture. Part 1: The structure and application process" is part of a series of five standards on Big Data Reference Architecture and is a Russian language adaptation of international technical report ISO/IEC TR 20547-1:2020 The document contains a description of the structure of a reference architecture for big data, and also provides a solution to the problem of mapping possible uses of big data in the reference architecture. The provisions of the national standard can be applied by organizations to describe the architecture of specific systems to work with big data and the implementation of these systems, taking into account the technologies used, as well as the roles / implementers and their needs.

To date, many methods and technologies for collecting, aggregating, manipulating, analyzing and visualizing big data have been created and adapted. These methods and technologies are drawn from various fields, including statistics, computer science, applied mathematics, and economics

REFERENCES

1. БОЛЬШИЕ ДАННЫЕ. ГОСТ РИСО/МЭК 20546-2019

2. Sridharan S., Purcell B. How Analytics Drives Customer Life-Cycle Management Vision: The Customer Analytics Playbook (2015) // Forrester Research Inc., Cambridge. USA. 16 p

3. Chekharin E.E. Big data: big problems // Psychological Science and Education. 2016. № 3 (21). URL: http://cyberleninka.ru/article/n/bolshie-dannye-bolshie-problemy (date of reference:25.01.2022).

4. James Combs, Yongmei Liu Angela Hall, David Ketchen How much do high-performance work practices matter? a meta-analysis of their effects on organizational performance. https://www.researchgate.net/publication/238606593_How_Much_Do_High-Performance_Work_Practices_Matter_A_Meta PERSONNEL PSYCHOLOGY 2006, 59, 501–528

5. Huselid, M.A. (1995) The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. Academy of Management Journal,38,635-672.http://dx.doi.org/10.2307/256741

6. Попазова О.А., Шихова Н.Н. Оценка и аттестация персонала. СПб.: Изд-во СПбГЭУ, 2019. 79 с.

7. Ушаков И. А., Котенко И. В., Крылов К. Ю. Анализ методик применения концепции больших данных для мониторинга безопасности компьютерных сетей // Материалы IX Санкт- Петербургской межрегиональной конференции «Информационная безопасностьрегионов России (ИБРР-2015)» (Санкт-Петербург, 28–30 октября 2015 г.). СПб., 2015. С. 75–76. (дата обращения: 23.03.2022)