Using the Big Data in the human resources management systems

Goker Oge*, Kesan Yusuf Capraz Applied School, Trakya University, 22800 Edirne, Turkey
Murat Topaloglu, Kesan Yusuf Capraz Applied School, Trakya University, 22800 Edirne, Turkey

Abstract

Measures like setting a destination, evaluation of performances, planning of the labour force about business, personnel recruitment, charging of personnel, informations about personnel, analytical of labour force and reporting, etc., are considered as vital problems of the human resources management. In this study, these criteria in a variety of data were collected with a large repository Apache Hadoop Distributed File System file system owned. Data entry and analysis were used Apache Pig and Java programming languages. The aim of the study is to help 'the owners of business' evaluate the abundance of data and to get rid of the management complexity via an application which is the biggest problem of management system in human resources.

Keywords: Big Data, management of human resources, Hadoop, Pig, map-reduce.

---

* ADDRESS FOR CORRESPONDENCE: **Goker Oge,** Kesan Yusuf Capraz Applied School, Trakya University, 22800 Edirne, Turkey.
 E-mail address: gokeroge@trakya.edu.tr /Tel.: +90-537-3908221

1. Introduction

The power of knowledge increases day by day with the rise of information technologies which also reshapes our lives. All of the shopping and service transactions have become recordable, especially with the smartphones coming into the picture. The size of the data saved keeps growing every day as a result of the soft infrastructure (Tuzunturk, 2010). The businesses have to keep up with the competition and make use of information and knowledge in an effective way to get ahead. They can also get some personal information through social media without permission related to their field. Trying to store such large data whose size cannot be calculated using ordinary methods means waste of time and resources. Since the 1980s, the size of the data in the world has been approximately doubling in every 40 months (Turkington, 2013). In 2000, the data stored in all around the world were estimated at 800,000 petabyte while it is expected that the amount of data will reach 35 zettabyte by 2020 (Dean & Ghemawat, 2008). With such a large amount of data in the ecosystem, the business world will have difficulty in finding enough people to employ. In this case, the most important role falls to the choices done by the human resources managers and the human resources management systems used by the organisation. The basic need of the organisations today is the analytics to derive the most meaningful and valuable information from the data that keeps growing rapidly.

Hadoop platform with its distributed architecture whose details were released by Google and which was turned into a project by Apache later on was developed in Java programming language (Lam, 2010). The programs on which the algorithms are written for data analysis are developed according to the Map Reduce programming model (Geylan, 2002). Hadoop makes the data process more secure, effective and scalable, thanks to its various properties. Hadoop Distributed File System (HDFS) provides a medium where the saved data can be distributed to more than one nodes and retrieved from those nodes whenever needed.

Map Reduce is a method used for the processing of huge amount of data (Hilbert & Lopez, 2011). After the Map function which is used to filter the data and the Reduce function which is used to get results are written, they are run on Hadoop. The Map's function is to classify and sort the data input and then transfer them to the Reduce function (James, 2014). That is, Hadoop is responsible for the simultaneous processing of threads by distributing them on a cluster and the reapproximation of the data (Dijcks, 2013).

2. What is human resources management?

To put it simply, human resources is the processes as a whole between the employer and the employees. At the same time, it serves as a unit to provide the human resources needed and to use these resources in an effective way. Human resources management aims to find and get the human resources that enable the organisation to compete today and in far future and to reach the goals. For that reason, human resources management is the discipline that plans and implements the policies and strategies that are necessary to obtain and maintain the human resources and to improve their performances (White, 2012).

2.1. The problems occurring during recruitment

In addition, the processes that the companies follow related to their field, recruitment, which is one of the main functions of human resources, a also of importance as it shapes the strategic development of the organisation. Human resources management also includes the identification and realisation of internal customer needs (Ouzounis, Syrris & Pesaresi, 2013). Placing the right person for the positions at the right time following the right steps directly affects the success and image of the company. Although this process is done much more professionally compared to the past with the use of human resources applications, evaluation centres and interview techniques based on competence, there are

still some problems occurring during the recruitment. These wrong and rights shape the company's success, move the companies to the front or behind the competition.

The problems occurring during the recruitment can be summarised as follows;

• Creating the job postings without considering the long-term plans of the organisation,
• Not utilising human resources consultancy services,
• Asking the wrong questions during the job interviews and not allocating enough time for the interviews,
• Choosing unexperienced people to do the interviews.

3. Materials and methods

3.1. The development of the algorithm for the recruitment problems

The data mentioned in large data analysis studies can be divided into two groups which are structured and unstructured. Structured data is classified and easy to analyse. The list on which the names of the applicants for a job post and the dates of application are shown can be an example of structured data while unstructured data are made of texts, visuals, schemes and they are more difficult to analyse compared to structured ones. For example, the CVs and covering letters of the applicants are types of unstructured data (Akgeyik, 2011).

In order to improve the decision making and recruitment interviews to getter better and reliable results and to deal with their current problems, organisations may need governance and data integration applications. In this way, receiving the data from the applicants, doing business analytics, forming a data warehouse, managing main data and admission processes will be held successfully. Today, an effective recruitment method is required to find the right applicant in the shortest time possible with the advanced selection criteria. It also lowers the cost and standardises the recruitment reports available any time anywhere which means an automated recruitment process from beginning to end.

In the newly developed application, it is assumed that an organisation receives job applications in a virtual environment. The applications includes name, surname, date of birth, city, military service status, mobile phone number, landline number, e-mail address, home address, alma mater, department, office applications, other applications, native language, foreign language, other foreign languages, etc. After the applications are sent to HDFS, the data is retrieved by Apache Pig, which is one of the Hadoop solutions, and the decision that is made is shown to the human resources management system administrator. To save time, the human resources manager is shown only the first five registries that are most suitable for the job posting and the working conditions in the organisation. The fields such as military service status, marital status, education, other applications, foreign language and second languages are given points, thus the applicants will be ranked by the scores they have accordingly.

The scoring for the applications will be as follows;

• Military service: 5 points for 'done', 0 points for 'not yet',
• Marital status: 5 points for 'married', 0 points for 'single',
• Driving license: 5 points for 'have', 0 points for 'don't have',
• Education: 40 points for doctorate, 30 points for master's degree, 25 points for bachelor's degree, 15 points for associate degree, 5 points for high school diploma,
• Other applications: if there are any other applications that the applicant specify, 10 points for each,
• Foreign language: 20 points for English, 10 points for French, 10 points for German,
• Second foreign language: if the applicants knows an additional foreign language, 15 points for each.

<div align="center">Table 1. Scoring Table</div>

Military service	Score	Marital status	Score	Driving licence	Score	Education	Score
Done	5	Married	5	Have	5	Doctorate	40
Not yet	0	Single	0	Do not have	0	Master's degree	30
						Bachelor's degree	25
						Associate degree	15
						High School	5
Foreign language	Score	Other applications	Score	Second foreign language	Score		
English	20	If there are any,	10	If there are any,	15		
French	10						
German	10						

By this means, all of the applicants will be graded out of 100 and the applicants with the first five highest score will be listed for the human resources manager. The human resources manager will not see those who are not on the list, therefore will not deal with thousands of applications.

3.2. The newly developed application

The data collected are added to the database and transferred to HDFS. After the applications received are saved in a .csv file, they are moved to HDFS and these files are run with the help of 'hadoop dfs-ls' command.

Apache Pig, as a data processing platform making procedural data flow writing on Hadoop possible for its users, has a structure that grants access to an upper layer without writing advanced Java codes for Hadoop's strong, distributed and flexible structure. The procedural structure of Apache Pig is as follows; A = Load Data, C = Transform A, D = Transform C and Dump D. With the load command, the data are retrieved from HDFS and it is transformed with the help of variables and it can be printed on the screen with the dump command.

Apache Pig seems insufficient for the development of the algorithm as Pig programming language is a procedural language with steps following each other from top to the down. All of the data in one step can be transferred to the next step but cannot be stored. That is, if the data are not saved during an intermediate step when the session ends, it is not possible to access the data.

Pig–f insankaynaklari.pig (insan kaynakları: human resources) command is run with Apache Pig and it automatically defines the programming jobs, map and reduce functions. Map-reduce function retrieves the incoming values with the defined jobs in map function and sends them to reduce function to be processed. After the data are processed with the reduce function, it is rewritten on HDFS. The file that is processed is saved as supergroup on HDFS. With the help of job controls, Apache Pig forms the necessary jobs and retrieves data from the localhost. The retrieved data enters the Map-Reduce functions instantaneously and the percentage completed appears.

After all of the data go through map-reduce functions, we can see infiles and outfiles on HDFS, necessary counters, defined and used jobs and the alert that tells the map-reduce function was successful. But most importantly, we can see the first five highest scores on the terminal screen, thanks to our algorithm.

4. Analysis and evaluation

As a result, the data obtained are seen as the first five people with the highest scores who were filtered with the extended data management and a high-performance parallel frame. Thanks to our

program working on HDFS, the human resources manages will interview the first five people instead of interviewing 18 persons one by one.

5. Results and discussion

The aim of this study is to develop an application that runs huge amounts of integrated data in a human resources management system in order to help the analysis and the decision making process after collecting human resources data that cannot be calculated exactly in a large data pool. The application accelerates the decision making process of human resources management system in a secure way with big data while improving the performance by determining the extent to which the staff need to improve for the future with the staff information obtained.

The application offers solutions using big data for the problems in human resources management by doing analyses for internal measurements while making the examination of external criteria at the same time. The human resources departments of the organisations can also identify and remove the blind spots by making the right decisions, taking necessary precautions. Therefore, the organisations will be more successful in terms of management.

In short, when the big data methods used in this study are integrated with the right analysis methods, they can help the companies for the strategic decisions, risk management and innovations to be made. Within the scope of the study, two different disciplines were synthesised in order to improve the goals and steps of the human resources management system using the newly developed application. Therefore, the human resources managers will be able to list the right person(s) in a short time by processing the staff information using big data solutions.

References

Akgeyik, T. (2011). Musteri odakli insan kaynaklari yonetimi. *Iktisat Fakultesi Mecmuasi, 55*(1), 853.

Dean, J. & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107–113.

Dijcks, J. P. (2013). Oracle. *Big Data for the Enterprise*. Retrieved June 27, 2014, from http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

Geylan, R. (2002). *Insan Kaynaklari Yonetimi 4* (vol. 6, p. 101). Eskisehir, Turkey: Anadolu Universitesi Isletme Fakultesi Unite.

Hilbert, M. & Lopez, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science, 332*(6025), 60–65.

James, J. (2014). *How much data is created every minute*. Retrieved April 12, 2014, from http://www.domo.com/blog/2012/06/how-much-data-is-created-every-minute/

Lam, C. (2010). *Hadoop in action*. Manning Publications Co..

Ouzounis, G. K., Syrris, V. & Pesaresi, M. (2013). Multiscale quality assessment of Global Human Settlement Layer scenes against reference data using statistical learning. *Pattern Recognition Letters, 34*(14), 1636–1647.

Turkington, G. (2013). *Hadoop beginner's guide*. Packt Publishing.

Tuzunturk, S. (2010). *Veri Madenciligi Ve Istatistik*. Uludag Universitesi Iktisadi ve Idari Bilimler Fakultesi Dergisi.

White, T. (2012). *Hadoop: The definitive guide*. O'Reilly Media, Inc.