DOI: https://doi.org/10.53052/9788367652018.03

Marcin BERNAŚ[1], Krzysztof WITOS[1], Aleksandra KŁOS-WITKOWSKA[1], Vasyl MARTSENYUK[1], Georgi DIMITROV[2], Dejan RANCIC[3], Oleksiy BYCHKOV[4]

RYNEK PRACY BIG DATA: PRZEGLĄD WYMAGAŃ W BRANŻY

Streszczenie: Badanie zostało przeprowadzone w oparciu o ogólnopolskie portale pracy, które prezentują oferty pracy z ich rynku. Ankieta została przeprowadzona online za pomocą narzędzi formularzy Google. Ze względu na różne formaty i specyfikacje portali w każdym przypadku dane zostały zebrane przez naukowców na podstawie wyszukiwania fraz. Badanie przeprowadzono w okresie od 1 września 2020 r. do 28 lutego 2021 r. W celu uzyskania szerokiego zakresu danych pytań wielokrotnych, z dodatkowymi opcjami pola otwartego, zaproponowano złagodzenie efektu zawężenia odpowiada na sugestie. Portale zostały wybrane na podstawie popularności i liczby ofert ogólnych. Dane ankietowe zostały przedstawione w formie ilościowej. Dane zostały zagregowane do kategorii na podstawie analizy składniowej, tj. różnicy formy, a nie znaczenia. Analiza ankietowa pozwoliła nam uzyskać listę twardych i miękkich umiejętności wymaganych na rynku pracy.

Słowa kluczowe: Big Data, rynek pracy, umiejętności twarde, umiejętności miękkie

BIG DATA LABOR MARKET: AN OVERVIEW OF THE DEMAND IN THE FIELD

Summary: The research was performed based on the national job portals that present the job offer from their market. The survey was performed online using google forms tools. Due to the various formats and specifications of portals in each case, the data was collected by scientists based on phrase search. The survey was performed during a period from the 1'th of September 2020 to the 28'th of February 2021. To obtain a wide range of data multiple question fields, with additional open field options, there were offered to mitigate the effect of narrowed answers suggestions. The portals were selected based on popularity and the number of general offers. The survey data was presented in quantified form. The data were aggregated to the category based on syntactic analysis i.e., the difference in form and not in meaning. The survey analysis allowed us to obtain the list of hard and soft skills that are required by the labor market.

Keywords: Big Data, labor market, hard skills, soft skills

[1] Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: (mbernas, kwitos, awitkowska, vmartsenyuk)@ath.bielsko.pl
[2] University of Library Studies and Information Technologies, Sofia, Bulgaria: g.dimitrov@unibit.bg
[3] University of Niš, Nis, Serbia: dejan.rancic@elfak.ni.ac.rs
[4] Taras Shevchenko National University of Kyiv, Kiev, Ukraine: oleksiibychkov@knu.ua

1. Introduction

The requirements of the current state of labor market and their impact onto competences of IT specialists have been elier studied in [1-10]. The given research was carried out in the context of project no. 2020-1-PL01-KA203-082197 entitled "Innovations for Big Data in a Real World" and is focused on the specialists in Data Science. Particularly we are interested in the competences in the area of Big Data. The survey was obtained by the scientists based on national job portals that present the job offer from their market. The survey was performed on-line using google forms tools. Due to various formats and specification of portals in each case the data was collected by scientists based on phrase search. The phrase used is "Big Data" statement. The survey was performed during a period from the 1'th of September 2020 to 28'th of February 2021. To obtain a wide range of data multiple question fields, with additional open field option, were offered to mitigate the effect of narrowed answers suggestions. The survey contains both open and close questions. To make a process of data collection unbiased no additional recommendation was added. No events were reported during that time that could influence the result. The job offers were selected at random on those portals. The portals were selected based on popularity and number of general offers. The survey data was presented in quantified form. The data was aggregated to the category based on syntactic analysis i.e., the difference in a form and not in a meaning (e.g., Master/MSc or X/N.A.). The open description was presented and analyzed without modification.
This survey is a part of the research within IO1 in connection with the objectives of project 2020-1-PL01-KA203-082197 "Innovations for Big Data in a Real World" (iBIGworld) under the Erasmus+ program. The questions in this study aim to research the needs and expectations of business organizations and to create a profile of a specialist in the field of Big Data. Despite direct Employers survey this survey allows to obtain additional information on specific skills for Big Data experts profile and specific job opportunities.

2. Collection and analysis of data

The data was acquired by four partner institutions scientists. In research 23 (41.8%) questionnaires were obtained from Bulgarian market, 20 questionnaires (36,4%) from Polish market, 9 (16.4%) from Serbian market and 3(5.5%) from the Ukraine market. In total 55 questionnaires were collected by 14 researchers.

3. Results

3.1. Place of work

The first question of the survey asks where jobs will take place. It is worth noting that four markets of project participants countries were analysed. The results are presented on Fig. 1.

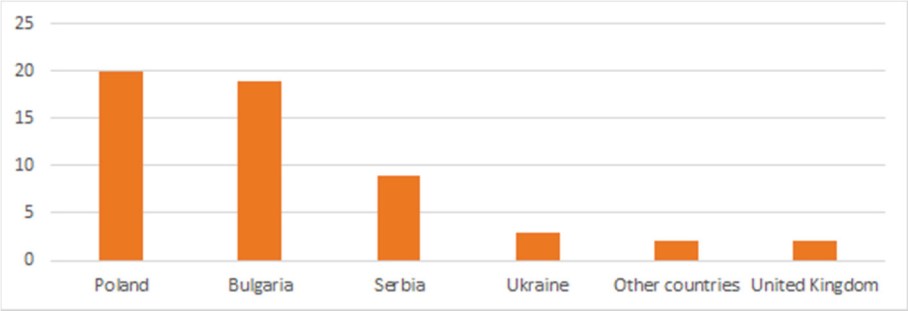

Figure 1. Countries of the job

Data description

According to the research data, the job that is offered in Poland, Ukraine and Serbia takes place in those countries. On the other hand, job offers in Bulgaria for 7,3% of offers take place abroad. The job is also in Denmark, United Kingdom, Italy, Spain and Germany. Nevertheless, all the offers take place in the EU area and UK.

Discussion

The survey shows that, on the market of research countries, there is a need for new employees with Big Data skills. Despite the limited number of offers the market is in constant need of new employees. The offers are present in the portal for a longer period of time. This shows how it is difficult to find expert employees in this field.

Main conclusion:
- There is a need for Big Data employees on the market
- The work place is in EU countries and the UK.

3.2. Position offered

The next question was aimed to analyse what kind of job offers the Big Data knowledge and skills are needed. The analysis was presented on Fig. 2.

Data description

According to the survey the Big Data skills are required in various occupations. What is worth to mention, there is no strict occupation considering Big Data scientist/ engineer (1 entry). The offers, where Big Data skills are required, are Data Scientist (32.7%), Data Engineer (21.8%) and Data Analyst (16.4%). In other occupations the Big data knowledge is also used but not to that extent. The occupations that are gaining popularity are Security (2 offers), tester (1 offer) and manager (1 offer).

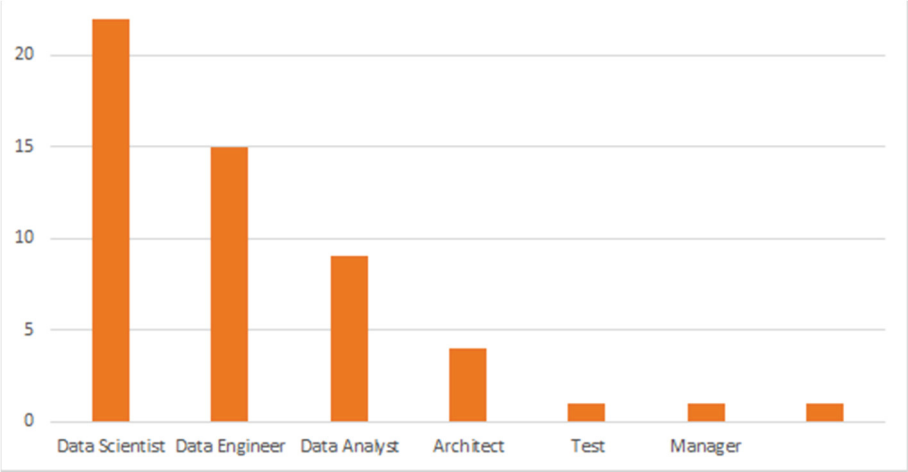

Figure 2. Positions in Big Data

Discussion

The Big Data specialist is considered as the person, whose knowledge exceeds data engineer and data analyst. Due to this, the offer also includes positions for architects. Thus, Big data specialists should have knowledge of a full pipeline, (full solutions in the project) where not only data processing and analysis is visible, but also the full system architecture presented. In turn, due to the fact that there are offers for the positions of a manager and people related to data security, it is also worth addressing these issues at least to a minimal extent as part of the course.
Conclusions:

- The Big Data knowledge is a supplementary one to Data Scientist, Data Engineer and Data Analyst occupations,
- The Big Data is a part of IT scientists and data analytics.

3.3. Type of employment

The question is meant to precise, what kind of employment is preferred for Big Data experts. The results are presented on Fig. 3.

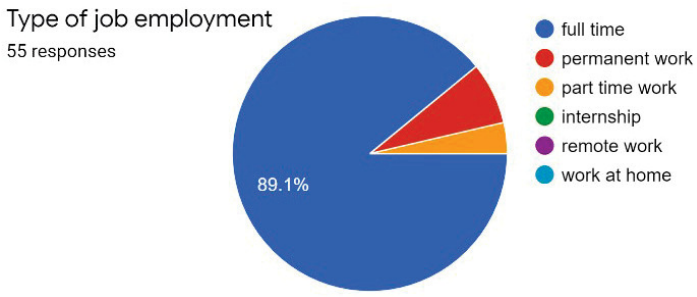

Figure 3. Type of employment in Big Data

Data description

The analysis shows that the majority of offers are presented for full time workers (89.1%) and permanent workers (7.3%) that give in total 96.4%. Only in case of 3.6 % the part time work is proposed.

Discussion

The type of job shows that this kind of field requires full engagement of employers in the process and cannot be realised as short-term part time work.

Conclusions:
- The difficulty of the subject requires full engagement.

3.4. Employers' profile

Under this question the profiles of the companies, which offer the job opportunities in the area of Big Data, are analysed. The size of companies is presented on Fig. 4.

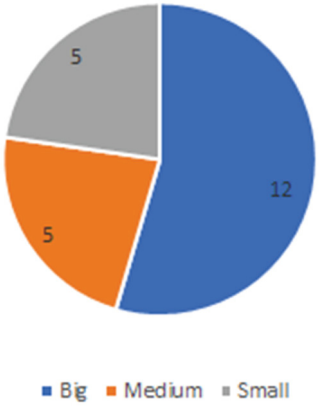

Figure 4. Size of the companies in Big Data

Additionally, in this question the profiles of companies were also analyzed (Fig. 5):

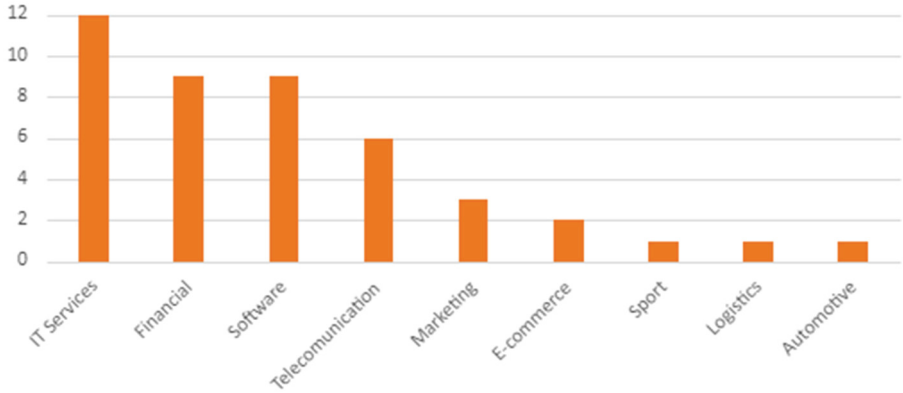

Figure 5. Profils of the companies

Data description

The companies and organizations indicated in the survey are mostly large companies/corporations employing several thousand employees. They operate mainly in the IT service sectors. The finance sector is also an important place, right after IT services. Another sector that stands out is the telecommunications industry. It is worth to note that IT services can be provided for various areas.

Discussion

Due to the variety of sectors applying Big Data processing takes place, the course (expert profile) should focus on the basics, which will certainly be used in these sectors. Based on the questionnaires, there is no just one sector we should focus on.

Conclusion:

- The general skills are needed, not dedicated for a specific area.

3.5. Education and qualification requirements

The question allows us to verify what kind of additional education is preferred for Big Data experts. Data is presented on Fig. 5:

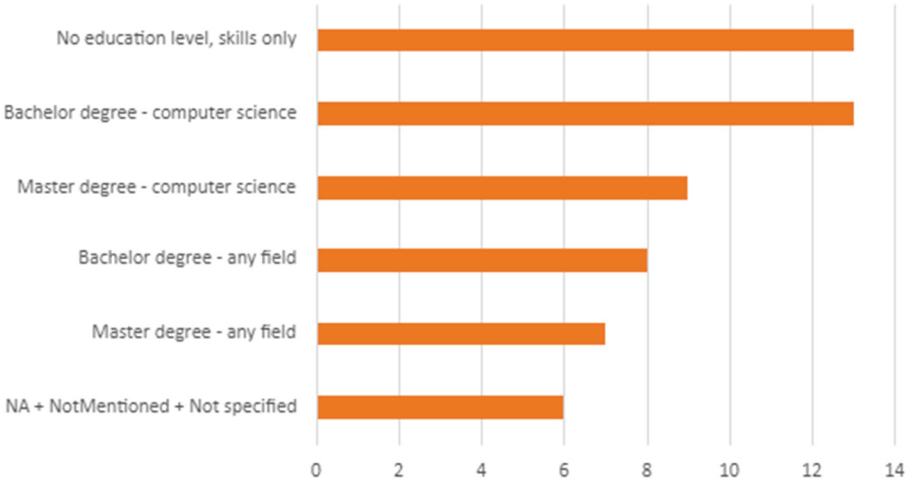

Figure 6. Educations requirements

Data description

The job offers are analysed to find the employee education profiles. The vast majority of offers contain requirements for minimum education in the subject of Big Data. In most cases, the BS level is sufficient to apply for a position. There is also a significant number of offers, where a professional title or academic degree is not required. The skills you have are important.

Discussion:

Due to Big Data scientists being closely related to data scientists (IT related), naturally the Big Data offers are connected with bachelor or master's degree programs. In analysis computer science is preferred over other fields of studies. Due to the significant number of offers, where education is not required, the employees show that the technical skills are equally important as the education, thus this should be taken under consideration, while preparing guidelines and profile of Big Data experts.

Conclusion:

- The Big Data course should be naturally offered as a part/extension of Computer science study.
- Can be offered as skills training as part of post diploma course.

3.6. Work experience requirements

The question illustrates the experience requirements for job offers for analysed countries, presented on Fig. 7.

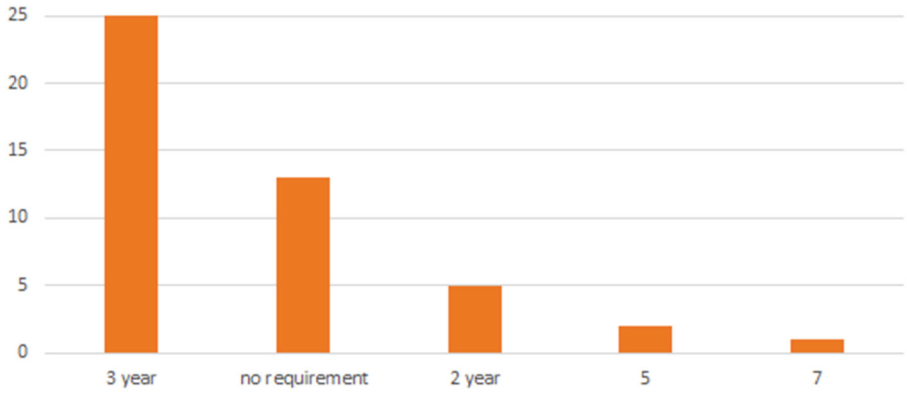

Figure 7. Work experience requirements

Data description

The data shows two types of employment requirements. The market is looking for experts with at least three (43.6%) or two (9.1%) years' experience. Second group shows the group where the experience is not required (23.6%). This shows that the majority of offers are dedicated for experts, however inexperienced workers are also needed. Last groups are seniors with 5 or 7 years' experience that are desirable to lead projects.

Discussion

The profile of a Big Data expert requires optimally 3+ experience and more, but the visible demand on employees with no experience in the field of Big Data indicates that the experience is a major factor and should not be neglected. The proposed course should focus on the basic issues of Big Data that an employee should have at the

beginning of his career in this industry, but also allow to gain experience e.g., by taking part in some project.

Conclusions:
- Both inexperienced and experienced Big Data specialist are needed.
- Course should be focused on basics.
- The course should cover some practical experience (e.g., project).

3.7. What programming languages does the seller know

Next question allows us to analyse the programming language trends covering Big Data topics. The result of analysis was presented on Fig. 8.

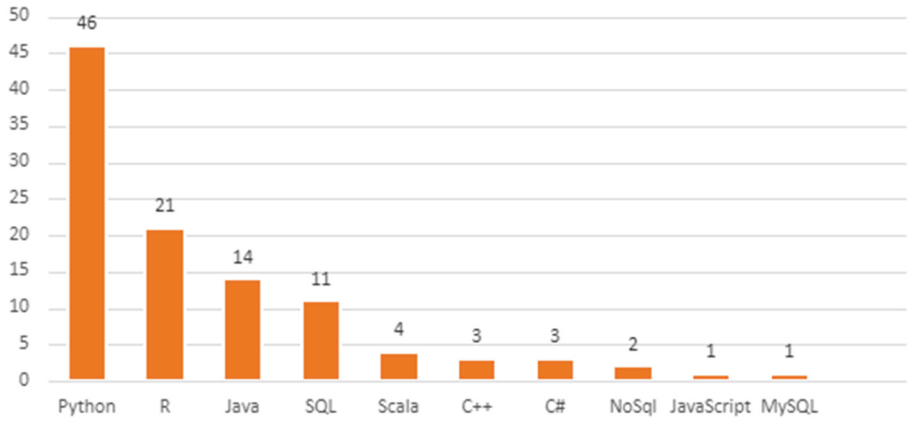

Figure 8. Programming languages requirements

Data description

Among a vast number of programming languages, the four of them are the most popular. They are, according to the survey, Python (81.8%) followed by R (38.2%), Java language (21.8%) and SQL language (20%). Other technologies were also mentioned as supporting technology, nevertheless without Scala (7.2%), C and C++ (5.5%) its requirement was specific only for one job offer (1.8%).

Discussion

The Big Data expert should include learning Python and, to a lesser extent, also elements of the R language. Due to the subject of databases, basic knowledge of the SQL language is also assumed. However, this SQL topic seems to be sufficiently covered in the basic field of computer science studies.

Conclusions:
- Python language should be incorporated as a needed skill,
- the R language, Java and SQL should be treated as part of a training program or perquisites for a course.

3.8. What general personal skills are required / offered

In the modern world not only hard skills but also soft skills are needed for future employees. The question shows what skills are needed for a job connected with Big Data (Fig. 9).

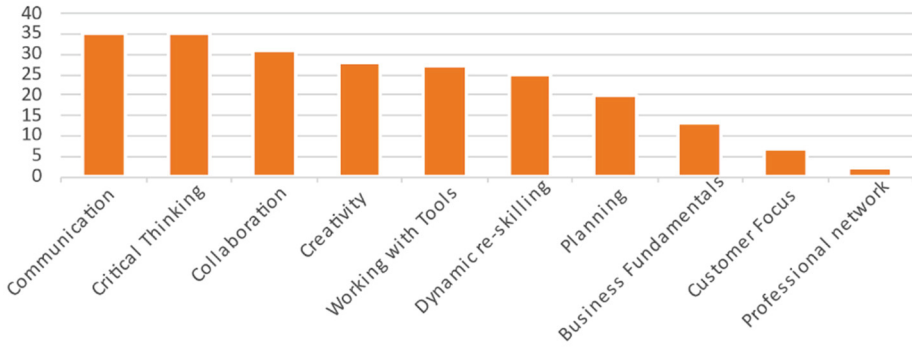

Figure 9. Soft skills related to a job in Big Data

Additionally, the job offers allows to extend soft skills (Fig. 10).

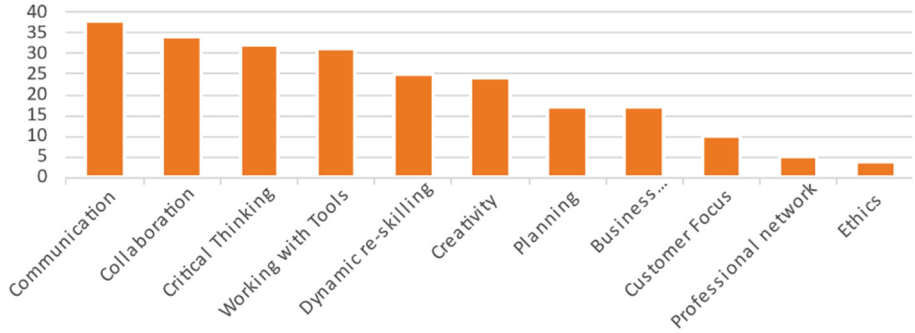

Figure 10. General soft skills

Data description

The survey shows that Big Data employee profile requires Communication skills seems to be very important, which together with critical thinking, rank first among these requirements. Cooperation and creative thinking are also very important. The knowledge of tools is in the next places. This question is related to the Personal Skills Required question and shows that the requirements for candidates are prerequisites, which will later be developed as part of the work performed for the employer. The requirements expected and offered coincide strongly together.

Discussion

The above results show that Big Data user profile should also take soft skills into consideration. The attention should be paid to the aspects of cooperation and creative thinking. In order to develop these skills, a sensible solution seems to be to implement

projects that would be implemented within a number of projects individually and in subgroups.

Conclusion:
- soft skills covering Communication, Critical thinking and collaboration must be developed within the course (expert profile),
- the skills: working with tools, reskilling and Creativity should be supported in the course (expert profile).

3.9. What competencies are needed for this job

The question is analysing on which part of Big Data issues the course should be focused and what part of issues should be addressed in the proposed curriculum. The results are presented Fig. 11.

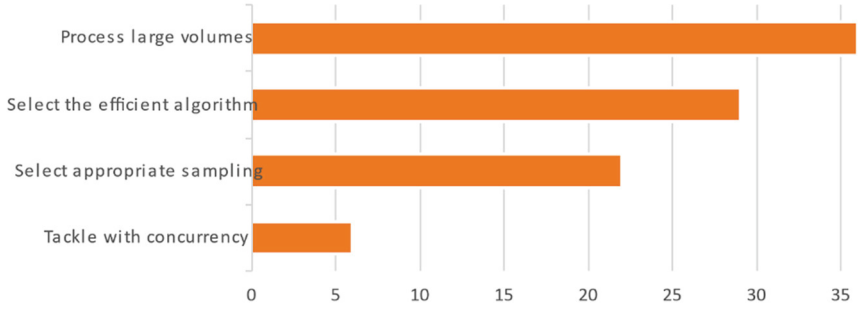

Figure 11. Competencies in Big Data

Data description

Among the required competences, the ability to process large data sets is distinguished. It is also important to choose the right algorithms which will allow us to receive expected information. The important information is that the concurrency is not a favoured feature anymore.

Discussion

Processing Big Data is a fundamental task in Big Data. The survey results indicate that the course should cover this range to an appropriate degree. It makes sense to introduce a real example (preferably a few) to the course, where the course participants will be able to practice at least one scenario on their own. The concurrency, due to a Big Data headway is already addressed by them thus the offers require specialist for data ingestion, processing and analysis.

Conclusion:
- The prepared curriculum should focus on processing Big Volumes of data and knowledge of appropriate algorithms to tackle with.

3.10. What related competencies are needed?

The question aims to find the technologies and specific areas that should be taken under consideration, while preparing Big Data curriculum. The data analysis is presented on Fig. 12.

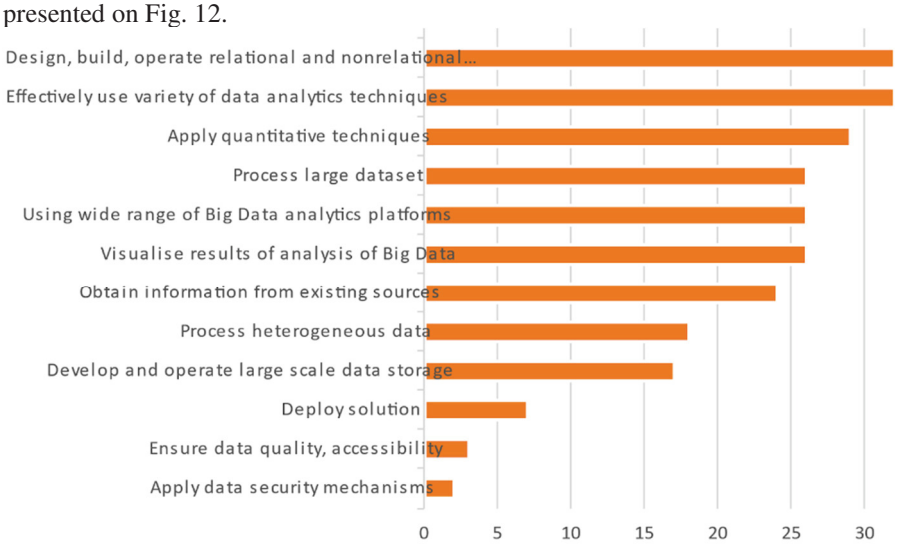

Figure 12. Related competencies needed

Data description

Knowledge of various data analysis techniques are indicated as related competences. They are also on the list of the most important skills: knowledge of relational and non-relational databases and the ability to visualize processed information.

Discussion

The course should contain as much information as possible about the various techniques of data analysis. Ideally, it should not be just a theoretical presentation but also practical so that the person who completed the course could already use them and not only know them theoretically.

The growing interest in non-relational databases was also reflected in the survey results. It is recommended that the items non-relational databases were also included in the course both in theory and in practise. Suggested solutions such as MongoDB, Neo4j or Azure Cosmos DB.

Conclusion:

- The course should provide description of various analytical techniques, processing SQL, NoSQL datasets, visualise data, using analytical platforms and process large datasets.

3.11. Big Data tools that the job seller uses

The question analyses the common tools that are used by the employers. The summary of results was presented on Fig. 13.

Figure 13. Big Data tools

Data description

The results show that most desired are open source Big Data analytics platforms like Hadoop, Spark or Data Lakes. They are followed by distributed commuting tools like MapReduce, Mahout, Lucene or NLTK. The particular solution is a pipeline proposed by Amazon AWS (21%), Hadoop (20%) and Microsoft Azure (14,5%). The analysis shows also the interest in streaming data processing (18,1%) like Flume, Kafka or Storm. In questionnaire there is mentioned particular technology from the presented pipelines like Scala, BigQuery or Pandas.

Discussion

The presented information shows that there are many technologies available on the market. There are classical open source solutions as Hadoop or Apache Spark. However, more and more become popular solutions based on clouds like: AWS, Microsoft Azure or Google solutions (like Big Query). Nevertheless, most cloud solutions despite their proprietary solutions support open source tools and standards.

Conclusions:
- the open source tools should be proposed which are available on cloud platforms,
- the full pipeline should be presented,
- the stress should be put on data analytics solutions,
- the streaming and cloud technology should be added if possible.

3.12. Description of the main responsibilities and obligations of the candidate for work position

The open questions allow us to compare the job responsibilities for each occupation. Total 55 descriptions were presented.

The first (word cloud) chart (Fig. 14) was prepared using R language script. All the descriptions were parsed automatically. On this chart we can see a visual representation of the most used words occurring in the job offers under analysis. The importance of each word defines its font size.

Figure 14. Words occuring in the job offers

The second chart (Fig. 15) was made manually, reading job offers one by one creating groups of the most common tools, skills etc. which can be found in the jobs offers we were analysing.

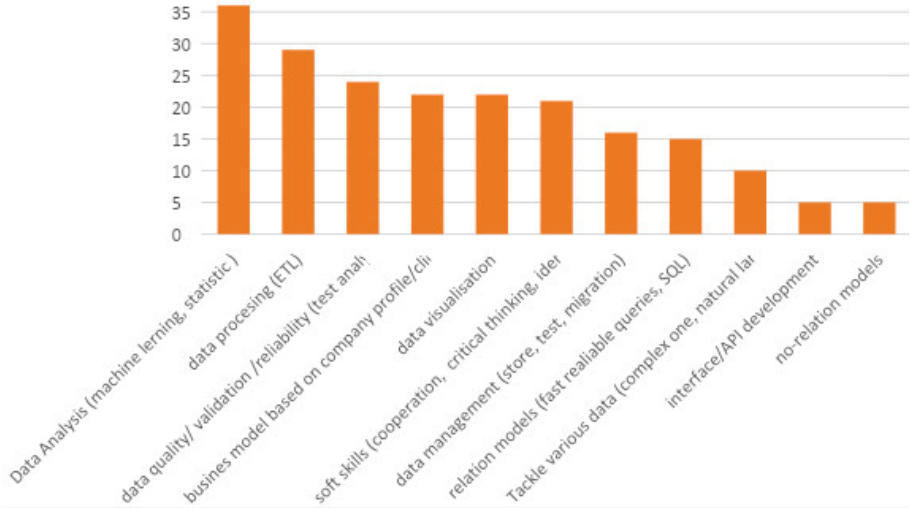

Figure 15. Groups of the most common tasks, tools, skills

Data description

The responsibility description usually mentions data analysis skills using statistics, machine learning (67%), but also ability to write own analytics tools based on existing and new methods. Second is ability to process data (54%) to form accepted by analytics. Using processing the stress is put on data quality, reliability and its validation (44%). 41% of responsibilities point out that knowledge of business models is as important as the technical skills. The stress is put on the ability to connect these two areas (obtain vital information). This element is strongly connected with soft skills, which are present in 39% of offers. Data visualisation is also present in 41% and describes the ability to present the obtained information in a simple form. Finally, the administration responsibilities were presented in (30%) of descriptions. The technical skills to manage structured data (28%) and unstructured one (9%) are also mentioned. Last two elements consider issues with various data types (19%) and development of interfaces to distribute obtained data/information.

Discussion

The full analysis of the description brings to a conclusion that on the market there is a big demand for Big Data specialists, that has a vast knowledge also in data processing, programming and data mining. The one group of specialists needed is a Big Data pipeline developer who is able to collect data, transform it, analyse and receive valuable data in some business context. In case of big solutions, the Big data management personnel is needed (warehouse, data maintenance and migration. Second group are ETL experts who obtain data and transform it for analytics. The analytics group is the biggest however and skills to analyse data (statistic, machine learning) are most needed. The description shows that some pipeline elements ingest, ETL are automated and the experts form data mining are needed. Last group are the analysis tools development teams. It is another step to provide advanced analytics tools for Big Data experts without advanced knowledge of analytical algorithms and solutions. The pipelines are Spark, Python-based solutions where platforms are Hadoop based or AWS or Azure one. The job offers are usually based in some business model and its specification should be known to employees. The responsibilities show the strong need for soft skills – especially collaboration and problem solving. The understanding of enterprise foundation is also essential.

Conclusions:

- the data analytics and data processing skills are most search on the job market,
- business model is important feature in majority of offers,
- soft skills are also important,
- the basic knowledge of pipeline elements is essential,
- the Hadoop ecosystem, AWS and Azure are preferred technologies,
- the Big Data is developing into analytics and processing areas, while the technical elements are automated,
- the job market is looking for issue solvers rather than strict technicians.

3.13. Offered payment

The offers were analysed based on the salary described for each job offer. The information was presented on Fig. 16.

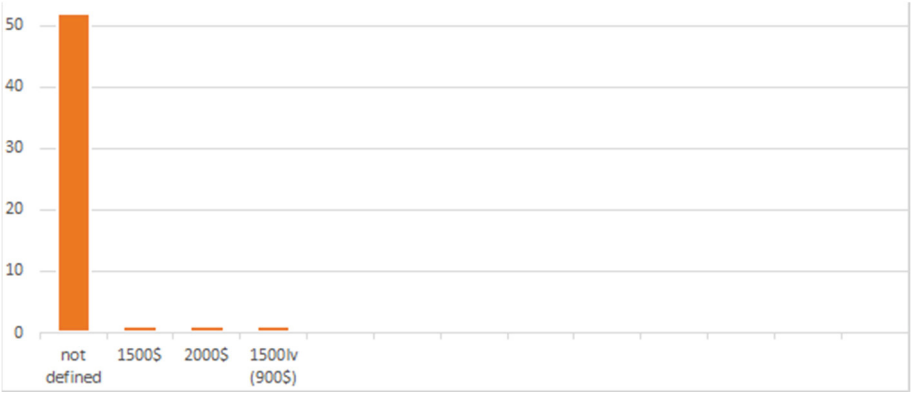

Figure 16. Salary offered

Data description

The job portals are not offering the information on salary. Only in the case of three samples the salary were presented as 1500$, 2000$ and 1500lv (900$).

Discussion

Due to the modern approach the salary is a part of the negotiation process. Most job offers did not contain data on the remuneration offered. Therefore, we used a different data source.
According to Hays's report [4] average salary in Poland for Big Data engineers is 25% higher than programmers and is on average 17 thousand PLN. (hays.com)

In the USA, for example, from $ 80k - $ 220k
(*https://www.northeastern.edu/graduate/blog/highest-paying-big-data-careers/)*

The earnings in Big Data are quite high, which may also result from the lack of specialists in the dynamically developing field of Big Data.
Preparing a course that would offer the acquisition of Big Data skills seems to be a legitimate direction/decision.

4. Conclusions

So, the research was performed based on the national job portals that present the job offer from their market. The portals were selected based on popularity and the number of general offers. The survey data was presented in quantified form. The data were aggregated to the category based on syntactic analysis i.e., the difference in form

and not in meaning. The survey analysis allowed us to obtain the list of hard and soft skills that are required by the labor market.

REFERENCES

1. GRIGORESCU A., ZAMFIR A.-M., SIGURDARSON H.T., LAZARCZYK CARLSON E.: Skill Needs among European Workers in Knowledge Production and Transfer Occupations (2022) Electronics (Switzerland), 11 (18), art. no. 2927, DOI: 10.3390/electronics11182927.
2. JONY S.S.R., KANO T., HAYASHI R., MATSUDA N., RAHMAN M.S.: An Exploratory Study of Online Job Portal Data of the ICT Sector in Bangladesh: Analysis, Recommendations and Preliminary Implications for ICT Curriculum Reform (2022) Education Sciences, 12 (7), art. no. 423, DOI: 10.3390/educsci12070423.
3. TOMMASI F., PERINI M., SARTORI R.: Multilevel comprehension for labor market inclusion: a qualitative study on experts' perspectives on Industry 4.0 competences (2022) Education and Training, 64 (2), pp. 177-189. DOI: 10.1108/ET-04-2021-0146.
4. TAVAKOLI, M., FARAJI, A., VROLIJK, J., MOLAVI, M., MOL, S.T., KISMIHÓK, G.: An AI-based open recommender system for personalized labor market driven education (2022) Advanced Engineering Informatics, 52, art. no. 101508, DOI: 10.1016/j.aei.2021.101508.
5. SCHLEGEL T., PFISTER C., HARHOFF D., BACKES-GELLNER U.: Innovation effects of universities of applied sciences: an assessment of regional heterogeneity (2022) Journal of Technology Transfer, 47 (1), pp. 63-118, DOI: 10.1007/s10961-020-09839-w.
6. KIENER F., GNEHM A.-S., CLEMATIDE S., BACKES-GELLNER U.: IT skills in vocational training curricula and labour market outcomes (2022) Journal of Education and Work, 35 (6-7), pp. 614-640. DOI: 10.1080/13639080.2022.2126968.
7. WYSKWARSKI M. Requirements for project managers—what do job advertisements say? (2021) Sustainability (Switzerland), 13 (23), art. no. 12999, DOI: 10.3390/su132312999.
8. SALAS-VELASCO M.: Mapping the (mis)match of university degrees in the graduate labor market (2021) Journal for Labour Market Research, 55 (1), art. no. 14, DOI: 10.1186/s12651-021-00297-x.
9. LI G., YUAN C., KAMARTHI S., MOGHADDAM,M., JIN X. Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis (2021) Journal of Manufacturing Systems, 60, pp. 692-706. DOI: 10.1016/j.jmsy.2021.07.007.
10. ALTONJI J.G., ZHONG L.: The labor market returns to advanced degrees (2021) Journal of Labor Economics, 39 (2), pp. 303-360. DOI: 10.1086/710959.