Choice Method of Analytical Platform for Smart City

Ihor Bodnarchuk, Oleksii Duda, Alexander Kharchenko, Nataliia Kunanets, Oleksandr Matsiuk and Volodymyr Pasichnyk

October 12, 2020

Choice method of analytical platform for smart cities

Bodnarchuk Ihor1[0000-0003-1443-8102], Duda Oleksii1[0000-0003-2007-1271], Kharchenko Alexander2[0000-0002-5868-3938], Kunanets Nataliia3[0000-0003-3007-2462], Matsiuk Oleksandr1[0000-0003-0204-3971] and Pasichnyk Volodymyr3[0000-0002-5231-6395]

[1] Ternopil Ivan Puluj National Technical University, Ruska str., 56, Ternopil, 46000, Ukraine, bodnarchuk.io@gmail.com, oleksij.duda@gmail.com, oleksandr.matsiuk@gmail.com

[2] National Aviation University, Kosmonavta Komarova ave. 1, Kyiv, 03058, Ukraine kharchenko.nau@gmail.com

[3] Lviv Polytechnic National University, St. Bandera str., 12, Lviv, Ukraine nek.lviv@gmail.com, vpasichnyk@gmail.com

Abstract. The smart cities development results in the expansion of the list of information technologies used for this purpose. These technologies are given below in the form of matrix and formal description. The proposed list of information technologies includes: Internet of Things, Fog Computing, Cloud Computing, Information Models, Intelligent Data Processing and Decision Support Systems, Mobile Technology, Geoinformation Technologies, GRID technology, Confidential Communication and Data Protection Technologies. While implementing the procedures for information technology support of the processes in smart cities, it is a common practice to use the models of analytical platforms which are on the open market. At present, their list is quite extensive. This fact creates uncertainty in the analytical platform selection. In order to select the analytical platform, the technique based on the hierarchy analysis method on the ground of procedure of pairwise comparisons of variants is proposed in this paper. The analysis was carried out among such alternative characteristics as computing resources and cloud infrastructure, availability and reliability, ana-lytics, safety and security, cost, application programming interface, support. The obtained results show the highest efficiency of the method of the IBM Bigdata Analytics platform. It should be noted that the proposed method is to be implemented taking into account the probable need for the expansion of the set of analytical platforms characteristics and parameters.

Keywords: Multicriteria choice, Smart city, Data Processing, Analytical Hierarchic Processing, Information resources.

1 Introduction

Prototypes of modern city information systems must meet most closely principles formulated on the World Summit on the Information Society (WSIS) and ensure effective information technology for support of processes running in resource and socio-communicational networks in the smart city with large population. For this purpose it is necessary to develop information technologies with implemented procedures of analytical processing of certain city data collections. Particularly such information technologies should give possibility to develop and utilize effective tools for analytical data processing implementation. A lot of such tools and means are based on the open analytical platforms providing their services both on commercial and free principles.System analysis of the processes, procedures and tools used for analytical processing of big data in smart cities is represented in paper [1]. Gupta [2] has analyzed data ecosystems and analytical data processing tools in smart cities that are integrated with different types of city systems and services. In paper [3], the authors note that the current investigation of the use of data and analytics means in cities is relatively inaccurate and fragmented and requires more holistic systematic approach. Picardal [4] presents the analysis of the information technologies that is used for city portal creation and for processes support in water supply networks in the Washington (USA). The above mentioned platform ensures flexible adding of functional components and system integration of new information technologies in it. Zschörnig in paper [5] presents the analysis of the original cloud information technology for IoT domains. The efficiency assessment of the offered prototype is carried out by the author. Arun in paper [6], investigating the architecture of IoT systems in smart cities, underlines the importance of the development and practice implementation of analytical data processing means as a separate structure layer. It makes it possible for the city to save the budget on acquiring of on-premises computing resources, with simultaneous decreasing of general financial expenses as payments only for hosting of consumed resources and services. There are a lot of (more then 60) analytic platforms represented on the market aimed for processing of urban data collections with wide list of available methods and tools. With this regard they are presented using various lists of different characteristics. Hence, the ambiguous situation arises in the procedures for selection the appropriate IT platform while developing IT component in smart cities with efficient data processing possibility. For such cases groups of experienced experts selected from available means the most "suitable" and the most "adopted" to the platform. As a consequence, the selection of completely functional, efficient, affordable and easy to use analytical platform is one of the urgent tasks of scientific researches and modern innovation information technology developments. Therefore, the objective of this paper is to facilitate the selection of information technology while designing the smart cities. The set goals result in the statement of the tasks concerning the formation of complete list of information technologies for smart cities and the development of a method for selecting the analytical platform for data processing in the smart city.

2 Information technologies of the smart city

The conclusion is based on the analysis of modern publications on the basis of the following information and communication technologies for smart cities formation [7]: Internet of Things (IoT) – IT_{IoT} , Fog Computing (FC) – IT_{FC} , Cloud Computing (CC) – IT_{CC}, Information Models (IM) – IT_{IM}, Intelligent Data Processing (IDP) and Decision Support Systems (DSS) – IT_{IDP} and IT_{DSS}, Mobile Technology (MT) – IT_{MT}, Geoinformation Technologies (GIS) – IT_{GIS}, GRID technology – IT_{GRID}, Confidential Communication and Data Protection Technologies – IT_{CCDP}. As the result based on the information technologies analysis and their functional coverage of main information processes the matrix is constructed. It is presented as a nested relation and shown in Fig. 1. Basic information technologies are given as an attributes, and types of information processes are given as the tuples.

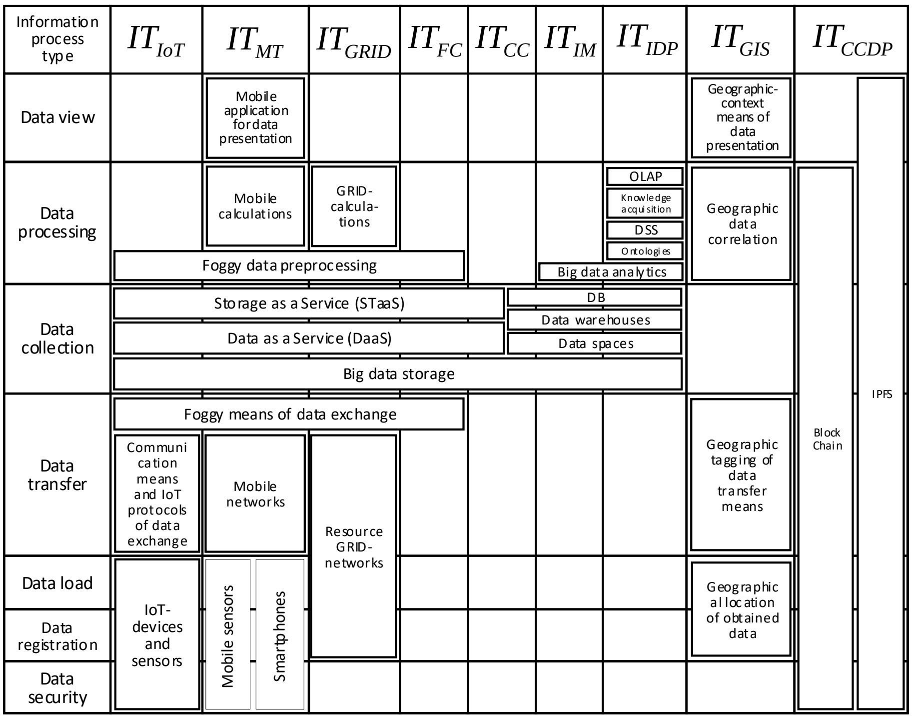

Fig. 1. Information technologies matrix of the smart city

Information technologies matrix is used for the determination of places, roles and connections between basic IT during smart city IT-platform (ITP) implementation. Normalized relation, given by information technologies matrix, are presented as follows:

$$ITP = \langle IT, D \rangle , \quad (1)$$

where $IT = \{IT_I\}$, $I = \{IoT, FC, CC, IM, IDP, DSS, MT, GIS, GRID, CCDP\}$ is a set of basic IT used for composing of ITP in modern smart cities;

$D = \{D_J\}$, $J = \{DS, DR, DL, DT, DC, DP, DV\}$ are the stages of data processing, particularly DR is data registration, DL is data load, DT is data transfer, DC is data collection, DP is data processing, DV is data view, DS is data security.

Detectors and sensors deployed on the bottom layer of the smart city information system (IS) (see Fig. 2) together with the means of data collection in socio-communicational environment are the main sources of heterogenous data sets generation.

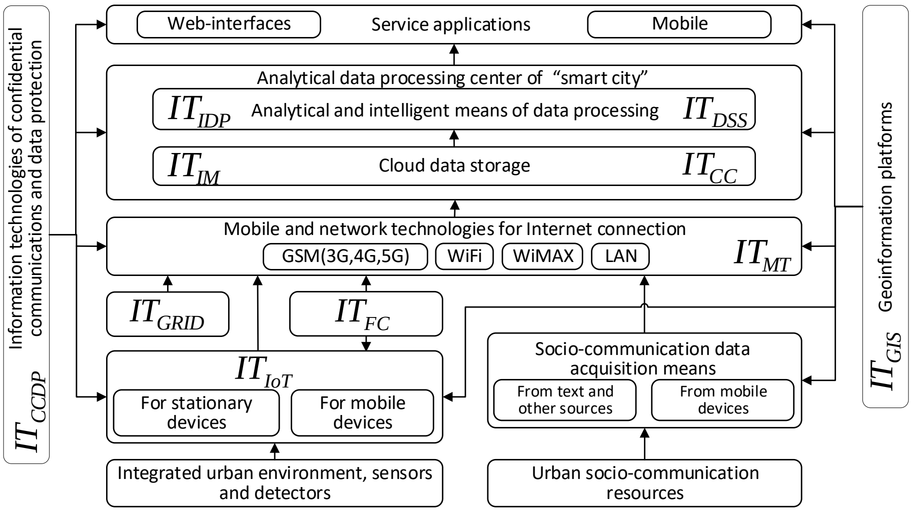

Fig. 2. Information technologies components that implement the smart city

Collected numerous big data sets of different types in information systems of the smart cities are processed typically by the center of analytical processing, which is as rule deployed on the base of cloud data storage [8]. Information sources associated with different components and subsystems of the smart cities make it possible to create big data sets that are not typically used rather effectively.

Modern infrastructure based on information technology enables to aggregate effectively generated data, to consolidate them and to process analytically in order to refine considerable the quality of base processes in urban systems.

3 Selection of the smart city analytical platform

It is proposed to select the analytical platform for smart city's needs with the application of the known expert method built on the procedure of pairwise comparisons of alternatives with the following steps:

Step 1. The three-layer hierarchy is built in order to get the possibility to apply the above mentioned method. There is the goal on the top level of the hierarchy and decision making is directed for this target. The second layer contains the set of criteria, with accounting of which the selection of the alternative analytical platform is conducted for analytical processing of urban data sets. Available alternatives compose the bottom level of the hierarchy. Decision making is conducted on the base of the composed vector of priorities and results in the selection of certain alternative among the available ones. The priority in our case is a real number that corresponds to each alternative. According to the heist priority, one alternative is selected and is treated as a taken decision. Corresponding three-layer hierarchic tree is shown in Fig. 3.

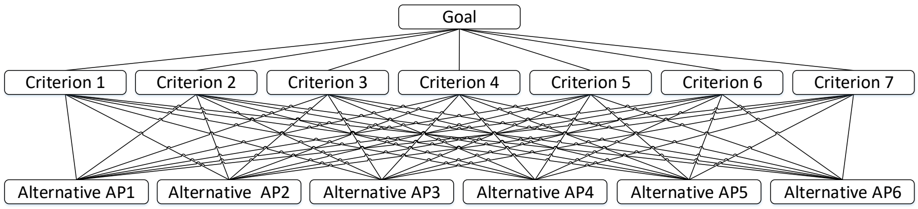

Fig. 3. General scheme of Analytical Hierarchic Processing method for the smart city analytical platform selection

Step 2. Composing the alternatives list. For instance, as the alternatives some popular ITP of analytical data processing are taken, particularly: 1) IBM Bigdata Analytics *(AP_1)* [9]; 2) HP Bigdata *(AP_2)* [10]; 3) Microsoft Bigdata *(AP_3)* [11]; 4) Oracle Bigdata Analytics *(AP_4)* [12]; 5) Google BigQuery *(AP_5)* [13]; 6) Cisco Bigdata *(AP_6)* [14].

Step 3. Assessment criteria listing on the base of the features analysis and possibilities of the selected platform use. The important factor for criteria listing, according to which the "best" ITP is selected, is the possibility of their adaptation to the needs of the smart city with large population. Particularly it also refers to Ternopil.

The set of criteria is formed as the result of system analysis of the selected APs functional possibilities. The pick of a certain analytic platform is made on the basis of these criteria. Seven characteristics of analyzed analytical platform are selected as the main criteria for alternatives weights calculation: computing resources and cloud infrastructure, availability and reliability, analytics, safety and security, cost, application programming interface (API), support.

Step 4. The selection of the scale for expert assessments. It is required to pick one of alternatives on the basis of the formed criteria set. The choice of the alternative is actually the calculation of elements priorities vector. Each element of this vector corresponds to a certain alternative. Thus, the decision making is based on the determination of the alternative with the greatest index. The scale of expert assessments is used for the implementation of Analytic Hierarchy Process (AHP): 1 is uniform importance (the importance of both objects is equal); 3 is weak importance (experienced experts judgments give the first object a slight advantage over the second); 5 is essential or significant importance (experienced experts judgments give the first object a big advantage over the second); 7 is very significant and obvious importance (first object superiority over second is very significant, explicit actually); 9 – absolute importance (first object superiority is more than convincing, actually indisputable); 2, 4, 6, 8 – intermediate values between adjacent scale values (compromised cases); inverse values – if one of the above values is obtained when comparing first object with second, then comparing the second object with the first is the inverse value.

The base of analytical platforms under analysis is that no one of them is oriented on the processing of urban data sets. The structure of the decision-making problem with AHP regarding to the selection of analytic platform is shown in Fig. 4.

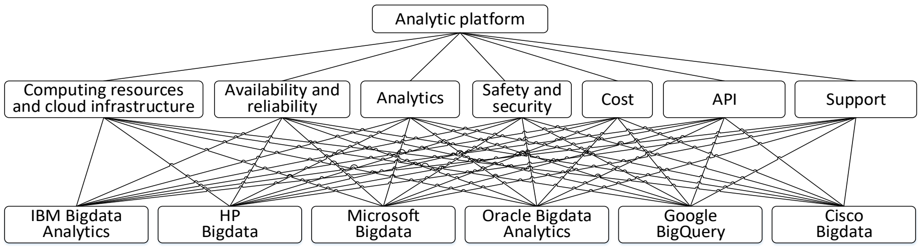

Fig. 4. The structure of the decision-making problem with AHP regarding to the selection of analytical platform for processes support in resource and socio-communication networks of smart cities

Step 5. The matrices of pairwise comparisons are built for each of the above listed criterion for AHP realization in order to select the analytical platform and corresponding numerical characteristics are calculated, particularly consistency index, the grates eigenvalue and consistency ratio. Each of the above mentioned matrices contains expert assessments values regarding to the couples of the analyzed analytical platform. Additional arguments are given for each criterion concerning the particularities of information systems development and its application for the processes support in resource and socio-communication networks of the smart cities when matrices of pairwise comparisons for AHP are built. The "Computing resources" criterion defines the integral characteristics of calculative possibilities of ITPs for their dynamic allocation and release. The matrix of pairwise comparisons for analytical platform selection for "Computing resources and cloud infrastructure" criterion is shown in Table 1.

The results of weights assessments calculations for "Computing resources and cloud infrastructure " criterion are shown in Table 2. The best alternative for "Computing resources and cloud infrastructure " criterion is the analytical platform "IBM Bigdata Analytics" because it has the greatest calculated value of its weight 0.3524. For the matrix of pairwise comparisons, composed for "Computing resources and cloud infrastructure " criterion the following parameters are calculated:

- the assessment of the greatest eigenvalue:

$$\lambda_{\max} = \sum_{i=1}^{n} w_i s_i \ , \qquad (2)$$

Table 1. The matrix of pairwise comparisons for alternatives selection for "Computing resources and cloud infrastructure" criterion

Alternatives	IBM Bigdata Analytics	HP Bigdata	Microsoft Bigdata	Oracle Bigdata Analytics	Google BigQuery	Cisco Bigdata
IBM Bigdata Analytics	1	2	3	5	2	7
HP Bigdata	0.5	1	3	5	3	5
Microsoft Bigdata	0.33	0.33	1	3	1	3
Oracle Bigdata Analytics	0.2	0.2	0.33	1	1	1
Google BigQuery	0.5	0.33	1	1	1	3
Cisco Bigdata	0.14	0.2	0.33	1	0.33	1
Total	2.68	4.07	8.67	16	8.33	20

Table 2. Alternatives weights for "Computing resources and cloud infrastructure" criterion

Alternatives	IBM Bigdata Analytics	HP Bigdata	Microsoft Bigdata	Oracle Bigdata Analytics	Google Big-Query	Cisco Bigd ata	Total	Alternative's weight
IBM Bigdata Analytics	0.3737	0.4918	0.3462	0.3125	0.24	0.35	2.114	0.3524
HP Bigdata	0.1868	0.2459	0.3462	0.3125	0.36	0.25	1.7014	0.2836
Microsoft Bigdata	0.1246	0.082	0.1154	0.1875	0.12	0.15	0.7794	0.1299
Oracle Bigdata Analytics	0.0747	0.0492	0.0385	0.0625	0.12	0.05	0.3949	0.0658
Google BigQuery	0.1868	0.082	0.1154	0.0625	0.12	0.15	0.7167	0.1194
Cisco Bigdata	0.0534	0.0492	0.0385	0.0625	0.04	0.05	0.2935	0.0489
Total	1	1	1	1	1	1	6	1

– consistency index:

$$C_I = \frac{\lambda_{\max} - n}{n - 1}; \qquad (3)$$

– consistency ratio:

$$C_R = \frac{C_I}{R_I}. \qquad (4)$$

The same value of random consistency index is used $R_I = 1{,}25$ for the following calculations of alternatives weights when $n = 6$.

The above mentioned calculated parameters for the matrix of pairwise comparisons for "Computing resources and cloud infrastructure " criterion are as follows:

– assessment of the greatest eigenvalue:

$$\lambda_{\max} = 0.3524 \cdot 2{,}68 + 0{,}2836 \cdot 4{,}07 + 0{,}1299 \cdot 8{,}67 + 0{,}0658 \cdot 16 + 0{,}1194 \cdot 8{,}33 + \\ + 0{,}0489 \cdot 20 = 6{,}2487; \quad (5)$$

– consistency index:

$$C_I = \frac{\lambda_{\max} - n}{n-1} = \frac{6.2487 - 6}{6-1} = 0.0497; \quad (6)$$

– consistency ratio:

$$C_R = \frac{C_I}{R_I} = \frac{0.0497}{1.25} = 0.0398. \quad (7)$$

It is obvious that $C_R = 3.98\% \leq 10\%$, so the matrix of pairwise comparisons for "Computing resources and cloud infrastructure" criterion is consistent.

Matrices of pairwise comparisons for the selection of ITPs by "Availability and reliability", "Analytics", "Safety and security", "Cost", "API" and "Support" criteria and calculated alternatives weights for the listed criteria are calculated in same way. Similar to "Computing resources and cloud infrastructure " criterion, the eigenvalues $\lambda_{\max}$, consistency indices C_I and ratio indices C_R are calculated for all of these criteria and shown in Table 3.

Table 3. Parameters of matrices of pairwise comparisons

Criterion	$\lambda_{\max}$	C_I	C_R
Computing resources and cloud infrastructure	6.249	0.05	0.04
Availability and reliability	6.243	0.049	0.039
Analytics	6.493	0.099	0.079
Safety and security	6.265	0.053	0.042
Cost	6.495	0.099	0.079
API	6.377	0.075	0.06
Support	6.432	0.086	0.069

The equation $C_R \leq 10\%$ is true for all criteria ensuring the consistency for each matrix of pairwise comparisons.

Step 6. Alternative weights assessment. The assessment of the importance grade with respect to each criterion is performed for alternatives weights estimation (see Table

4). The results of alternatives weights assessments calculation with respect to criteria are shown in Table 5.

Table 4. Matrix of pairwise comparisons of alternatives with respect to criteria

Criteria	Computing resources and cloud infrastructure	Availability and reliability	Analytics	Safety and security	Cost	API	Support
Computing resources and cloud infrastructure	1	2	6	4	3	5	5
Availability and reliability	0.5	1	5	5	3	6	3
Analytics	0.17	0.20	1	0.50	0.5	4	0.5
Safety and security	0.25	0.20	2	1	1	4	1
Cost	0.33	0.33	2	1	1	3	1
API	0.20	0.17	1	0.25	0.33	1	0.33
Support	0.20	0.33	2	1	1	3	1
Total	2.65	4.23	19	12.75	9.83	26	11.83

Table 5. Alternatives weights with respect to the main criteria

Criteria	Computing resources and cloud infrastructure	Availability and reliability	Analytics	Safety and security	Cost	API	Support	Total	Criteria weights
Computing resources and cloud infrastructure	0.3774	0.4724	0.3158	0.3137	0.3051	0.1923	0.4225	2.3992	0.3427
Availability and reliability	0.1887	0.2362	0.2632	0.3922	0.3051	0.2308	0.2535	1.8696	0.2671
Analytics	0.0629	0.0472	0.0526	0.0392	0.0508	0.1538	0.0423	0.4489	0.0641
Safety and security	0.0943	0.0472	0.1053	0.0784	0.1017	0.1538	0.0845	0.6653	0.095
Cost	0.1258	0.0787	0.1053	0.0784	0.1017	0.1154	0.0845	0.6898	0.0985
API	0.0755	0.0394	0.0526	0.0196	0.0339	0.0385	0.0282	0.2876	0.0411
Support	0.0755	0.0787	0.1053	0.0784	0.1017	0.1154	0.0845	0.6395	0.0914
Total	1	1	1	1	1	1	1	7	1

Step 7. Results of weighting. Weighted results of the analytical platform selection are given in Table 6.

Table 6. Weighted results with respect to criteria for analytical ITP selection

Platform \ Criteria	Computing resources and cloud infrastructure	Availability and reliability	Analytics	Safety and security	Cost	API	Support
IBM Bigdata Analytics	0.1208	0.1039	0.0254	0.0271	0.0446	0.0194	0.0225
HP Bigdata	0.0972	0.0748	0.0056	0.0263	0.0174	0.0084	0.0395
Microsoft Bigdata	0.0445	0.0282	0.0036	0.0193	0.0049	0.0035	0.003
Oracle Bigdata Analytics	0.0226	0.0129	0.0053	0.0082	0.0106	0.004	0.0111
Google BigQuery	0.0409	0.0344	0.0082	0.098	0.0146	0.004	0.0052
Cisco Bigdata	0.0168	0.0129	0.0161	0.0043	0.0065	0.0018	0.0101
Consistency index	0.0171	0.013	0.0063	0.005	0.098	0.0031	0.0079
Total	0.3427	0.2671	0.0641	0.095	0.0985	0.0411	0.0914

The best alternatives of ITPs with respect to the given criteria and corresponding weights are shown in Table 7.

Table 7. The best alternatives and corresponding weights of analytical ITPs choosing

Criterion	The best alternative	Weight
Computing resources and cloud infrastructure	IBM Bigdata Analytics	0.3524
Availability and reliability	IBM Bigdata Analytics	0,3389
Analytics	IBM Bigdata Analytics	0,4026
Safety and security	IBM Bigdata Analytics	0,3961
Cost	IBM Bigdata Analytics	0,2853
API	IBM Bigdata Analytics	0,4528
Support	HP Bigdata	0,4326

4 Results

In order to assess the reliability of the obtained solution while selecting the alternative, we use the consistency index, containing information about the violation of numerical (cardinal) and transitive consistency of matrices. The limits of application of the hierarchies analysis method are defined if the consistency index is less than 0.1. In the carried out investigation, the calculated consistency index is

0.033, which indicates the high level of the obtained solution reliability. The results of weights calculations are presented in Table 8 and Fig. 5.

Table 8. Alternatives and their weights

Alternative	Weights
IBM Bigdata Analytics	0.0462
HP Bigdata	0.0386
Microsoft Bigdata	0.018
Oracle Bigdata Analytics	0.0122
Google BigQuery	0.0177
Cisco Bigdata	0.0101

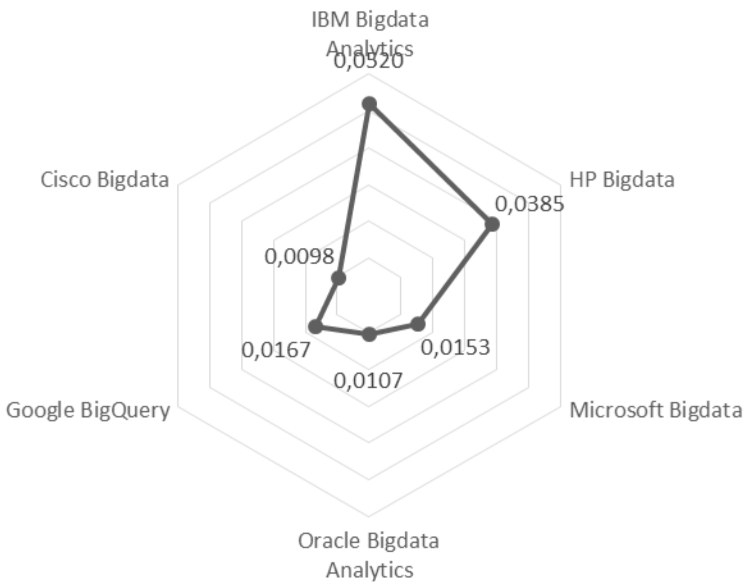

Fig. 5. Diagram of alternative analytical platforms weights

The recommendation for the selection of "IBM Bigdata Analytics" analytical platform are given on the basis of the above mentioned calculations, because investigated alternative has the greatest weight and fits for analytical processing of urban data sets.

5 CONCLUSION

The list of information technologies for the smart cities, which includes: Internet of Things, Fog Computing, Cloud Computing, Information Models, Intelligent Data Processing and Decision Support Systems, Mobile Technol-ogy, Geoinformation

Technologies, GRID technology, Confidential Communication and Data Protection Technologies is formed in this paper. It is offered to implement the selection of the best variant with the application of the method formed on the basis of hierarchies analysis. It is based on the procedure of pairwise comparisons of variants.

The method offered by the authors is applied for solution of the problem of efficient selection of the analytical platform for smart cities in the context of efficient creation for development information systems in Ternopil, that are carried out by the team of researchers from Ternopil National Technical University and National University "Lviv Polytechnic".

References

1. Soomro, K., et al.: Smart city big data analytics: An advanced review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9.5 (2019): e1319.
2. Gupta, A., Panagiotopoulos, P., Bowen, F.: An orchestration approach to smart city data ecosystems. Technological Forecasting and Social Change 153 (2020): 119929.
3. Ruhlandt, R., et al. Drivers of 'Data and Analytics' Utilization Within (Smart) Cities: A Multi-Method Approach. EPOC 2018 (2018).
4. Picardal, C., et al.: Bellevue Smart: Development and Integration of a Smart City. Journal-American Water Works Association 112.2, 28–37 (2020).
5. Zschörnig, T., et al.: A cloud-based Analytics-Platform for user-centric Internet of Things domains–Prototype and Performance Evaluation. Proceedings of the 53rd Hawaii International Conference on System Sciences (2020).
6. Arun, A.: Architecting IOT for Smart Cities. Smart Cities in Application. Springer, Cham, 141–152 (2020).
7. Tabachyshyn, D., Kunanets, N., Karpinski, M., Duda, O., Matsiuk, O.: Information Systems for Processes Maintenance in Socio-communication and Resource Networks of the Smart Cities. In: Advances in Intelligent Systems and Computing III, vol. 871, 192–205 (2019).
8. Duda, O., Kunanets, N., Matsiuk, O., Pasichnyk, V.: Cloud-based IT Infrastructure for "Smart City" Projects. In: Dependable IoT for Human and Industry: Modeling, Architecting, Implementation. River Publishers, 389–410 (2018).
9. IBM – Hadoop – Big data analytics, https://www.ibm.com/analytics/hadoop/big-data-analytics, last accessed 2020/07/01.
10. Hewlett Packard Enterprise DATA ANALYTICS SOLUTIONS, https://www.hpe.com/emea_europe/en/solutions/data-analytics.html, last accessed 2020/07/01.
11. Big Data and Advanced Analytics Solutions | Microsoft Azure, https://azure.microsoft.com/en-us/solutions/big-data/, last accessed 2020/07/01.
12. Big Data Products – Big Data Analytics|Oracle, https://www.oracle.com/big-data/products.html, last accessed 2020/07/01.
13. BigQuery Data Analytics, https://cloud.google.com/bigquery/, last accessed 2020/04/03.
14. Big Data and Analytics – Cisco, https://www.cisco.com/c/en/us/solutions/data-center-virtualization/big-data/index.html, last accessed 2020/07/01.