Nowoczesne Systemy Zarządzania
Zeszyt 12 (2017), nr 2 (kwiecień-czerwiec)
ISSN 1896-9380, s. 119-132

Instytut Organizacji i Zarządzania
Wydział Cybernetyki
Wojskowa Akademia Techniczna
w Warszawie

Modern Management Systems
Volume 12 (2017), No. 2 (April-June)
ISSN 1896-9380, pp. 119-132

Institute of Organization and Management
Faculty of Cybernetics
Military University of Technology

"Big-Data" systems in improving modern organizations

Public speeches of senior staff of the army as the safety component cultural

Piotr Zaskórski
Military University of Technology, Faculty of Cybernetics
piotr.zaskorski@wat.edu.pl

Wojciech Zaskórski
Military University of Technology, Faculty of Cybernetics
piotr.zaskorski@wat.edu.pl

Abstract: An attempt has been made to specify a determinant of effectiveness of a modern organization in this article. One of the determinants is information effectiveness. Gathering various types of information has been the basis for enhancing modern operating entities. Very large information resources can be subject to multifaceted exploration and discovery of knowledge. Extracting/generating maximum amount of knowledge from systems which gather polymorphic information of very high volume (so called "Big-Data" systems) in the form of numerical data, text files, images (video) is possible due to various mechanisms of multifaceted analysis. It is especially relevant in modern, distributed, process oriented and multi-entity organizations which require integrated operation for execution of complex, common processes, drawing from a uniform knowledge base.
Keywords: Big-Data, organization, system, effectiveness.

„Any sufficiently advanced technology is indistinguishable from magic"
Arthur C. Clarke, author of 2001: *A Space Odyssey*

Introduction

Gathering information, of all shape or form, becomes challenging for modern organizations. Information becomes a strategic resource, requiring not only safeguarding, but most of all, suitable explorability across multiple aspects. Extracting knowledge to the maximum possible extent becomes a target for BI-class systems, but primarily for those system which gather multifaceted data of extreme proportions, so called "Big-Data" systems, which possess robust functionality of numerical data analysis, text files and images (videos).

A phenomenon of informational asymmetry can be observed in modern distributed organizations (Zaskórski, 2005, pp. 201-244), which can be a attributed to restricted access to their own or environmental data. This effect can be applied to the disproportion between the abilities and the needs of a particular organization, and either conscious or unwitting access restrictions to particular data by other users (resource holders). Informational asymmetry can indicate a state, in which we do not possess sufficient information to implement a particular undertaking, not solely because of restricted information access, but also as a result of inefficient usage of information potential. Such phenomenon may also be a result of operations of business environment or operations of the opponent in military systems. Most often though, it is a result of restrictions, or even lack of know-how when dealing with extracting knowledge from information systems one possesses. This in turn, leads to loss of information advantage by lowering one's own potential, even leading to the handing over of initiative to the competition.

Information resources and their accessibility distinguish some organizations from others and become the requisite for effective management process, ranging from planning based on knowledge and knowledge extraction to supervising and coordinating of efforts, and objective result evaluation. Thus, limiting information asymmetry by thorough usage of available information resources can increase effectiveness and accuracy of decision-making processes.

Negative effects of information asymmetry cannot be eliminated but can be limited by the use of better and broader information access (i.e. "cloud" computing[1]), or by advanced data analysis techniques. In the long term, the dynamic developments in information technology in business and its strategic role require engagement of persons[2] who not only possess thorough knowledge of broadly defined information technology solutions, but also knowledge of management and economics.

1 So called "cloud" computing as a variation of virtual information service (usage of not only outside computing infrastructure but also of applications and information gathered therein, stored in non-easily identifiable places, but logically available).

2 Chief Information Officer (CIO) – IT manager in large organizations.

1. Attributes of modern organizations

A modern organization (company, enterprise) is most of all a compound process system, characterized by elasticity and dexterity concerning its adaptation capabilities in changing environment. Dexterity, for this purpose, means operating capability in the context of on-going, effective reactions to change. Dexterity is de facto preconditioned by the elasticity of the system and thus the capability of adapting organizational structure and extracting from it those components which can attribute to reorienting tasks and those who perform them according to competence and availability of various resources.

Elasticity of a system process cannot infringe on its function and task-oriented integrity. Hence, singular tasks and goal completion should be integrated with goals of the whole organization. The higher the level of integration of minute goals with global ones, the higher the level of synergy of the whole system. Integration – particularly, integration of information – does not rely on mindfulness of cohesive operations within a changing environment and elasticity of organization behavior alone, but also about maximizing utilization of own potential. Consequently information integrity is bound to the accessibility of data resources (which are uniform and up to date), to various authorized users. Accordingly, so called informativeness and decisiveness of operation systems (of modern organizations, business or otherwise) is conditional upon the content, orderliness and dedicated access (unrestricted in time and space) to integrated information resources. Elasticity and dexterity of operations is possible on the condition that, at any time or place, authorized operators have access to identical information.

Nowadays, a distribution of task operators and of various types of tangible and intangible assets becomes one of the attributes of a modern organization. This coerces the flattening of structures and imposes orientation towards operational structures, which in turn require a dynamic and elastic access to information, determining their efficient functioning (fig. 1). Value of a modern organization is calculated by the capability of creating multi-entity structures, scattered across territories but integrated by common, dedicated information (knowledge base). Cooperation of multiple entities according to criteria of competence (network-centric structures) is often implied by a so called value chain – a relation between efforts and effects resulting from experience "curve".

Creation of multi-entity organizations is attributable to virtualization of operations. This indicates that modern organizations may constitute in cyberspace on the basis of information about common goals and the capability of mutual complementation of tasks, according to the criteria of optimizing value towards a common goal.

Multi-entity operations require common criteria of assessment of individual entities in the virtual model. Hence, the necessity of emphasizing hypertext structures, which rely on access to common knowledgebase, among those focusing on a common value system. Individual entities cooperating in virtual space should thus be treated as fractals of a singular unity, created for implementation of uniformly defined processes which are described by resource usage and operation result data collection. This requires recording and gathering of great quantities of transaction data and historical data, to raise information based decision-making effectiveness. Every flavor of data may be of use. No information of any sort is disposable. Accordingly, there is a need of organizing and maintaining so "Big-Data" systems, which gather any amount of data of various types that may be of use at an indeterminate time.

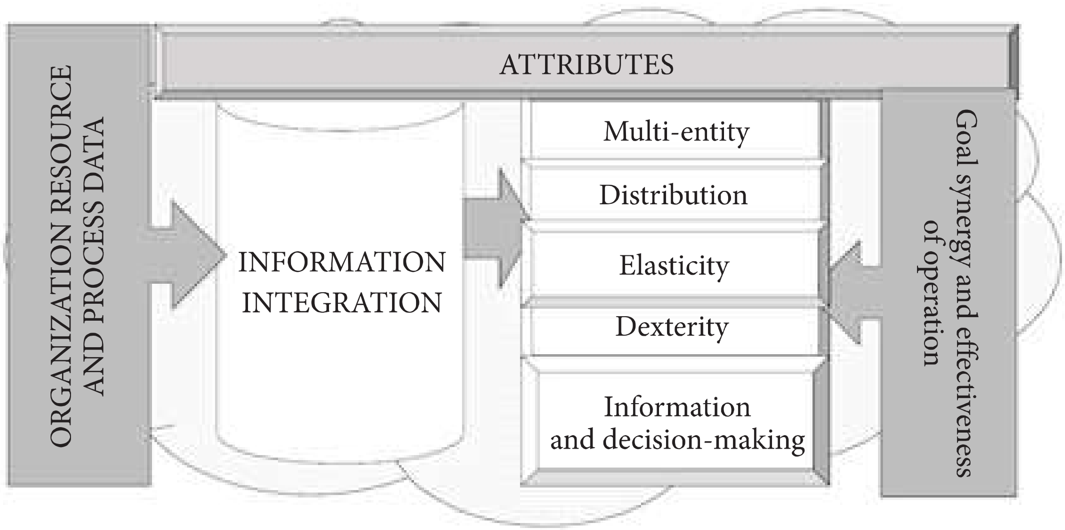

Fig. 1. Functional environment and attributes of a modern organization
Source: self-study based on: Zaskórski, 2012, pp. 24-33

2. Functionality of "Big-Data" systems

"Big-Data" systems are constructs of very large amounts of various data in the form of structured files (so called columnar databases or other ordered data structures), as well as flat text files or static or dynamic images (as well as videos). A valuable analytical expansion of structured data, stored in traditional databases, are OLAP--class systems and data warehouses, as immutable collections of data. Operations of data aggregation can be performed on such collections. Furthermore, a broad exploration of such data may be performed to procure knowledge using various models (algorithms) by use of Data-Mining (Olszak, Ziemba, 2003; Zaskórski, 2012, pp. 193-258). Advanced algorithms may lead to spectacular and detailed conclusions (i.e. identification of metadata describing a particular person). An example of the aforementioned is the research carried out by the Massachusetts Institute of Technology[3]. This may mean that it is not necessary to know specific identity records (last name, address, social security number etc.) to pinpoint (conclude, on the basis of systematic monitoring of transaction data) preferences and other important information about residence, employment etc. Such data does not have to be procured by illegal means, as it is voluntarily given away during electronic transactions. The problem is to gather the data and run appropriate analytical models (i.e. association, aggregation, dispersion, time correlation). In a sense, it is evidence that in many situations, one is able to procure knowledge from ordinary information flow. Hence, the difficulty keeping pace with the developments, which may be used to limit informational asymmetry or to stimulate it by altering routine operations. Often investments are made in various security systems (keys), while repeated, routine actions, registered over a longer period of time may be enough to gather sensitive information about a single user, more so about a whole organization. This is the power, and threat, offered by "Big-Data" systems.

"Big-Data" systems and correlated artificial intelligence methods expose cause--and-effect relationships in a multi-dimensional environment. Those systems view phenomena and behaviors well above three-dimensional reality. "Big-Data" systems are able to analyze a collection of various actions or even entirety of all actions in all possible dimensions and predict our behavior on the basis of multi-criteria analysis, conducted simultaneously and paralelly. Resources of data and metadata passed over to public institutions (banks, network operators, insurance providers and providers of various services) can predict (with probability reaching 90%) subsequent operations of an entity or even whole processes, using the so called inference engine.

One of the main premises, associated with "Big-Data" systems is the gathering of all possible data (fig. 2) of target sizes up to hundreds of terabytes of even petabytes. No data is worthless and its preservation may considerably enrich the spectrum of analysis. Upon gathering, the data is stored in an arbitrary database, ready to be explored according to necessity. Own resources as well as competition (opponent) resources become important assets of organizations, further enhanced by functionality of "Big-Data" systems. Concurrently, as mentioned earlier, information asymmetry, enforced by a particular entity towards its environment is perceived as an asset. This means that such an entity will strive to impair the value of its exposed data towards its competitive environment (including "Big-Data" systems of competition) by:

- implementing e-mail encryption,
- frequent changes to e-mail addresses while limiting personal data profiles on the Internet, which allow identification of persons or entities,

[3] MIT scientists analyzed "(...) anonymous credit card transactions of 1,1 million people. Using a new analytic formula, they needed only four bits of derivative information – metadata, such as the location or time span – to recognize individual consumer behavior of 90% of participants, even when data was bereft of names, bank account numbers or other obvious identifying attributes (…)" (Source: zipnews.pl/2015013113768-anonimowe-transakcje-karta).

- selecting metadata, which are not unique to the particular organization when compared to other users,
- delaying or relocating data on the hard drive, because every computer connected to the Internet should be considered insecure and requiring safeguarding,
- limiting access to social media, which can easily intermingle with private resources,
- limiting emission of data bearing recurrent metadata, including credit card transactions.

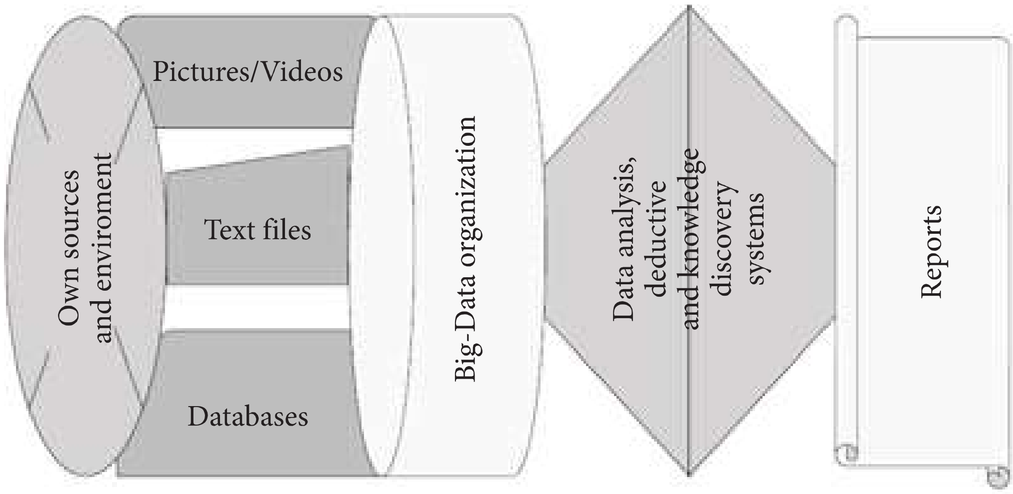

Fig. 2. General idea of "Big-Data" systems
Source: self-study based on: www.microsoft.com/en-us/cloud-platform/data-warehouse-big-data

The quantity of information stored in computer systems is rising rapidly. This is further exacerbated by the use of mobile devices, where unstructured data (video, audio, images – with no possibility of using database relation mechanisms) is gathered. Additionaly, more and more often, ad-hoc data work is required in business practice. Data for this work is procured from unstructured heterogenic sources. Effective analysis of such data may decisively determine dominance over competition. It is important to note, that it has become expected to process such large volumes of data in real-time with real-time result sharing. This causes ordinary data warehouses using structured data models must be replaced with "Big-Data" solutions, which enable:

- effective processing of large volumes of data, using modern columnar databases or integrated solutions[4] (appliance),

[4] These solutions include IBM Neteeza, Sybase IQ, HP Vertica or Oracle Exadata.

- integration of heterogeneous information within a single tool, in real time[5] (data streaming)
- analysis of unstructured data using advanced technologies of video streaming analysis, voice transcription and tools for redefining ETL data for Business Intelligence systems,
- interactive visualization, visual analysis and prognosis report creation[6].

Implementation of such systems and their use as virtual services allows for better understanding of available information and quicker access to such information. Worth noting is the expansion of capabilities regarding multi-aspect analysis, as well as reduction of required labor for advanced, professional data and information analysis, along with deduction and knowledge gathering processes. Of particular importance are "Big-Data" systems based on "cloud" computing services (the Internet), which provide functional possibilities[7], such as:

- storing large amounts of data (volume),
- maintaining high dynamic of data (velocity),
- maintaining high variety of data (variety),
- assessment (concurrent verification, value) of data.

In 2012, GartnerGroup pointed out that "Big-Data" consists not only of large volumes of information of high dynamic and variety, but also new, dedicated models of analysis and processing of such volumes, with effective and potent support of decision-making processes, by discovering knowledge and new phenomena (deduction) using optimization of action/processes. Furthermore, specialist personnel (engineers) engaged with "Big-Data" processing should understand business goals of the enterprise and be aware of crucial flow and analysis of data pertaining to effective execution of said goals. Such experts are, most of all, responsible for the operation of gathering and processing of source data and its evaluation in the context of usefulness of new sources of information. They should also be responsible for designing, storing and processing of the incoming data stream.

3. OLAP systems as a component system of organization and "Big-Data" exploration

One of the solutions for analysis of large amounts of unstructured, historical data is an OLAP (Januszewski, 2008; Olszak, Ziemba, 2003) system. The technological

---

5 IBM Data Capture, Verint, HP Authonomy among others, which allow for Internet and social network (Facebook) monitoring.

6 Solutions alike IBM Cognos, Data Watch and QlikView.

7 According to GartnerGroup it is a 4V model (4W in the Polish version), which is; first use your own data, deduct and skillfully analyze by experts, enhance own data with market information using dictionaries and reference databases; verify data, hypotheses and conclusions.

tools related to this class of systems are data warehouses[8], main purpose of which is to support analytic decision-making processes using specialized analytical procedures. Data warehouses are a collection of large amounts of data, stemming from heterogenic sources of OTLP (Januszewski, 2011; Zaskórski, 2012) class. Data warehouses (fig. 3) are large databases, subject to aggregation using OLAP functions and knowledge "extraction" by DM ("Data Mining"). Decision-making processes (DSS[9] systems – Decision Support Systems) may be assisted in this manner using retrospective (Zaskórski, Zaskórski, 2005) and prospective models including event parallels and situation development trend discovery.

Preparation of data and the implementation of analytical/decision-making tools is a complex and time-consuming process, which requires systematic collection and integration of geographically and temporally scattered data, while unifying its form and content. The dynamics of modern reality and challenges facing technological development coerce changes in the operation model of a modern manager, by multifaceted data analysis. Data exploration systems are enhancing his capabilities by allowing access to synthetic and ordered information, reflecting the dynamics and essence of occurrences and processes within the organization as well as its environment. OLAP class systems are treated as systems with "built-in BI intelligence"[10].

Operations related to gathering and processing information gain unique significance and ensure a more effective use of the potential of an organization, all the while reducing information asymmetry. Pursuing enhanced effectiveness of operations requires the invocation of knowledge and experience of the organization and its environment. BI class systems are characterized by a high level of variety and a high number of potential users. Their main components are the aforementioned data warehouses (including so called data stores) integrating distributed business environment and maintaining a thematic, integrated, growing (but permanent) data collection.

Selection of data and its substantial correctness and usefulness are the basis for its usability in analytical decision-making processes. Data warehouse mechanisms can reflect the method of current/operational data collection and analysis for a selected time period with up to date results and historical/retrospective data pertaining to a longer time.

---

8 Data warehouse system according to a dictionary by dr. Z. Ryznar is a:
- Database set, created around a specific architecture;
- Metadata repository with a management system located within the warehouse;
- Warehouse management software;
- Multifaceted database analysis tools (OLAP) and software for intelligent data-mining.

9 OLAP/DM systems may also support knowledge base creation in expert systems and AI systems (model databases).

10 Can be widely used in administration, education, military etc.

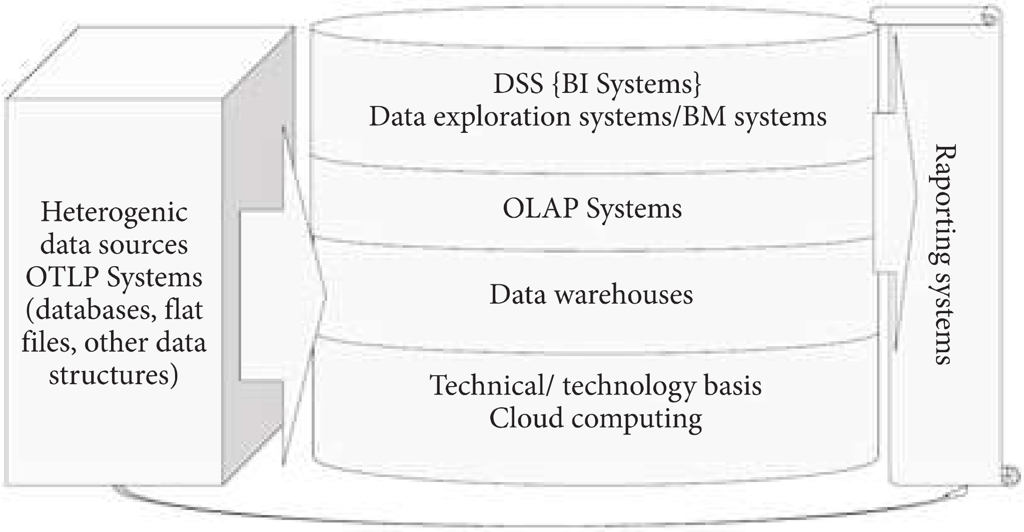

Fig. 3. General architecture of data analysis and exploration
Source: self-study

Analytic processing is a multi-dimensional data analysis initiated by the business/end user, based on complex mechanisms of deduction and reporting (Gorawski, 2003). Such mechanisms are a collection of tools which enable analysts and managers to quickly and interactively access data and visual information, which in turn can determine situation (including prognosis) of an organization, based on:

- analyses based on complex search algorithms of data collections aggregated (in so called cubes/aggregations) according to set measurements,
- interactive reporting dedicated to a wide spectrum of users, including generating answers to questions (assessment-prospective models).

Depending on the problem and the variation and number of dimensions of the conducted analysis – the subject of the analysis may be related to various measurements. Actual information needs arising from the needs and competences of the members of the process will profile the scope and form of the analysis. Aggregations located in the cube may be subject to further operations related to:

- winding (unwinding) consisting of moving to a higher level of hierarchy of the dimension (operation opposite to unwinding);
- rotation (turning around) which changes the perspective of data visualization and enabling their presentation across any dimension;
- selection, projection and cropping, which indicate a narrowing of the scope of the subject of analysis by specified value (selection) or size (projection);
- ranking, consisting of arranging elements of the dimension by value (dimensions).

We can generally conclude that in management practice, it is expected of OLAP systems to search for answers to routine questions. More and more often though, there are searches for rules, patterns, trends and relationships which are basis for generating knowledge about occurrences in an origination or its environment (the user may often be unaware of the occurrence of particular relationships or regularity). In such cases we deal with a process of searching for hidden knowledge, which could be applied to enhance effectiveness of operations. Here we relate to models of advanced exploration (fig. 4) of large subsets of data to discover rules and patterns which are useful in decision-making and prognosis of future states and behaviors. It is worth noting that the process of data exploration should always be a creative process, where the human element cannot be underrated.

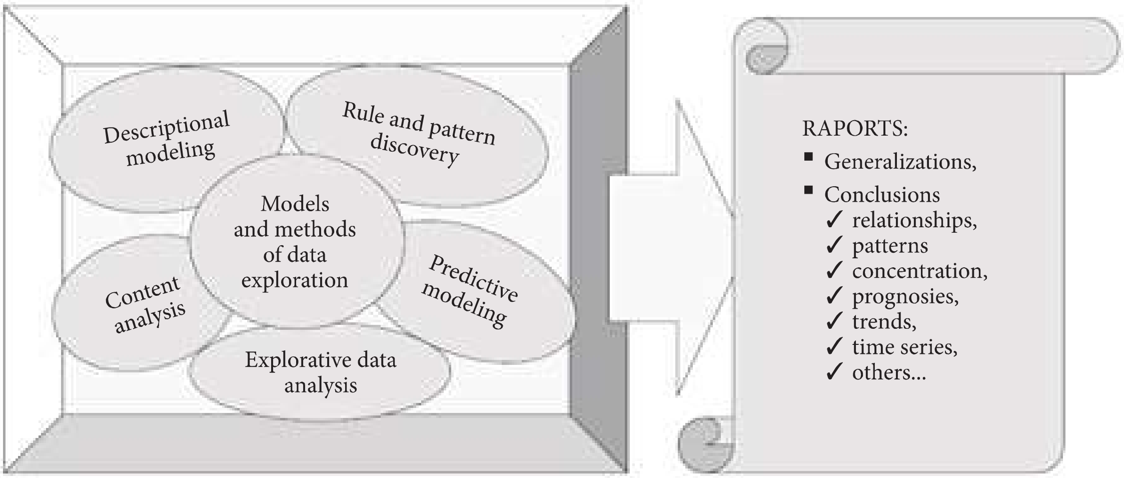

Fig. 4. Models and methods of data exploration
Source: self-study on the basis of: Hand, Mannila, Smyth, 2005, pp. 279-375

Advance exploration models[11] of data are a logical component supplementing the functionality of data warehouses and OLAP-class systems. An important part of this process is the awareness of the analytical-systemic and business prerequisites for the undertaken research goals. A detailed selection process of data (attributes) is an imperative. Models and techniques of advanced data exploration are based on various methods[12] (fig. 4) of Data Mining, usually related to formulating hypotheses and their verification[13]. Such models reflect the complexity of the process of advanced data exploration and are related to:

- explorative analysis of data, implied as finding the relationships between data/values (according to set measurements) using visualization;
- descriptive modeling, based on the characteristics of all available data using cluster analysis and models of relationships between data;

[11] (CRISP-DM-1996 r.), Cross-Industry Standard Process for Data Mining standard.
[12] Handa, Mannili and Smyth; five primary methods of data mining.
[13] Data Mining can usually be associated with the search for the answer to the question „I do not know, what I do not know".

- predictive modeling, which strives to predict the relationship of one parameter on the basis of known values for other variables within the same class of aggregation (regression methods[14], classification[15];
- content searches in the aspect of available data related to the discovered pattern;
- discovering patterns and rules – mentioned earlier – usually concerning a chosen aspect by use of so called associative rules[16].

OLAP systems, integrated with the implementation of advanced data exploration models can generate knowledge about certain phenomena and relationships. Thus Data Mining systems convert certain resources into knowledge, which is useful in the context of executed operations. Typical OLAP tools are useful for entities who "know[17] what they don't know", while DM tools help in cases where the entity "doesn't know what it doesn't know" (Januszewski, 2011).

4. Application of "Big-Data" in decision-making systems

As stated earlier, gathering of large amounts of data and its multi-aspect exploration may be an significant factor in limiting information asymmetry, which is discovering knowledge others do not possess – thus outstripping potential competition (opponent). One of the primary challenges in managing a contemporary organization (especially a scattered one) is its information integration and providing for the possibility analytically process available resources on-line and "discover" knowledge. Use of such knowledge can substantially influence effectiveness of the whole organization within a changing environment, taking into particular consideration, reliable planning.

Modern organizations rely on processing structures, which use network models (including network, hypertext), which notably increase the value of resources of "Big-Data" type information, with all data available through a WWW platform. Such a coordination of efforts (dynamic monitoring of process execution and their correction, based on historically objective assessment) lifts many limitations in access to data and its dynamic use. Using "Big-Data" systems along with all active functions (models and tools of discovering knowledge from structured, unstructured, text, images etc., including DM models) can directly influence decision accuracy

[14] An extrapolation of trends on the basis of gathered data, incl. Kenyes, Brown, Holt and Winters models, which allow for specifying a quantifiable variable for a future occurrence with varying accuracy.

[15] Assigning data to an appropriate class on the basis of its attributes and specifying and differentiating classes i.e. using decision tree techniques, nearest neighbor algorithm or neural networks.

[16] i.e. fuzzy set logic, based on quantitative measurement of probability (if *A* is *B* with probability of *p*).

[17] Needed aggregations according to specified measurements and for whom and when.

and in turn the effectiveness and the functioning of modern organizations. Systems of this class may be particularly significant in the context of the function diagnosis, prognosis and regulation, the results of which are useful in the planning process.

Shared information resources are a component which integrates information for operators of particular processes and enhances consistency of operation within a geographically scattered environment. Gathering and systematizing of own data, information and knowledge about processes occurring within the organization, along with the knowledge possessed by participating virtual operators and experts may determine the model of cohesive decision-making by multiple entities in the scope of:

- planning processes and decision-making processes in multiple time spans,
- monitoring, control and assessment of quality and pinpointing sources of possible disturbances.

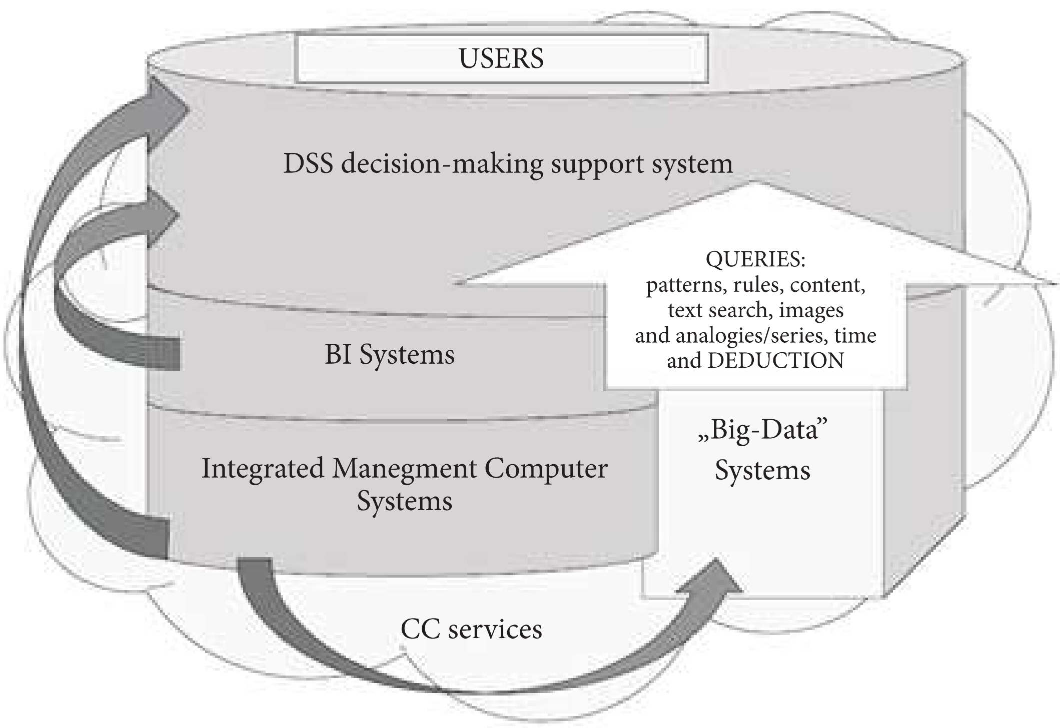

Fig. 5. “Big-Data” systems as a DSS platform

Source: self-study on the basis of: Hand, Mannila, Smyth, 2005, pp. 483-545

OLAP/DM systems currently in operation are a crucial component of an organization which learns, and uses not only proprietary information, but also information from the environment (including cooperating entities or competition/ opponent). Such resources, in an integrated environment (i.e. computing cloud) enhance operational capabilities of an organization. This is reflected in modern structures, which possess broad decision-making competence among process owners, independence and responsibility of employees and teamwork models during execution of each process.

Information integration and access to multi-aspect analyses – uniformly and homogenously ordered information resources – promote effective local and global operation. Implementation scope of cross-engineering is a direct derivative of:

- standardized operation procedures for execution of common processes;
- unifying technological information standards.

The concept of a modern distributed, but process oriented, organization is nowadays mostly a function of informational efficiency. The human element supported by knowledge “discovery” tools enhances the capabilities of a modern organization in real world operations. It is prerequisite by effectiveness of the usage of such tools and the awareness level of process operators, concerning usefulness and power of those tools. Decision-making processes and DSS-class[18] systems as a whole are determined by universal utilization of common (often dedicated) resources, upon the establishment of commonly accepted information technology standards of gathering, saving and storing data.

Conclusions

Modern organizations are organizations which “learn” – gather specific information and perfect their operations on the basis of both personal and outside experience. Data and information concerning organization operations within the background of the behavior of the environment can present an objective image of decision-making determinants. DSS systems are limited in effectiveness by the quality of data and information and representative nature of knowledge gathering and generation mechanisms. “Big-Data” systems directly influence this quality by their capability to select resources for particular tasks and decision-making processes. The power of “Big-Data” systems is stimulated by the multifaceted results of real world observation.

BIBLIOGRAPHY

[1] Gorawski M., 2003, *Ocena efektywności architektur OLAP*, [w:] Z. Huzar, Z. Mazur (red.), *Problemy i metody inżynierii oprogramowania*, Wydawnictwa Naukowo-Techniczne, Warszawa.

[2] Hand D., Mannila H., Smyth P., 2005, *Eksploracja danych*, Wydawnictwa Naukowo-Techniczne, Warszawa.

[3] Januszewski A., 2011, *Funkcjonalność informatycznych systemów zarządzania. T. I. Zintegrowane systemy transakcyjne*, PWN, Warszawa.

[4] Januszewski A., 2008, *Funkcjonalność informatycznych systemów zarządzania. T. II. Systemy Business Intelligence*, PWN, Warszawa.

[18] Decision-making assistance on the basis of analogy, generating knowledge and deduction.

[5] Larose D.T., 2006, *Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych*, PWN, Warszawa.

[6] Olszak C., Ziemba E., 2003, *Systemy Business Intelligence narzędziem wspomagającym pracę menedżerów*, [w:] B.F. Kubiak, A. Korowicki (red.), *Human-Computer Interaction*, Uniwersytet Gdański, Gdańsk.

[7] Van Ufford D.Q., 2002, *Business Intelligence – The Umbrella Term*, Amsterdam (electronic version).

[8] Zaskórski P., 2005, *Strategie informacyjne w zarządzaniu organizacjami gospodarczymi*, Wojskowa Akademia Techniczna, Warszawa.

[9] Zaskórski P., 2012, *Asymetria informacyjna w zarządzaniu procesami*, Wojskowa Akademia Techniczna, Warszawa.

[10] Zaskórski P., 2012, *Wirtualizacja organizacji w „chmurze" obliczeniowej*, „Ekonomika i Organizacja Przedsiębiorstwa", Nr 3.

[11] Zaskórski P., 2013, *Systemy klasy BI platformą współczesnych organizacji*, [w:] *Współczesne zarządzanie – różnorodność problemów i sposobów ich rozwiązywania. Część 4. W objęciach nowoczesnej technologii*, Wojskowa Akademia Techniczna, Warszawa.

[12] Zaskórski W., Zaskórski P., 2005, *Retrospective models in operations planning*, VII NATO Regional Conference on Military Communications and Information Systems, Zegrze.