High-level Architecture For a Digital Oilfield: Features of the Transition to Data-driven Decision Management

Aleksei Savelev
Division for Information Technology
Tomsk Polytechnic University
Tomsk, Russia
sava@tpu.ru

Abstract— **This paper is devoted to the design of a distributed heterogeneous data warehouse of a digital oilfield. With increasing amounts of data collected from intelligent controllers and sensors, the lack of mechanisms for combining data from different sources and providing them to consumers affect the overall management efficiency. In addition, without it is impossible making the next logical step – the effective application of intelligent analysis methods. The paper describes the high-level architecture, as well as subsystems of operational management and decision support. The presented data are intermediate results of the project "Digital oilfield heterogeneous distributed data warehouse for informational support of decision-making processes".**

Keywords— *business analytics, data mining, decision-making, digital oilfield.*

I. INTRODUCTION

Today, due to the development of digital oilfield (DOF) technologies, there are virtually no production and management tasks in the field of oil production, the solution of which would not be automated to one degree or another. At the same time, there is a growing interest in the study and implementation of data mining technologies to support decision-making and improve the overall management processes efficiency. However, the effectiveness of implementing these depends on the quality of the previous automation and control stages solutions. However, the distribution, implementation, and use of individual information solutions often occur outside the framework of a rational automation strategy. Due to the complexity of the digital oilfield, due to the nature and composition of the appropriate management tasks, the lack of integration between the multitudes of existing software solutions prevents the formation of single information space of the enterprise. Thus, the effective use of data analysis tools and the consequent transition to data-based management are complicated. Separately, we should note that the oil production companies recognize the data integration problem. Several software solutions automate the main tasks of oilfield development management. The main applied field of research in the digital oilfield area is solving problems of forecasting and modeling the state of development, including the current monitoring of oil well equipment [1-5].

Authors of papers [6-7] considered the existing groups of Energetics consortium international standards as a way to ensure the uniformity of data presentation. The advantage of these standards is their data schemes, covering a wide range of oil industry elements. With an increase in the volume of data generated by the industry-recognized by most authors, various ways are offered to increase the efficiency of working with them. The authors of [8–9] proposed the use of cloud technologies subject to the availability of reliable communication channels with high bandwidth. The use of Big Data technologies, as well as machine learning and artificial intelligence concerning specific problems and tasks of the oil industry, are discussed in [2, 4, 10-14].

Despite the progress in this research area of, there are several unsolved problems, including data integration, knowledge extraction from the existing information volume, complexity, and efficiency of processing the available information [15]. In other words, there is no unified approach to storing and managing large amounts of data from heterogeneous, distributed sources and then extracting knowledge from them to support decision making and creating appropriate analytical problem-oriented algorithms and systems.

Summarizing the above, we can identify the following most pressing problems in the application of data mining and artificial intelligence methods in oil field development management:

- The lack of uniform software solutions for the digital oilfield distributed heterogeneous data sources integration.

- The volume of data collected in the field and the mining process has moved into the Big Data category, the effective use of which in the management process requires the development of specialized software tools.

As a solution to these problems, we propose heterogeneous distributed data warehouse creation to unify the processes of interaction between the components of the digital oilfield infrastructure, ensuring the unity of the data collection, processing and storing processes, as well as intellectualizing the oilfield development management.

II. DIGITAL OILFIELD INFRASTRUCTURE MODELING

Within the framework of the research, we created a software-hardware stand for simulating the infrastructure of a digital oilfield. A hardware-software stand is a system of three main subsystems: data generation, data aggregation, data processing, and storage. The stand includes the following data sources:

- Sources of low-level data collected from cluster controllers and sensors. Data source include four main data sets (5560, 3368, 15040 and 22144 records) both in the two most commonly used formats in the industry – wellsite information transfer standard markup language (WITSML) and log ASCII standard (LAS).

- Instrumentation and automation equipment database. Information is presented as a Microsoft SQL database and contains records of 4000 models of equipment, as well as of the regulated time for maintenance and repair work.

- Reference and reporting information stored as .doc, .pdf files. This set of files is the "passport" of the oil field and contains an extensive list of textual, tabular, and graphical information describing the oil fields starting from the exploration stages.

Low-level data sources support an improved drilling data transfer protocol – Energestics transfer protocol (ETP), which implements a more reliable data transfer channel at the program level. The stand generally allows studying the peculiarities of the technical interaction between DOF information infrastructure components. The stand is applicable in the tasks of specialists training, as it allows to visually demonstrating the interaction of different software and technical solutions within the infrastructure of one oil company, including the associated features of synchronization, integration and data import/export. Besides, the stand provides flexible options for setting up a test environment for testing the interaction with elements of the external infrastructure, load modeling, system testing.

III. ATTRIBUTE DEPENDENCY MODEL

The problem of efficient use of heterogeneous and distributed sources of information is the difficulty of ensuring timely access to the full list of data, the result of processing and analysis of which is necessary for management processes.

As already noted, one of the problem solutions is the creation of systems and integration tools that provide a unified presentation of data. The amount of information generated by the elements of the digital oilfield infrastructure makes it virtually impossible to integrate at the physical level. Integration at the logical level requires a single, global scheme that allows combining different ways of defining dependencies between data, as well as implement different strategies for finding solutions.

To form a unified scheme for presenting digital oilfield data, we propose to use the attribute dependency model, which is part of the object-oriented system analysis methodology. The network of functional dependencies describes the links between data located in various sources. The network is a non-strict hierarchy, which may have several roots. The attribute dependency model is used to determine the values of some

attributes based on the values of other attributes. There are two main classes of tasks to be solved using the model:

- The interpretation task. The initial data are the values of the primary attributes. It is necessary to find the values of some target attributes that correspond to the given source data. This type includes the tasks of medical diagnostics, diagnostics of technical equipment, forecasting tasks. The usage of direct output methods can solve such problem.

- The task of finding a feasible solution (inverse problem). The source data is the specific (specified) value of a non-primary attribute. It is necessary to find the corresponding values of the primary attributes. This type includes the tasks of planning, target management. To solve the inverse problem, use methods of reverse output.

The task of finding the optimal solution can be solved in two ways. In one way, as a series of solutions to the problems of interpretation for various alternative variants defined by multiple combinations of primary attributes. Alternatively, as a series of solutions to the problem of finding feasible solutions for different values of the efficiency criterion and target constraints. Methods for solving the optimization problem use the techniques of direct and inverse derivation, complementing them with procedures for generating variants and checking restrictions. Fig. 1 describes an example of the attribute dependency model structure.
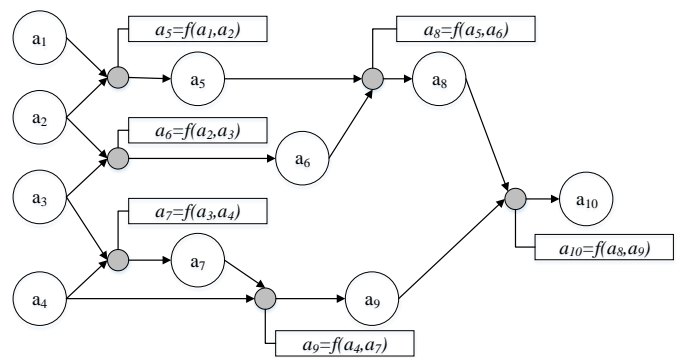

Fig. 1. An example of the attribute dependency model structure

Thus, the attribute dependency model makes it possible, first, to create a diagram of data interconnections stored in various sources. Secondly, the meta description of model attributes — data types — makes it possible to automate the tasks of searching and providing data access, as well as evaluating the feasibility of storing attribute values that are calculated (dependent on primary ones).

IV. HETEROGENEOUS DATA WAREHOUSE HIGH-LEVEL ARCHITECTURE

The study of the structure and types of data located in the different information sources described in section II, as well as scientific, technical and regulatory documentation analysis, made it possible to formulate the following main theses:

1. Technical interaction unification of heterogeneous data sources, subsystems, and software solutions today are the focus

of the DOF-research. There are several issues of data integration at the physical level. The WITSML and PRODML standards are considered as the basis for integration.

2. In other areas of activity that are not related to oil production, in particular, in the framework of the ATLAS experiment at CERN, working with heterogeneous sources is based on the development of solutions for the collection, storage, and analysis of metadata.

3. Solutions and approaches that integrate technical cooperation standardization tools for production systems and decision support tools based on the analysis of information from distributed heterogeneous sources have not been found within the framework of the DOF-concept.

Partly, we can explain the current situation by the following: the use of data mining tools for decision support purposes requires well-functioning solutions in the areas of data integration, including low-level ones, within a unified information environment.

Thus, we formulate the following basic principles for the development of the heterogeneous distributed data warehouse architecture for digital oilfield:

- The principle of technical interaction unification - the architecture should provide for the availability of physical integration solutions for low-level data obtained using industry-standard solutions and equipment. The integration of low-level data is necessary to provide information support for the processes of operational accounting and control over the state of oilfield development.

- Principle of single information space - the primary tool for the tasks of strategic management should be an appropriate decision support system, based on the

knowledge base, describing the relationship of all data from heterogeneous and distributed sources. The unity of the information space should be ensured by an appropriate logical model in combination with the principles of its change and maintenance.

Fig. 2-4 show the first version schemas of the heterogeneous distributed data warehouse architecture for digital oilfield consisting of operational management subsystem architecture (Fig. 2), decision support subsystem architectures (Fig. 3) and high-level representation of the heterogeneous distributed data warehouse architecture (Fig. 4).

TABLE I. FIG. 3 DATA FLOW DESCRIPTIONS

Data flow number	Data flow description
1	The list of target attributes that determine the desired state of the system
2	List of primary attributes defining the value of target attributes
3	Request for a list of data sources that store basic attribute values
4	Check request whether the search query is new
5	List of data sources and base attribute value extraction rules.
6	Extract the values of the primary attributes that determine the value of the target
7	List of base attribute values
8	Calculated values of target attributes and estimation of their errors

Thus, a mixed approach to the organization of a heterogeneous data warehouse, which consists in sharing the operational management subsystem and the decision subsystem, while maintaining the existing infrastructure of the digital field, improve the quality of decisions and implement the transition to a data-driven decision management model.
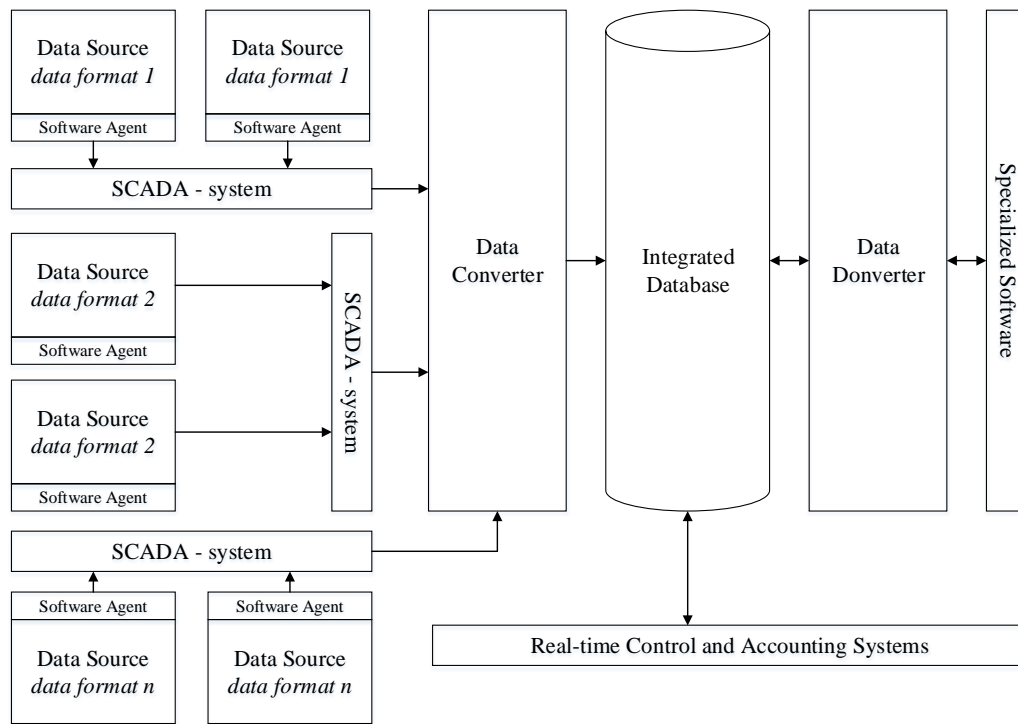
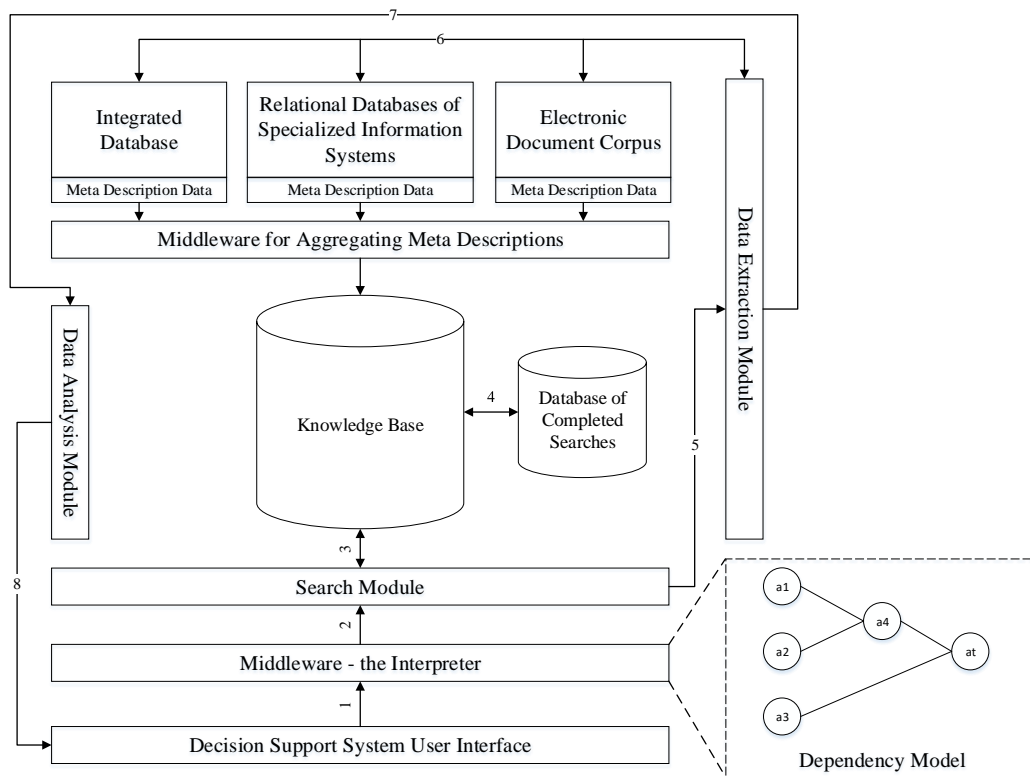Fig. 2. High-level architecture of the operational management subsystem

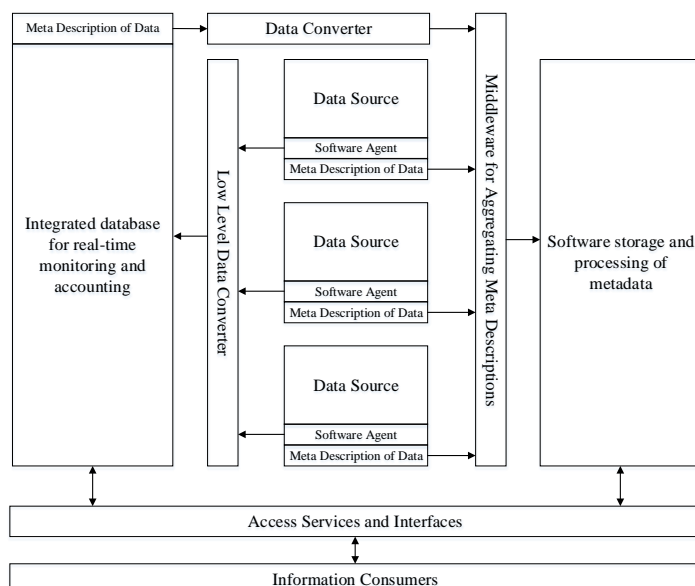Fig. 3. High-level architecture of the decision support subsystem

Fig. 4. High-level architecture of a heterogeneous distributed data warehouse for digital oilfields

V. Conclusion

The transition to a data-driven digital oilfield management model implies an analytical module using machine learning tools. The development of such a module is a complex task that requires a comprehensive integration of data sources to build the most accurate model of the analyzed processes. Because DOF data sources usually exist autonomously from the analytical module, the priority is to create a system for integrating data from heterogeneous sources into a single repository for quick and efficient search and retrieval of data needed by an expert. Within the framework of the current stage of the work, we propose a high-level architecture of a heterogeneous data warehouse for a digital oilfield, combining subsystems of operational management and decision support. The implementation of the software prototype according to the developed architecture will allow going over to the tasks of designing an analytical module for the data coming into the knowledge base using machine learning tools.

References

[1] H. Qin and Z. Han, "Stochastic Resource Allocation for Well Control With Digital Oil Field Infrastructure" IEEE Systems Journal, vol. 12, pp. 1295–1306, 2016.

[2] H. Hu, L. Fan and X. Guan, "Application on Crude Oil Output Forecasting Based on Gray Neural Network", Proceedings of the 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, pp. 533–537, 2017.

[3] F. Alenezi, and S. Mohaghegh, "A Data-Driven Smart Proxy Model for A Comprehensive Reservoir Simulation", Proceeding of 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT), 2016.

[4] A. Preveral, F. Trihoreau and N. Petit, "Geographically-Distributed Databases: A Big Data Technology for Production Analysis in the Oil&

Gas Industry", Proceedings of the SPE Intelligent Energy Conference & Exhibition Utrecht Netherlands, 2014.

[5] O. Patri, A. Panangadan, C. Chelmis, R. McKee and V. Prasanna, "Predicting Failures from Oilfield Sensor Data using Time Series Shapelets" Proceedings of the SPE Annual Technical Conference and Exhibition Amsterdam Netherlands, 2014.

[6] Q. Zhong and H. Du, "Design on the platform system of well-site information center based on WITSML", Proceedings of the 2015 International Industrial Informatics and Computer Engineering Conference, 2015.

[7] W. Mckenzie, F. Morandini F, P. Verney, L. Deny, J.-F. Rainaud, R. Eastick, G. Masters and M. Endres, "Integrating WITSML, PRODML, and RESQML Standards for Cross-Domain Workflows", SPE Intelligent Energy International Utrecht Netherlands, 2012.

[8] G. Zhao, W. Di and R. Wang, "Prospects and Potential Application of Cloud Drilling", IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 201–205, 2017.

[9] R. Alguliyev, M. Ramiz and M. Hajirahimova, "Big Data Integration Architectural Concepts for Oil and Gas Industry", Proceedings of the IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016.

[10] M. Brule, "The Data Reservoir: How Big Data Technologies Advanced Data Management and Analytics in E&P", Proceedings of theSPE Digital Energy Conference and Exhibition, 2015.

[11] D. Cameron, "Big Data in Exploration and Production: Silicon Snake-Oil, Magic Bullet, or Useful Tool?", Proceedings of the SPE Intelligent Energy Conference & Exhibition, 2014.

[12] R. David, S. Rao and N. Reddicharla, "Managing and Orchestrating Multi-Vendor Intelligent Oil Field Technology Environment to Enable Efficient Next Generation Production and Reservoir Management Workflows", Proceedings of the SPE Middle East Intelligent Oil and Gas Conference and Exhibition, 2015.

[13] R. Perrons and J. Jensen, "Data as an asset: What the oil and gas sector can learn from other industries about "Big Data", Energy Policy, vol. 81, pp. 117-121, 2015.

[14] P. Chanana, T. Son and U. Bhakne, "Emerging Technologies and Workflows in Digital Oil Field", Proceedings of the Offshore Technology Conference, 2016.

[15] J. Crompton, "The Digital Oil Field Hype Curve: A Current Assessment the Oil and Gas Industry's Digital Oil Field Program", Proceedings of the SPE Digital Energy Conference and Exhibition, 2015.