# Adaptation of the attribute dependency model for designing a heterogeneous digital oilfield data warehouse

**A O Savelev**[1]
[1] Tomsk Polytechnic University, 30 Lenina Ave., Tomsk, 634050, Russia

E-mail: sava@tpu.ru

**Abstract.** The development of digital oilfield technologies has led us to the fact that there are virtually no production and management tasks left without any automated software solutions. However, the distribution, implementation, and use of individual IT-solutions often occur outside the rational automation strategy. At the same time, due to the complexity of the management object, due to the nature and composition of the tasks being solved, the lack of integration between the multitude of existing software solutions prevents the enterprise single information space formation. Thus, with the actual availability of critical information, the operational and strategic management tasks solution occurs without considering it. As a solution, the article considers an approach to creating a heterogeneous data warehouse based on an attribute dependency model that describes the relationship and interdependencies of data stored in distributed and heterogeneous sources. The article describes the interim results of a project to develop the architecture and software heterogeneous distributed data warehouse prototype for digital oilfield infrastructure.

## 1. Introduction

A distinctive feature of modernity is a significant increase in the amount of information generated by various systems, environments, and communities. The oil industry is no exception. The collecting data volume and their structure suggest that the issues of information processing in the field of oil production have moved into the category of “big data” – data sets whose size exceeds the capabilities of traditional databases for recording, storing, managing and analyzing information.

With increasing amounts of data collected from intelligent controllers and sensors, the lack of integration data mechanisms from heterogeneous sources and providing the results to management level affects the overall decision-making efficiency.

Many authors address issues of improving oilfield development management efficiency using digital technologies. The main applied Digital Oilfield research topics is solving forecasting and modeling the state of oilfield development problems, including the current monitoring of oil well equipment [1-5]. Authors [6-7] consider the existing groups of the Energestics consortium international standards as a way of ensuring the uniformity of data integration and presentation. The advantage of the existing groups of standards is their data schemes, covering a wide range of oil industry data.

With an increase in the volume of data generated by the oil industry, recognized by most authors, various ways are offered to increase the efficiency of working with them. The authors of [8] propose to use cloud technologies, which is of interest if there are reliable communication channels with high bandwidth. The use of Big Data technologies, as well as data analysis, machine learning and artificial intelligence in relation to specific problems and tasks of the oil industry, is discussed in [2, 4, 9-14]. Despite the progress that exists in this research area, there are a number of unsolved problems [15], including data integration, knowledge extraction from the large information volume, complexity and efficiency of processing the available information. In other words, there is no typical solution to the

Journal of Physics: Conference Series **1333** (2019) 032074 doi:10.1088/1742-6596/1333/3/032074

problem of storing and managing large amounts of data collected from heterogeneous, distributed sources of information.

## 2. Attribute Dependency Model

The problem of efficient use of heterogeneous and distributed information sources is the difficulty of ensuring timely access to the full list of information, the result of processing and analysis of which is necessary for management processes. As we have already noted, one of the solutions to the problem 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 have physical level integration. Integration at the logical level requires a single, global scheme that allows combining different ways of defining dependencies between data, as well as implementing different strategies for finding solutions.

To form a unified scheme for presenting digital field data, the author proposes to use the attribute dependency model, which is part of the object-oriented system analysis methodology [16].

Links between data located in heterogeneous sources are described through a network of functional dependencies [16]:

$$T^{A}(c) = < A(c), R^{fd}.$$

The network is a loose hierarchy that can have several roots. In the origins of the network are the basic attributes, the values of which do not depend on the values of other attributes [16]:

$$\{a_i\}^{B} = (a_i \in A(c) | \neg\exists (a_i, \dots a_n)\ (a_i, \dots a_n) R^{fd} a_i\}.$$

These include attributes, the values of which sets by the developer, contained in regulatory documents, collected from the corresponding sensors and controllers. The remaining attributes directly or indirectly depend on the base ones. The network attributes contain target attributes whose values determine the target state of the system [16]:

$$\{a_i\}^{G} = \left(a_i \in A(c) \middle| \neg\exists a_j\right)\ (a_i, \dots a_n) R^{fd} a_j\}.$$

A subset of target attributes, as a rule, includes performance and quality characteristics. Figure 1 shows an example of a functional network of dependencies for the task of deciding whether to carry out a geological and engineering operation – hydraulic fracturing (HF). Table 1 shows the decoding of the attribute values of the model.

**Table 1. Description of model attributes**

| Attribute | Description | Attribute type |
|---|---|---|
| $a_1$ | Scheduled time of the work | *base* |
| $a_2$ | Waiting time for the work start | *base* |
| $a_3$ | Oil production before HF | *base* |
| $a_4$ | Expected oil production after HF | *base* |
| $a_5$ | Mineral Extraction Tax | *base* |
| $a_6$ | Domestic oil price | *base* |
| $a_7$ | Accumulated loss during repair | *base* |
| $a_8$ | Well workover costs | *base* |
| $a_9$ | HF cost | *dependent* |
| $a_{10}$ | Price for 1 ton of oil without taxes | *dependent* |
| $a_{11}$ | Total costs | *dependent* |
| $a_{12}$ | Actual production increase | *dependent* |
| $a_{13}$ | Payback period | *target* |

If the expected value of the target attribute – the payback period – meets the efficiency conditions, a positive decision is taken to conduct the HF.

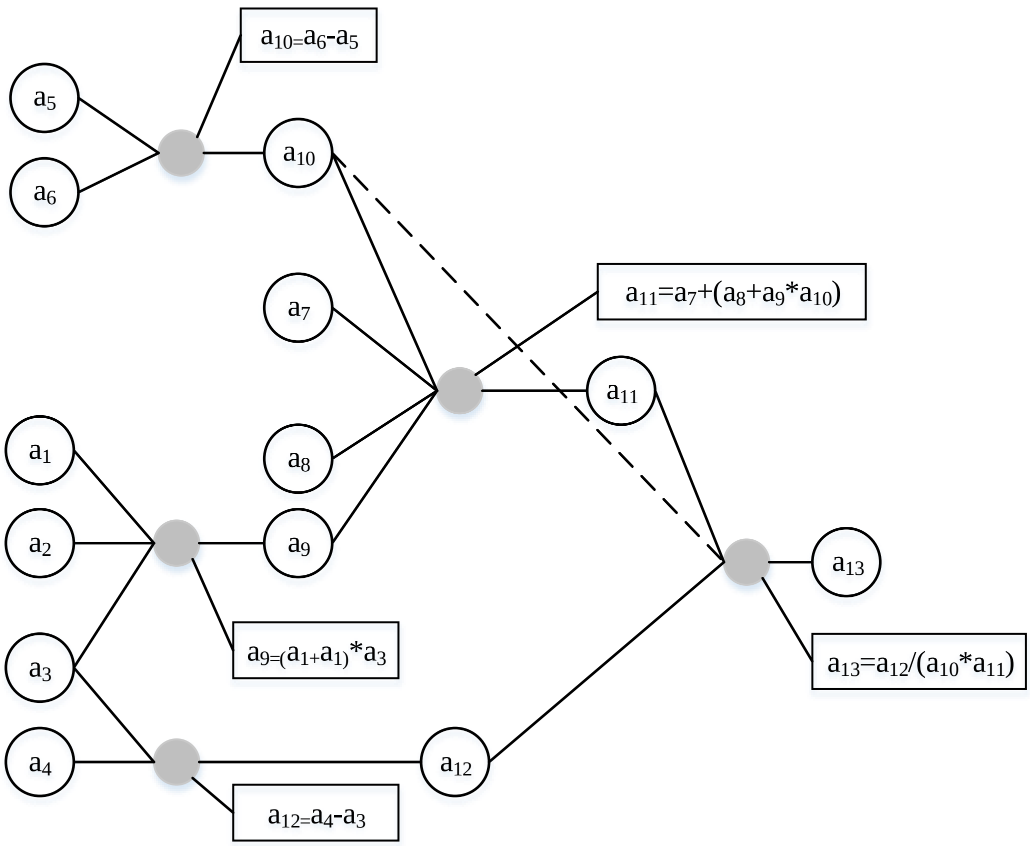


**Figure 1.** An example of an attribute dependency model for estimating the payback period of HF

## 3. Attribute meta-description format

The attribute dependency model allows you to describe the dependency of data stored in heterogeneous distributed sources. However, this description is not enough to automate the search tasks necessary for decision-making information. The author proposes to supplement the model with a meta description format for each type of data used. For readability, the format is presented as a fragment of an XML document.

```
<Attribute>
        <!-- Description of basic data type information -->
        <General>
                <ID> … </ID>
                <Name> … </Name>
                <MeasurementUnit> … </MeasurementUnit>
                <ValidRange>
                        <from> … </from>
                        <to> … </to>
                </ValidRange>
        </General>
        <!-- Description of data sources  -->
        <DataSources>
                <DataSource>
                        <ID> … </ID>
                        <Type> … </Type>
                        <AccessLine>  <!-- Automated Access Options -->
                        </AccessLine>
```

```
        </DataSource>Место для уравнения.
    </DataSources>
    <!-- Decision making tasks information -->
    <DecisionTasks>
        < DecisionTask>
            <ID> … <ID>
            <AttributeType> <!-- base or target -->
            </ AttributeType>
        </ DecisionTask>
    </ DecisionTasks>
</Attribute>
```

The proposed meta-description format is not final. As part of further research, the author plans to implement a search module to automate access to basic attributes for various decision-making tasks. According to the test results, the meta-description format will also be improved.

## 4. Results and Discussion

The attribute dependency model of digital oilfield data stored in heterogeneous, distributed sources of information will allow solving the tasks:

1. Interpretations, where the values of the basic attributes are initially specified and the values of the target attribute corresponding to the source data are searched for. Thus, the model can be used in the process of developing problem-oriented software solutions, namely when creating computational methods, determining their input and resulting parameters.
2. Search for a feasible solution. With a known value of the target attribute, you must find the valid values of the base.
3. An analysis of the attribute dependency model will reveal for each of the management tasks a set of basic attributes, respectively, eliminating the need to store data that is not basic to any of the tasks. This will optimize storage systems.

It is also advisable to note the complexity of the proposed approach. There is a need in rules set, techniques, and software, which will allow forming a meta-description of data for each of the sources. Depending on the type of source and data access interfaces, this can become extremely time-consuming and lengthy processes.

The presented results are intermediate. In the course of further research, the author proposes developing rules set for automated meta-descriptions creation of data stored within heterogeneous sources.

## 5. Acknowledgments

The reported study was funded by RFBR according to research project № 18-37-00311.

## References

[1] Qin H. and Han Z. 2016 *IEEE Systems Journal* **12** 1295–1306
[2] Hu H , Fan L and Guan X 2017 *Cloud Computing and Big Data Analysis* (Chengdu: IEEE)
[3] Alenezi F and Mohaghegh S 2016 *4th Saudi International Conference on Information Technology* (Riyadh: IEEE)
[4] Preveral A, Trihoreau A and Petit N 2014 *SPE Intelligent Energy Conference & Exhibition* (Utrecht: SPE)
[5] Patri O P, Panangadan A V, Chelmis C, McKee R G and Prasanna V 2014 *SPE Annual Technical Conference and Exhibition* (Amsterdam: SPE)
[6] Zhong Q and Du H 2015 *Proceedings of the 2015 International Industrial Informatics and Computer Engineering Conference* (Cambridge: IEEE)
[7] Mckenzie W, Morandini F, Philippe, Deny L, Rainaud J-F, Eastick R, Masters G and Mack D Endres 2012 *SPE Intelligent Energy International* (Utrecht: SPE)
[8] Zhao G, Di W and Wang R 2017 *Cloud Computing and Big Data Analysis* (Chengdu: IEEE)

[9] Alguliyev R M, Ramiz M. Aliguliyev R M and Hajirahimova M S 2016 *10th International Conference on Application of Information and Communication Technologies* (Baku: IEEE)
[10] Brule M R 2015 *SPE Digital Energy Conference and Exhibition* (Texas: SPE)
[11] Cameron D 2014 *SPE Intelligent Energy Conference & Exhibition* (Utrecht: SPE)
[12] David R M, Rao S R and Reddicharla N 2015 *SPE Middle East Intelligent Oil and Gas Conference and Exhibition* (Abu Dhabi: SPE)
[13] Perrons R K and Jensen J W 2015 *Energy Policy* **81** 117–121
[14] Chanana P, Son T M and Bhakne U 2016 *Offshore Technology Conference* (Kuala Lumpur: Offshore Technology Conference)
[15] Crompton J 2015 *SPE Digital Energy Conference and Exhibition* (Texas: SPE)
[16] Silich M P 2005 Information technology of the complex socio-economic systems design based on the object-oriented modeling methodology (Tomsk: Doctor of Technical Sciences Thesis work) 91–100