

Proposing of Single Entity Design Pattern in Big Agricultural Positioned Data Sets (ADS)

Jan Rajtr, Pavel Šimek, Jan Pavlík

Department of Information Technologies, Faculty of Economics and Management, University of Life Sciences Prague, Czech Republic

Abstract

With emerging usage of positioned devices such as drones, cell phones or IoT, the amount of data that can be collected expands drastically. At any given time, there is usually at least one nearby device that has positioning capabilities. Smart phones, smart TVs, personal computers, or even cars contain localization features. These vast amounts of data require a lot of effort in analysis and understanding in order to be properly utilized, which is especially true for the field of agriculture, where proper analysis can yield tremendous improvements in terms of production.

Current computer technologies offer plenty options for such analysis. However, not every agricultural subject has access to a mainframe with performance in petaflops to perform complicated analyses of such big data in a timely manner.

The defined design patterns for creation of data offers potential for speeding up the analysis of ADS on personal computers.

This article describes known and used creational patterns and compares their benefits regarding ADS and offers possible usage and improvements.

Keywords

Big data, agricultural, designing patterns, software engineering.

Rajtr, J., Šimek, P. and Pavlík, J. (2018) "Proposing of Single Entity Design Pattern in Big Agricultural Positioned Data Sets (ADS)", *AGRIS on-line Papers in Economics and Informatics*, Vol. 10, No. 4, pp. 65-69. ISSN 1804-1930. DOI 10.7160/aol.2018.100407.

Introduction

Generally, the design pattern may be described as a generalized solution to a programming problem with particular purpose. The design patterns are used as a high-level programming idiom especially by professional software engineers.

The reasons why usage of design patterns is common are many - because most of the programming problems are often comparable to those that have already been solved, creating new algorithms that tackle the same issue in a similar manner is a waste of time. On the other hand, it is always up to the software engineer to decide if the given programming problem should be solved with design patterns, because their usage may not resolve in improvement (with requirements of readability and understanding of its architecture) (Dascalu et al., 2005).

Big Data is defined by several characteristics apart from size, particularly, the volume, velocity,

variety, and veracity of the data (Coble et al., 2018). Big data is often mistakenly thought of as a huge database storage. This, however, is not precise, because every kind of data has a connection or relation to its origin, usage and utilization. This system approach leads to questions such as: what kind of relations are between data, what is the usage of data, what is the real information which the data bear. And this system conception leads to the emergence of design patterns. Every single one of these questions can lead to more complex systems which use more processing time and cost more money. So the overall complexity of the system is of crucial importance as it is directly linked to the required resources. Correctly designed and engineered system can be measured, and its complexity can be calculated (Pavlic et al., 2008).

Emerging usage of autonomous positioned devices such as smartphones, IoT devices and so on, leads to new problems, questions and different approaches

to seek the answers. The Internet of Things (IoT) has opened productive ways to cultivate soil with the use of low-cost hardware (sensors/actuators) and communication (Internet) technologies. Remote equipment and crop monitoring, predictive analytic, weather forecasting for crops or smart logistics and warehousing are some examples of these new opportunities (Ferrandez-Pastor et al., 2018). Also, remote sensing, as one of the sources for big data, is generating earth-observation data and analysis results on a daily basis, utilizing wide variety of technologies, ranging from satellites, manned/unmanned aircrafts, to ground-based structures. Agricultural remote sensing is one of the backbone technologies for precision agriculture, which considers within-field variability for site-specific management instead of uniform management as in traditional agriculture (Huang et al., 2018). The Big Data by itself is changing and the procurement of data and its analysis has becoming more structured and complex. The challenge of analyzing and processing a huge amount of data is becoming increasingly important in this fourth industrial revolution era (Guardo et al., 2018). Many researches point to optimization and effectiveness of whole processes such as data transfers, querying, mining or other methods in Big Data and particularly in agricultural sector can be divided into two data groups internal (private) data and external data (Stočes et al., 2018). Other sources present part of application assumption of agriculture network monitoring system and design the big data processing and analysis module (Chen et al., 2018). Also, the aspect of real time availability of data is more and more obvious in web-based systems. All of these requirements are important and change the perspective these issues are being looked at (Singh et al., 2018).

Materials and methods

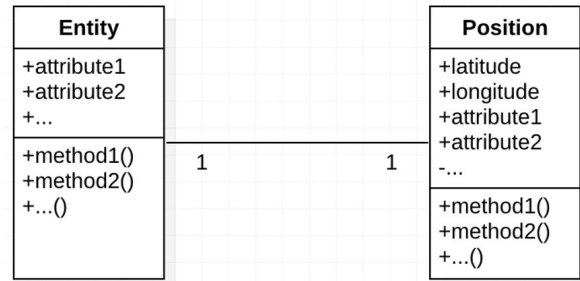
Let the Agricultural positioned data (ADS) sets be defined in its simplicity as a data which carry the position of its creation or position of discrete behavior (globally named as state). One ADS entity can be imagined as one simple entity with its information, behavior (state) and position of its state. Basic principle on which would be agricultural data set described is that every entity has its own geographical position. This concept has become common in past years thanks to emerging number of devices aggregating data with geographical positioning (Zhou and Li, 2018).

This data is usually stored in cloud or distanced databases. Every representation of entity in ADS

consists of two parts:

- structural data information – state (constants, behavior, methods)
- position – global geographic position

In the traditional object paradigm this information set would be described as the following system:



Source: authors

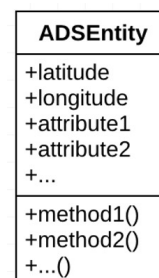
Figure 1: Traditional ERM/ORM.

The usage of this relation is to prevent duplicity inside relation-based databases and utilize the relation logic between objects more similarly to object-oriented databases. However, this traditional approach may be extremely dangerous in terms of big data. Even if the ADS system entity has always one and only one position, system must allocate memory and processing power to query (or hold relation) between two entities. In most cases SQL or noSQL systems are sufficient in these situations. But in case of millions of records, this approach may lead to significant slowdown. Key question would be how this pattern behaves in large data sets known as Big Data (Zhu, 2018).

The complexity of such system may be described in two ways – polynomial and data complexity. (Papadimitriou et al., 1999) Approximated mathematic description of data query complexity (Q) for the example above would be:

$$Q = f(n) = 2^n$$

ADS entity as part of simple one entity system stores the position data inside each entity. In object form the following scheme can be used:



Source: authors

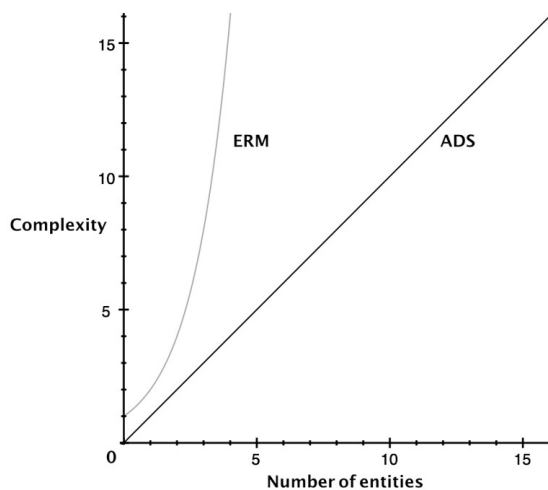
Figure 2: ADS entity.

Results and discussion

The ADS approach leads to linear n complexity because every ADS entity has the contextual position data available in one query or instance. Accordingly, the approximated mathematic description of data complexity would be following:

$$Q = f(n) = n$$

When comparing the complexity to the traditional ERM, it is obvious how the system can be simplified. With increasing amounts of data, the complexity of ADS is always linear and can be easily predicated. The graph of the complexity shows this relation with traditional ERM model.



Source: authors

Figure 3: Complexity comparison of traditional ERM/ORM to ADS.

The whole idea of measuring complexity of the system could be seen from three perspectives – computational, psychological, and representational. The ADS may be defined as a non-traditional design pattern which by itself - in software engineering - can make process of finding solution easier and a more understandable (Mangalaraj et al., 2014).

Additional minor questions may arise. Firstly – how to resolve duplicity and redundancy in data? For example, let the given ADS consist of weather measurements from multiple (hundreds to thousands) of IoT devices. The question of duplicity and redundancy of data – in this case position for every device – must be thought about in a complex way. Can be a location really be generalized on level of information base? Because what if the IoT device is movable? Such as smartphone, smart watch or drone. Should it

be really generalized to separate entities? Next and more important question is – how could the duplicity problem be solved algorithmically? Should the DBS check if there is same location in the database? In this case, the system would use valuable computer processing time for such queries. With the use of ADS this problem can be easily eliminated. The last question is – will the removal of redundancy be a significant benefit for data size? How much processing time is it worth to reduce the data size by let's say 1%? Such calculations are however out of the scope of this paper.

The last thing is that the usual information context of the position in ADS type systems is pointless without other data or context. This means that positional data in terms of ADS does not describe anything useful (apart from the position itself). To understand and use the location data we would need more information or metadata to describe the position itself. But by doing that, the location would become its own entity and should be stored separately, which would break the ADS paradigm by increasing the structure complexity.

Finally, the goal to understand ADS on a deeper level leads to other areas, such as research in spatial databases, tree indexing algorithms, synthetic databases and so on. It is important to mention, that recent researches show that finding highly effective algorithms and querying methods may not be possible (Roumelis et al., 2017).

Conclusion

Proposed approach to store position in ADS can be used and is recommended to use in atomic systems where ADS entities are of discrete character. This means that there is no prediction for continuous changes of position data in ADS entities. For example, atomic measurements from IoT, timed analysis and so on.

ADS could be represented in many forms in GIS analysis and system automatization processes. It is also important, that ADS stand on the premise of abandoning database relation between position and information. This means that it may not meet the conditions of Boyce-Codd forms or equivalent relation or object-based data schemes. However, this does not prevent the ADS concept or at least some parts of it to be utilized in design patterns of object programming and software engineering. If the positional data were to be separated from other information into special sets (like tables in SQL), it would lead only to extreme data swelling with no speedup or system

improvement (except elimination of duplicity and redundancy). Proposed solution offers simple and fast approach to solve issue of growing data complexity in big positioned data sets. Importance of similar approaches is increasing regarding the different architectures and kind of meta-data which are stored for upcoming big data mining (Guerrero, 2017).

Corresponding authors

Ing. Jan Rajtr

Department of Information Technologies, Faculty of Economics and Management

Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Prague – Suchbát, Czech Republic

Phone: +420 22438 2049, Email: rajtr@pef.czu.cz

Acknowledgements

This article was created with the support of the Internal Grant Agency (IGA) of FEM CULS in Prague, no. 20181010 – „Identification of deficiencies and limitations of remote sensing data processing and its effective use“.

References

- [1] Dascalu, S., Hao, N. and Debnath, N. (2005) “Design patterns automation with template library”, *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, pp. 699-705. ISBN 0-7803-9313-9. DOI 10.1109/ISSPIT.2005.1577183.
- [2] Chen, J., Zhong, Y. and Lam, A. (2018) “Research on Monitoring Platform of Agricultural Product Circulation Efficiency Supported by Cloud Computing”, *Wireless Personal Communications*, Vol. 102, No. 4, pp. 3573-3587. ISSN 0929-6212. DOI 10.1007/s11277-018-5392-3.
- [3] Coble, K. H., Mishra, A. K., Ferrell, S. and Griffin, T. (2018) “Big Data in Agriculture: A Challenge for the Future”, *Applied Economics Perspectives and Policy*, Vol. 40, No. 1, pp. 79-96. ISSN 2040-5790. DOI 10.1093/aepp/ppx056.
- [4] Ferrandez-Pastor, F. J., Garcia-Chamizo, J. M., Nieto-Hidalgo, M. and Mora-Martinez, J. (2018) “Precision Agriculture Design Method Using a Distributed Computing Architecture on Internet of Things Context”, *Sensors*, Vol. 18, No. 6, Article no. 1731. ISSN 1424-8220. DOI 10.3390/s18061731.
- [5] Guardo, E., Di Stefano, A., La Corte, A., Sapienza, M. and Scata, M. (2018) “A Fog Computing-based IoT Framework for Precision Agriculture”, *Journal of Internet Technology*, Vol. 19, No. 5, pp. 1401-1411. ISSN 1607-9264. DOI 10.3966/160792642018091905012.
- [6] Guerrero, J. I., Garcia, A., Personal, E., Luque, J. and León, C. (2017) “Heterogeneous data source integration for smart grid ecosystems based on metadata mining”, *Expert Systems with Applications*, Vol. 79, pp. 254-268. ISSN 0957-4174. DOI 10.1016/j.eswa.2017.03.007.
- [7] Huang, Y., Chen, Z. X., Yu, T., Huang, X. Z and Gu, X. F. (2018) “Agricultural remote sensing big data: Management and applications”, *Journal of Integrative Agriculture*, Vol. 17, No. 9, pp. 1915-1931. ISSN 2095-3119. DOI 10.1016/S2095-3119(17)61859-8.
- [8] Mangalaraj, G., Nerur, S., Mahapatra, R. and Price, K. H. (2014) “Distributed Cognition in Software Design: An Experimental Investigation of the Role of Design Patterns and Collaboration”, *MIS Quarterly: Distributed Cognition in Software Design: An Experimental Investigation of the Role of Design Patterns and Collaboration*, Vol. 38, No. 1, pp. 249-274, ISSN 0276-7783. DOI 10.25300/MISQ/2014/38.1.12.
- [9] Papadimitriou, CH. and Yannakakis, M. (1999) “On the Complexity of Database Queries”, *Journal of Computer and System Sciences*, Vol. 58, No. 3, pp. 407-427. ISSN 0022-0000. DOI 10.1006/jcss.1999.1626.
- [10] Pavlič, M., Kaluza, M. and Vrček, N. (2008) “Database Complexity Measuring Method”, *Proceedings of the 19th Central European Conference on Information and Intelligent Systems*, pp. 577-583. ISBN 978-953-6071-04-3.

- [11] Roumelis, G. Vassilakopoulos, M. Corral, A. and Manolopoulos, Y. (2017) "Efficient query processing on large spatial databases: A performance study", *Journal of Systems and Software*, Vol. 132, pp. 165-185. ISSN 0164-1212. DOI 10.1016/j.jss.2017.07.005.
- [12] Singh, A., Garg, S., Batra, S., Kumar, N. and Rodrigues, J. J. P. C. (2018) "Bloom filter based optimization scheme for massive data handling in IoT environment", *Future Generation Computer Systems*, Vol. 82, pp. 440-449. ISSN 0167-739X. DOI 10.1016/j.future.2017.12.016.
- [13] Stoces, M., Masner, J., Kanska, E. and Jarolimek J. (2018) "Processing of Big Data in Internet of Things and Precision Agriculture", *Agrarian Perspectives XXVII.: Food Safety - Food Security, Proceedings of the 27th International Scientific Conference*, pp. 353-358. ISBN 978-80-213-2890-7, ISSN 1213-7979.
- [14] Zhou, X. L. and Li, D. Y. (2018) "Quantifying multi-dimensional attributes of human activities at various geographic scales based on smartphone tracking", *International Journal of Health Geographics*, Vol. 17, Article no. 11. ISSN 1476-072X. DOI 10.1186/s12942-018-0130-3.
- [15] Zhu, H. J. and Zhu, L. H. (2018) "Real-time positioning of a specific object in the big data environment", *EURASIP Journal on Wireless Communications and Networking*, Article no. 43. ISSN 1687-1499. DOI 10.1186/s13638-018-1043-3.