Data Management with AI Methods in Distributed Environment

Włodzimierz Funika[1,2], Filip Szura[1,2], Jacek Kitowski[1,2]

[1] AGH University of Science and Technology, Faculty of Computer Science Electronics and Telecommunication, Department of Computer Science, al. A. Mickiewicza 30, 30-059 Krakow, Poland
[2] AGH University of Science and Technology, ACC CYFRONET AGH, ul. Nawojki 11, 30-950 Krakow, Poland

Emails: {funika,szura,kito}@agh.edu.pl

Keywords: data management, artificial intelligence, knowledge engine, fuzzy logic, rules

1. Introduction

Data management is a very important issue in distributed data systems, e.g. data grids. Most of the related issues are connected with the location of the data in the environment. The location of data should be aimed to optimize the most significant parameters and provide the reduction of the cost and network load connected with data migration between sites. Due to that it is important to perform these actions very fast without the administrator's interference. The solution which is going to be presented should allow reducing the cost of communication since nowadays data centers are able to offer and manage vast amounts of data but the links between data centers as well as between the data center and the end user still may cause bottlenecks. Moreover, links between data centers and computing units may become a bottleneck if the data will be too far from the computing node. Due to this it is important to manage the data so that they should be placed "locally".

2. Related Work

As mentioned above the data management becomes a very important issue addressed in a number of papers. A first interesting solution, Autopilot [1] is aimed to manage a heterogeneous computational grid. This solution is described by authors as an infrastructure oriented towards dynamic performance tuning. A next interesting approach is Google Bigtable [2]. It is described as a distributed storage system for managing structured data that is designed to scale to a very large size (petabytes of data) across thousands of commodity servers. In particular, a Bigtable is a sparse, distributed, persistent multidimensional sorted map. On the physical level, Bigtable uses the distributed Google File System (GFS) [3]. Another data management approach is Amazon Dynamo [4] which is a highly available and scalable distributed key/value based datastore built for supporting internal Amazon's applications. In the Dynamo system, each data item is replicated at N hosts where N is a parameter configured "per-instance".

3. Description of a problem solution

In this paper we are going to present an extension to our approach to data management which is called KODMAS (KnOwledge based Data MAnagement Support) which was presented in [5]. This approach exploits fuzzy logic which is supported by rules. KODMAS uses knowledge on possible actions which can be performed when needed. In KODMAS we combine an expert system with static defined actions with a dynamically changing knowledge offered by fuzzy logic. The mentioned expert system is implemented using rules that define relevant actions. The second part of KODMAS's knowledge is implemented using fuzzy logic. This part is able to modify its knowledge and enables self-learning. After an action execution the system observes the monitored infrastructure to state if the performed action was helpful or not. After that the system is able to modify its knowledge. Apart from the learning method previously used, KODMAS exploits separate knowledge (KW-I) to tune the fuzzy logic data (KW-M). This modification is done by an expert system which is based on rules. The knowledge of the described system is depicted in Fig.1. With KW-I knowledge we are closer to recognize the context in which the change of system behavior was observed. Fig. 1 shows a system response on a data incoming from the monitored and managed infrastructure. It also shows the modification of the fuzzy logic engine when the system detects that the performed action influences the infrastructure. The outcome are the actions proposed by the system.

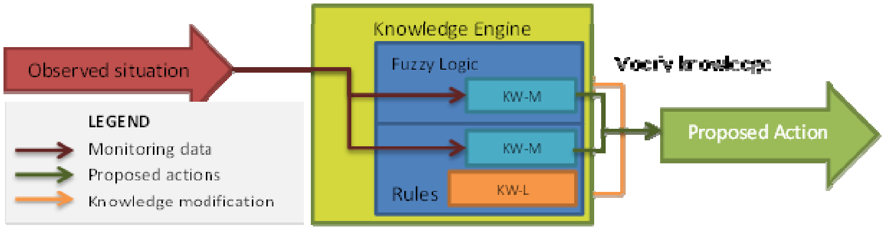

Fig. 1. Knowledge-based processing of storage-related workflow.

4. Results

During the tests of the extended system we observed that the introduced expert system which deals with fuzzy logic-based learning is able to perform the same actions as the learning phase involved in the fuzzy logic engine. But this solution allows also for defining a kind of politics which allows for learning in the knowledge engine with an extended set of learning data. This learning is based not only on the definition of action state and observed parameters but it takes into account also the type of monitored entity, the location in the fuzzy set that relates to the extent of improvement or deterioration of the system state, which is dependent of the actions performed.

5. Conclusions

In this paper we aimed to present an extension to KODMAS, which is focused on improving the learning phase of our knowledge engine that supports the management of data storage. This approach should allow to modify knowledge to better exhibit the needs of the monitored infrastructure. The knowledge modification is customizable to the observed situations, monitored entity, and current knowledge. The presented method of learning also enables to change and tune the knowledge that describes the learning phase even when the system is running. The use of rules also provides a human-understandable way to tune the method of learning and its parameters.

Acknowledgements:This research is partly supported by the European Union within the European Regional Development Fund program as part of the PLGrid PLUS Project POIG.02.03.00-00-096/10 (`http://plgrid.pl/plus`).

References

1. R. L. Ribler, H. Simitci, D.A. Reed:. The autopilot performance-directed adaptive control system. Future Gener. Comput. Syst., 18(1):175–187, September 2001.
2. J. Dean et al.:. Bigtable: A Distributed Storage System for Structured Data web site: `http://research.google.com/archive/bigtable.html`
3. H. Gobioff et al.: The Google File System, web site: `http://research.google.com/archive/gfs.html`
4. D. Hastorun D. et al.:. Dynamo: amazon's highly available key-value store, ACM SIGOPS Operating Systems Review - SOSP '07, 2007.
5. W. Funika, F. Szura, J. Kitowski:. Automation of System Monitoring Based on Fuzzy Logic or Rules; Comparison of Two Designed Approaches with Regard to Computational Infrastructures, LNCS vol. 7136, Springer, 2012, pp. 142-156.