

## MOTIVATION

- Rapid growth of data volumes which requires data storage monitoring
- Necessity to meet QoS parameters demanded by users w.r.t. data access
- Cost reduction due to data replication and communication between sites
- Ability to automate replica monitoring and management

## BENEFITS FROM USE OF AI METHODS

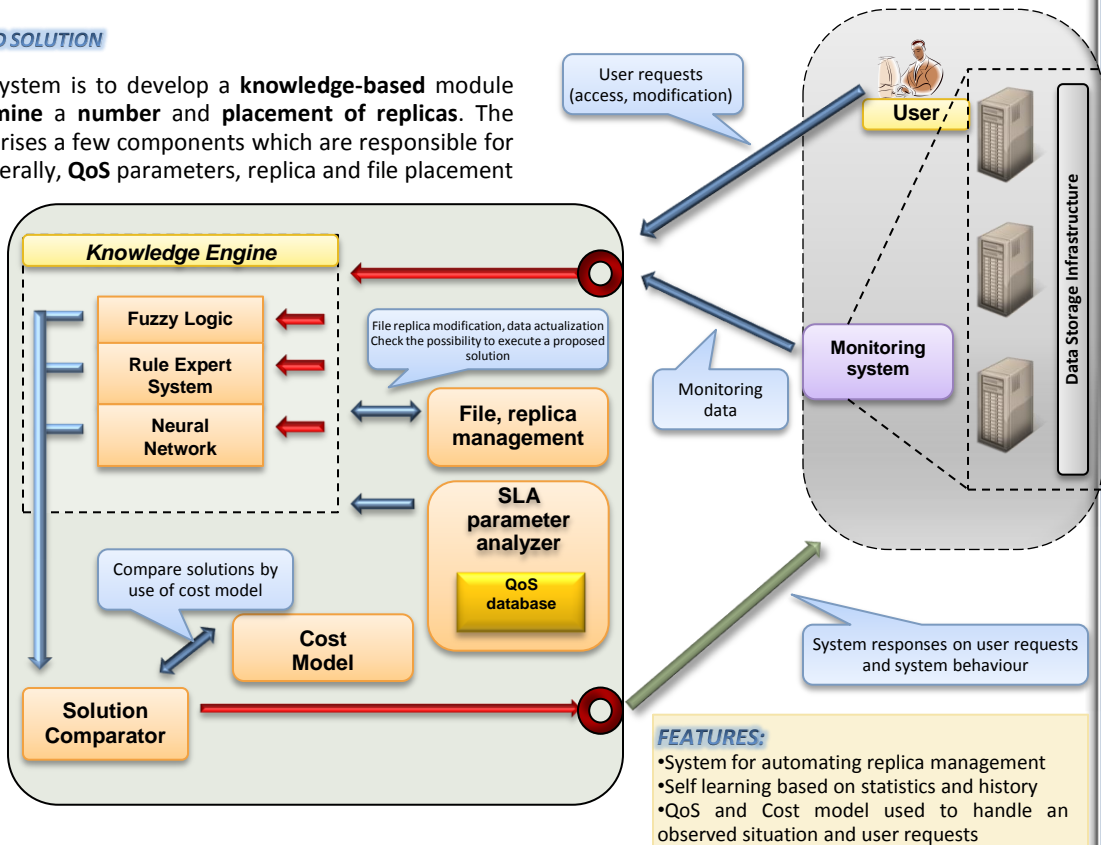
- Matching data-bound solutions to the observed situations
- Self learning to acquire new best suitable solutions
- Prediction of data storage system's behavior
- Cost effective data management and storage

## DESCRIPTION OF PROPOSED SOLUTION

The main idea for this system is to develop a **knowledge-based** module which is used to **determine a number and placement of replicas**. The presented solution comprises a few components which are responsible for management of SLA, generally, **QoS** parameters, replica and file placement and integration.

The Knowledge of this system is used when the user modify data or access their files. It is also used when the system detects probable QoS breaks. The system uses knowledge to define solutions, e.g. creating new replicas.

To do that it uses an SLA parameter analyzer which checks the defined contracts and tries to predict which one may be endangered.



### FEATURES:

- System for automating replica management
- Self learning based on statistics and history
- QoS and Cost model used to handle an observed situation and user requests

In the presented solution the exploited knowledge is described using **three approaches**. They are: **fuzzy logic**, **rule expert system** and **neural network**. This solution can also combine both the fuzzy logic and rule-based approach within the knowledge engine. In addition to the knowledge the system also uses a cost model in which the cost is evaluated when the knowledge engines return their responses to the user (solutions). It determines which of the solutions is more preferable to be applied and answers the following question „is this solution (i.e. ,the actions implied) cost-effective?”

## FUTURE WORK

- Extend the cost model function which will enable to take into account the cost of data transmission between the user and their data resources
- Extend the knowledge of the system, e.g., by functions which allow for better utilization of neural network
- Implement a context approach which is aimed to allow evaluating if the solution applied is helpful in terms of infrastructure or not.

## REFERENCES

- [1] K. Ranganathan and I.T. Foster: "Identifying Dynamic Replication Strategies for a High-Performance Data Grid", in Proc. GRID, 2001, pp.75-86.
- [2] D. Krol, B. Kryza, K. Skalkowski, D. Nikolow, R. Slota, J. Kitowski: QoS Provisioning for Data-Oriented Applications in PL-GRID, In Proc. of Cracow Grid Workshop - CGW'10, October 11-13 2010, ACC-Cyfronet AGH, 2011, Krakow, pp. 142-150
- [3] W. Funika, F. Szura, J. Kitowski, Agent-based monitoring using fuzzy logic and rules, in Computer Science Annual of AGH-UST, vol. 12, 2011, pp. 103-113, AGH Press, Krakow, 2011; ISSN 1508-2806

## ACKNOWLEDGMENTS

The research presented in this paper has been partially supported by the European Union within the European Regional Development Fund program no. POIG.02.03.00-00-007/08-00 as part of the PL-Grid Project ([www.plgrid.pl](http://www.plgrid.pl)) and ACC Cyfronet AGH grant 500-08.