A Resource Prediction System using Data Mining

Włodzimierz Funika[1,2], Tomasz Kwiecień[1], Marcin Nowak[1]

[1] AGH, Faculty of Computer Science, Electronics and Telecommunication, Dept. of Computer Science, al. Mickiewicza 30, 30-059, Kraków, Poland,
[2] AGH, ACC CYFRONET AGH, ul. Nawojki 11, 30-950, Kraków, Poland,
emails: {mnowak, tkwiecie}@student.agh.edu.pl, funika@agh.edu.pl

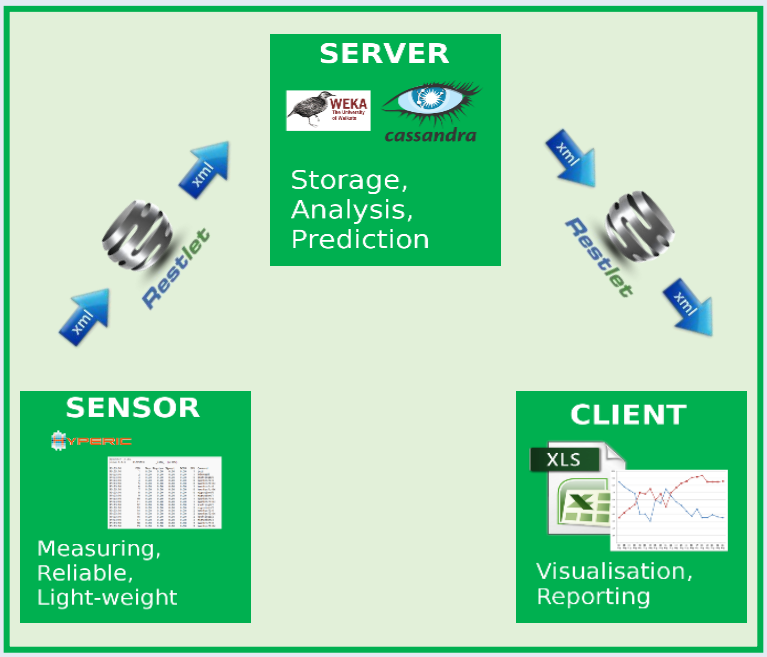
Motivation

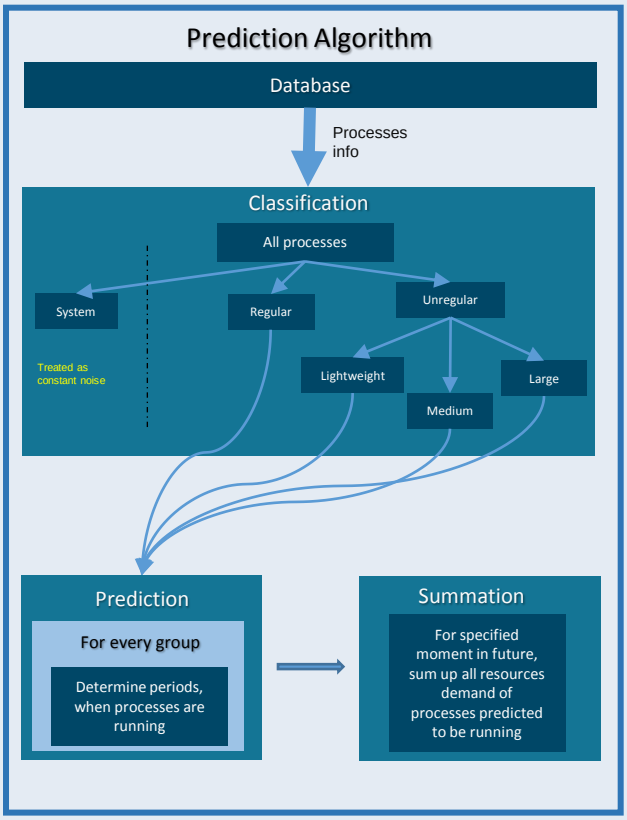
- Large number of big systems with enormous data sets
- Variable demand for CPU and data storage resources
- Problems with predicting the CPU and memory usage and other related values

Scientific goals

Design and implementation of a system that is aimed to:

- Organize data about resource usage
- Predict future resources demand
- Visualize predicted/measured values





Solution

System composed of three modules:

- sensor - measuring current consumption using Hyperic Sigar API [1] and sending results to the server
- server – storing data in Apache Cassandra Database [2] and predicting future values
- client - GUI application visualizing the measured values and predicted results

Summary

Currently project is in the implementation phase. Our plan for the nearest future is to identify statistical, heuristic and machine learning methods, which will be suitable for prediction of resource consumption. At this point, the WEKA library [3] is under testing for our research purposes. It provides a set of various data analysis algorithms, such as clustering or classification/regression. The analysis of the problem gives high hopes that for the issue presented, it is feasible to create a system that will be able to give concrete results, which in turn could be used in determining the schedule of big data computations on large multiprocessor platforms.

References

1. Hyperic Sigar API web site: http://www.hyperic.com/products/sigar
2. Apache Cassandra web site: http://cassandra.apache.org/
3. WEKA web site: http://www.cs.waikato.ac.nz/ml/weka

Acknowledgements

This research is partly supported by the European Union within the European Regional Development Fund program as part of the PLGrid Core Project POIG.02.03.00-12-137/13 (http://www.plgrid.pl/projekty/core) and the AGH-UST grant no. 11.11.230.124.