Artur KRUPA, Bartosz SAWICKI

Warsaw University of Technology

doi:10.15199/48.2015.12.35

High-resolution scatter analyse using cloud computing

Abstract. *Cloud computing is the newest approach to solve computationally challenging problems. It is oriented on optimization of processing costs using low-budget, standard computers. Algorithmic scheme for such problems is MapReduce. We will show how to use MapReduce architecture to efficiently solve high number of independent analysis needed for scatter plots. Presented case study is based on simple student problem solved using FEM. High-resolution scatter plot image introduce new quality in visualization of results.*

Streszczenie. *Chmura obliczeniowa (ang. cloud computing) to najnowsze podejście do rozwiązywania problemów złożonych obliczeniowo. Jest to architektura zorientowana na optymalizację kosztów przetwarzania przy użyciu niskobudżetowych, standardowych komputerów. Algorytmem obliczeniowym dla takich problemów jest MapReduce. W niniejszym artykule pokażemy jak wykorzystać MapReduce do efektywnego rozwiązywania dużej liczby niezależnych analiz, które zostaną zobrazowane przy pomocy wykresu zmienności. Zaprezentowany przykład jest prostym studenckim problemem MES. Wysokiej rozdzielczości analiza wprowadzaj nową jakość w wizualizacji wyników. (**Wysokiej rozdzielczości analiza zmienności parametrów przy wykorzystaniu chmury obliczeniowej**).*

Keywords: cloud computing, electric field simulation, scatter analyse, MapReduce model.
Słowa kluczowe: chmura obliczeniowa, symulacja pola elektrycznego, analiza zmienności parametrów, model MapReduce.

Introduction

The aim of the scatter analysis is to visualize variability of solutions as a function of input parameters. To prepare such plot, problem has to be analysed many times for different values of parameters. Scatter plot helps to understand relations within the problem, so it is one of the most popular scientific tools.

Data required for scatter plot could be acquired from different sources. The simplest one is to observe natural variability of measured quantity. This approach is usually taken for biological or environmental studies with high natural variability [1]. Another way is to build experiment in which parameter is changed in specified range. Such solution is popular when technical objects are analysed. And finally modelling techniques could be used to sweep parameters over wide range of variability.

Numerical modelling techniques allows to significantly speed up process of parameter sweeping. Especially that each simulation is independent, so it is easy to parallelize them. Among many schemes and architectures used for parallel computing, terms „cloud computing" and „volunteer computing" are today the most active [2]. While the cloud computing is commercial product built around cost-effectiveness, in volunteer networks computational power could be nearly free for researchers [3]. Socially important problems such as climate prediction or cancer therapies are able to attract millions of users willing to donate their computational time. Cloud computing and volunteer computing are both based on distributed architecture, so similar effects can be observed.

In this paper it will be shown how simple scatter analysis performed during student course "Electromagnetic Fields" could be refined by increasing its resolution by usage of computations based on cloud comping architecture.

Student problem no. 5

Scatter analysis is widely used in many scientific problems. As an example simple student problem was chosen. "Student problem no. 5" is studied during laboratory of electromagnetic fields at Faculty of Electrical Engineering at Warsaw University of Technology.

Measurements stand is presented on Fig. 1. This practical exercise is about internal structure identification based on current field measurements [4].

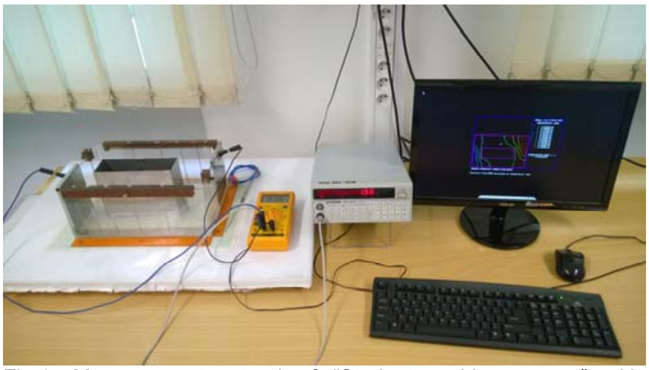

Fig.1. Measurements stand of "Student problem no. 5" with important elements (from left): model, multimeter, function generator and PC platform.

Laboratory model, presented on Fig. 2 is a container made from transparent material filled with distilled water (dielectric). Circuit is powered by voltage source which is connected to two electrodes placed at opposite sides (near opposite corners) of narrower edges of the container. Coordinates of the left-bottom corner of the metal frame are parameters, which are non-linearly related with resistance of the model. Distribution of electric potential inside the container is visualized on Fig. 3.

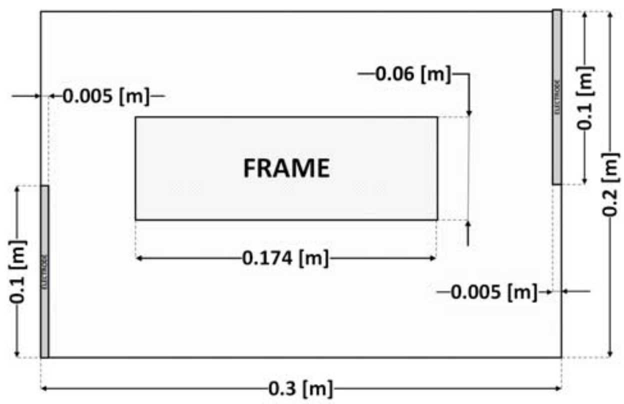

Fig.2. "Student problem no. 5" - scheme of the model

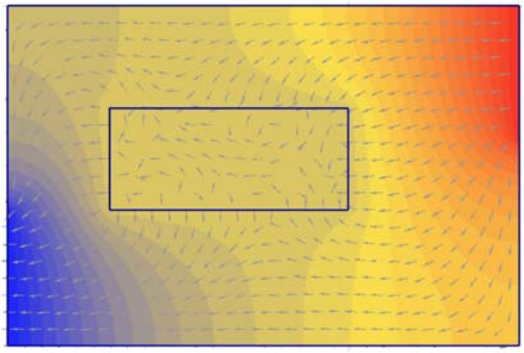

Fig.3 Solution of electric scalar potential (colormap) and electric field E (vectors) for analysed problem.

Presented model is used by students in laboratory to practice three aspects of inverse problems. The first part is intended to investigate sensitivity of electric field inside container. The second part of exercise is designed to show use of Newton's method to solve a one dimension inverse problem and localize an object inside container based on given current value. In the third part students use a computer to solve Newton's method of above problem, but in two dimensions. Based on potential values measured on electrodes placed over walls of the container, students gets reconstructed position of the object and compares it with the real one.

Manual scatter analysis

Important part of the lab is variability analysis using scatter plot. Students do a measurements of the model to determine a sensitivity map of a target function in specific points inside the container. Target function is defined by equation:

(1) $$F(x,y) = (I - I_z)^2$$

where: x,y – coordinates of the frame, I – measured current, I_z – current value given by the teacher.

Using method of disturbances, gradient of target function is calculated and visualized by the scatter plot.

Another, but similar task is to prepare a map of total resistance. Students use a multimeter to read a current values and then, having voltage source from generator, are able to find a resistance. This procedure is repeated for several positions of the frame.

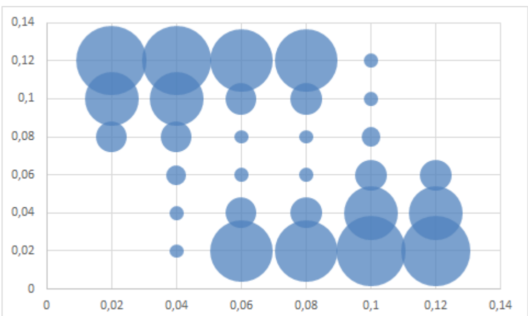

Fig.4. Typical scatter plot prepared by the students with 36 measurements points

The typical result is presented on Fig. 4. As one can see scatter plot (also referred as a bubble plot) is a set of circles for which radius related with value of visualized result. Centre of the circle comes from (x,y) coordinates of the frame.

Manual measurements has many advantages for teaching engineers, but it is not possible to significantly accelerate and automate measurements. To obtain stable results is necessary to take into account time needed to stabilize the water level. On the other hand, too long duration of the experiment causes evaporation and electrolytic phenomena, which interfere with its results.

The above circumstances explains why students can take only several measurements during lab, and why obtained variability image (scatter plot) has very low resolution. This kind of result is acceptable for experienced technicians, but for beginners it is hard to create feeling of continuous function.

Computer-aid scatter analysis

In previous section we specify the main problem. We explained why resolution of analysis is important and why measurements are a time consuming tasks.

Manual processing is limited, so natural solution is to take advantage of computing machines. Measurements procedure is replaced by numerical simulation of model described by partially differential equations based on Finite Element Method. We developed simulation software, which on a typical modern PC requires about 5 seconds to solve single forward problem.

This way we obtained system capable to run many simulations for different input parameters (coordinates of the metal frame). To avoid troublesome declaration of resolution at the begging of analysis, Monte Carlo method is used for generation of input data [5]. Random distribution of parameter allows to smoothly increase resolution by adding subsequent simulations.

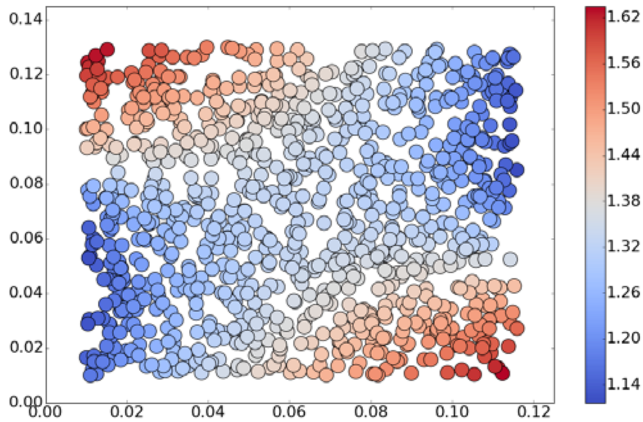

Fig.5. Scatter plot of 1.000 simulations. Input parameters are generated by uniform distribution Monte Carlo method.

On Fig. 5. scatter analysis using 1000 simulations is presented. Total computational time was approx. 5000 seconds (1.5 h) on standard PC. Value of resistance is displayed using color, what is advisable for large number of points. Comparing with Fig. 4, resolution is much better, but imagination is still required to see full image.

High-resolution scatter analysis

How can we increase resolution further? If we would like to have 1mm resolution, number of forward problems reach 14.300 instances. This would take about 20 hours of computations time, or over 3 days of continuous manual human work. Natural solution for this issue is deployment of parallel computer processing.

Set of independent simulations can be easily parallelized, this class of problems is even called "embarrassingly parallel". Nevertheless fluent management of thousands of machines solving millions of tasks is still challenging [4].

As already mentioned in the introduction cloud computing is new, promising, and flexible concept of parallelization. Usually cloud computing is used to process large datasets, but its heart (MapReduce algorithm) can be also used to efficiently manage thousands of computational tasks [6][7].

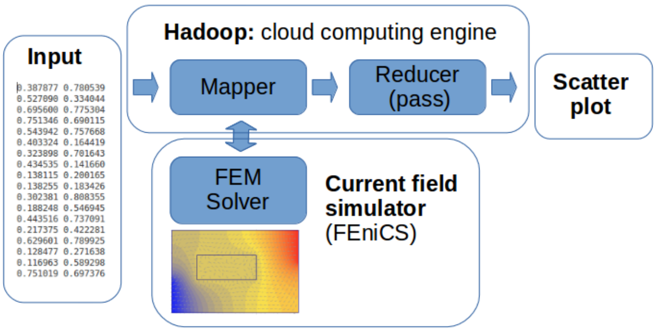

Fig.6. Scheme of the cloud computing system based on Hadoop implementation of MapReduce

We proposed a computation environment ready to solve defined problem. Architecture of developed cloud computing system is presented on Fig. 6. MapReduce framework consists of two main processing stages, mapper and reducer. Programmers are able to provide any software that will be run as mapper or reducer. Straight forward approach is to communicate mapper with the problem simulator and treat input data as the problem parameters [8]. Then MapReduce framework will spread calculations over all available computing nodes. Proper use of reducer stage is the key for optimal implementation. For scatter analysis no reduction is required, so reducer is only passing through. Using this method we can compute points and generate an electric field with determined accuracy. If we declare density of our grid, we can compute better view of our field. This makes our computations depend on our possibilities and wants.

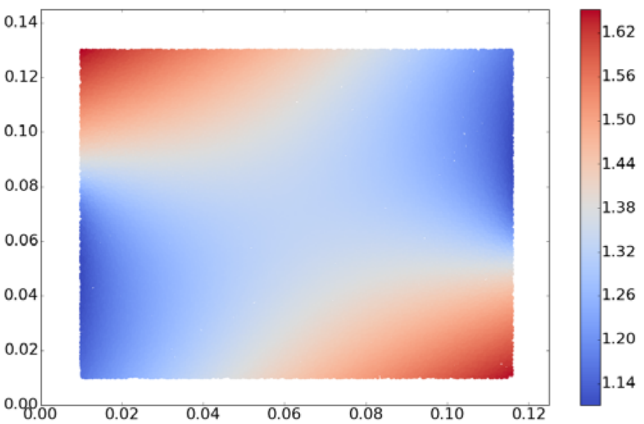

Fig.7. Scatter plot of 100.000 simulations using cloud computing based on MapReduce framework

Final result of scatter analysis is presented on Fig 7. Number of simulations is 100.000, where parameters were chosen using Monte Carlo method. Cluster consisting of 24 computational cores controlled by Hadoop solved this problem in 6 hours. It's worth to mention that cloud computing scheme allows to delegate our simulations into external service providers, and significantly reduce time required to complete the job.

High-resolution image presented on Fig. 7 is the final solution. Smooth colour transitions on the colormap are natural for function. Even inexperienced viewer can now analyse and properly understand continuity of variability.

Summary

The scatter plot is one of the important tools for analysing electromagnetic problems. Based on real student laboratory problem, methodology of cloud based high-resolution scatter plot has been presented. MapReduce model proved to be efficient and applicable to solve thousands of independent simulations.

High-resolution plots provides very different quality comparing with manual scatter analysis. Image with continuous plot helps students to really see investigated function and deeper understand its variability.

REFERENCES

[1] B. Sawicki, Uncertainty of numerical simulations in bio-electromagnetic problem, *Przegląd Elektrotechniczny*, no. 07, pp. 49-51 (2015)
[2] G. D'Angelo, M. Marzolla, New trends in parallel and distributed simulation: From many-cores to Cloud Computing, *Simul. Model. Pract. Theory,* 49 (2014), pp. 126
[3] D. Kondo, B. Javadi, P. Malecot, F. Cappello, D. P. Anderson and U. C. Berkeley, Cost-Benefit Analysis of Cloud Computing versus Desktop Grids, *Parallel & Distributed Processing 2009, IPDPS 2009, IEEE International Symposium on. IEEE, 2009* (2009), p. 1-12
[4] S. Filipowicz, S. Krzemiński, J. Sikora, J. Starzyński (red.), S. Wincenciak, Laboratorium Podstaw Elektromagnetyzmu, *Oficyna Wydawnicza PW*, (2005)
[5] H. Miras, R. Jimenes, C. Miras, C. Goma, CloudMC: a cloud computing application for Monte Carlo simulation. *Physics in Medicine and Biolog*, 58 (2013), N125–33
[6] B. S. Kim, S. J. Lee and T. G. Kim, MapReduce Based Experimental Frame for Parallel and Distributed Simulation Using Hadoop Platform, *in Proceedings 28th European Conference on Modelling and Simulation*, (2014)
[7] Radenski, Atanas, Big Data, High-Performance Computing and MapReduce, *in Proceedings 15th International Conference on Computer Systems and Technologies – CompSysTech'14*, (2014), p. 13–24.
[8] A. Krupa, B. Sawicki, Massive simulations using MapReduce model, *Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska, Accepted for publication* (2015)

***Authors**: M.Sc. Artur Krupa, Warsaw University of Technology, IETiSIP, 75 Koszykowa St., 00-662 Warsaw, POLAND, Email: Artur.Krupa@ee.pw.edu.pl; Ph.D. Bartosz Sawicki, Warsaw University of Technology, IETiSIP, 75 Koszykowa St., 00-662 Warsaw, POLAND, Email: Bartosz.Sawicki@ee.pw.edu.pl.*