С Е К Ц І Я 1

ЕКОНОМІКА ПІДПРИЄМСТВ ТА ІНВЕСТИЦІЙ
BUSINESS ECONOMICS AND INVESTMENT

S E S S I O N 1

Графічний процесор загального призначення як альтернативна інвестиція в IT

Катажина Мровца

Факультет менеджменту,
Краківський університет економіки,
ПОЛЬЩА, Краків, вул. Раковіцка 27,
E-mail: mrowca.kasia@gmail.com

В наш час практично всі середні та великі компанії використовують інформаційні технології (IT). Надзвичайно важко втриматись на ринку без технологічної підтримки. IT повсюди, вони є більш розповсюдженими і набагато дешевшими, ніж були в кінці XX ст. На початку ери оцифровування інвестиції в технології видавались досить простими, але зараз це питання значно ускладнилось.

Перш за все менеджер компанії повинен відповісти на запитання, чи варто інвестувати кошти в IT. Якщо інвестиція таки потрібна, то наступним важливим завданням є вибір технології. Кожній інвестиції повинен передувати економічний аналіз. У деяких випадках не варто переходити зі старих технологій на нові, оскільки це не принесе компанії значних прибутків. Більше того, інвестиції в IT є ризикованими, їх введення може зайняти навіть декілька років і не завжди буде успішним. Згідно досліджень Standish Group лише 10% інвестицій в IT були здійснені у запланований термін та згідно запланованих витрат. Більше третини таких проектів було припинено [17]. Незважаючи на те, що введення нових IT в компанії є ризикованим, воно також може бути і вигідним.

Останнім часом зростає потреба обчислення наборів даних. Тому компанії інвестують у власні центри зберігання та обробки даних (ЦЗОД) або у хмарні обчислення. У даній роботі наведено альтернативні можливості – рішення на основі графічних адаптерів. Це багатообіцяюча технологія, вона значно дешевша і не росповсюджена у Європі. Саме тому її впровадження може бути досить прибутковим для компанії.

Графічний процесор загального призначення – це графічний адаптер, який може обчислювати нерафічні набори даних. Графічний процесор може обробляти дані швидше, ніж центральний процесор завдяки паралельній архітектурі ядра та певній кількості ядер. Центральний процесор має лише кілька фізичних ядер в той час, коли графічний процесор може мати їх сотні (Мал. 1, джерело [15])[2][6][11].

Рішення, що базуються на технології графічного процесора загального призначення, стають щораз популярнішими, зокрема у США та Азії. Суперкомп'ютери з графічними процесорами загального призначення є дешевними порівняно з групами комп'ютерів, що мають лише центральний процесор. Наприклад, BNP Paribas Corporate та Investment Banking замінили 500 стандартних ядер центрального процесора лише двома графічного процесора загального призначення. Нова платформа економить електроенергію (комп'ютер з графічним процесором споживає 2кВт, 500 центральних процесорів споживали 25кВт). Нова архітектура може обробляти одну тисячу мільярдів обчислень за секунду. Поєднавши обчислювальну функцію та низьке споживання електроенергії, рішення на основі графічного процесора уможливлюють стократне збільшення кількості обчислень з розрахунку на Ват. [2][4][15].

Рішення на основі графічного процесора загального призначення є майбутнім IT. Дана технологія є важливою не лише для бізнесу, але й для дослідників. Ця інновація є обов'язковою. Без сучасних рішень європейські підприємства не можуть залишатись конкурентоспроможними на ринку. Технологія на основі графічного процесора загального призначення є не лише інновацією, це також спосіб підняти конкурентоспроможність на ринку.

Переклад виконано в Агенції перекладів PIO
www.pereklad.lviv.ua

GPGPU as an alternative IT investment

Katarzyna Mrowca

Management Faculty, Cracow University of Economics,
POLAND, Cracow, Rakowicka 27,
E-mail: mrowca.kasia@gmail.com

IT is evolving. The technology is becoming more and more available because of increasing accessibility, decreasing prices and user-friendly interface. Technology based on GPGPU (General Purpose Graphic Processing Unit) is an innovation which brings huge economical profits. In spite of this fact it is not popular in Europe. The article is focused on analysis of drawbacks and benefits of GPGPU computing. Moreover, in the article there are described examples of success implementations of GPGPU based solutions into a business. Aim of those deliberations is to prove a thesis that innovative GPGPU technology can bring significant profits. Research methods adopted in the article are literature analysis, observations and experiments made by computer simulations.

Keywords –I nnovation, GPGPU, HPC,

I. Introduction

IT is everywhere. It is hard to imagine companies without IT support. The implemented technology is basically the same – corporations and big companies are using similar systems, delivered mostly by the same vendors. Certainly, the companies are not able to work without IT, but does IT still create a competitive advantage? Can IT be an innovation?

First of all the company manager should answer the question if it is worthy to invest money into IT. If investment is needed, the second important issue is which technology chooses. Every investment should be preceded by economic analysis. In some cases it is better to not switch from old technology to new one, because it will not bring significant profits for company. Moreover, IT investment are risky, the implementation can take even few years and not always is successful. According to Standish Group research only 10% of IT investment was realized in planned time and expenses. More than one third of such projects were stopped [17]. However, the introduction of new IT solution into company is hazardous can be also profitable.

Recently, the need of computation large data sets is growing. In order to fulfill it, companies invest in own data centers or cloud computing. This paper will show alternative possibility – solutions based on graphic cards. This technology has high potential, is considerably cheap and not widely used in Europe. That is why, implementation of it can be profitable for company.

II. IT background

At the beginnings of IT the technology itself was extremely expensive. Moreover, computer literate staff was hard to find. The first super-computers in early 80's were worth more than 5 million dollars. One decade later the price decreased to 1 million $. Although the technology was more advanced, the value of a single machine fell down. Nowadays, the high performance computers can be bought for less than 100,000$ [6]. Furthermore, highly competent employees with excellent computer literacy are available on the labour market. Those factors create a good basis for investment. Surprisingly, in spite of the fact that the technology is accessible almost for everyone, its popularity is quite low.

It is easy to notice, that the high-tech technology nowadays is considerably cheap. The question is if it is still innovative enough to create a competitive advantage. The computers with more than one core, introduced at the beginning of the first decade of XXI century, have quickly become considerably cheap and they do not boost computing performance in a spectacular way. Although, the GPGPU (general-purpose computing on graphics processing units) technology is also not 'expensive' and gives huge opportunities, it is not popular. The market share of solutions based on GPU (Graphics Processing Unit) computing is quite small in Europe [12]. The GPGPU is not a new technology; it has been developed since 2000. In spite of the fact that it achieves enormous performance boost, it is not common. Switching from the CPU based technologies to GPU based solutions could bring some benefits in niche environment. [1][2].

III. Basic description of the technology

Graphics Processing Unit (GPU) computing or general-purpose computing on graphics processing units (GPGPU) is a new way of exploiting graphic card processors. Ordinary GPU processes graphic data only. A new generation unit is able to handle computation made by CPU typically. Such a type of GPU is produced by nVidia (CUDA) or ATI (Stream). The working concept of both is more or less the same, but construction and performance are completely different. One of the similarities is high performance, significantly higher than in case of CPUs. The GPU is able to process data faster because of core parallel architecture and the number of cores. The CPU has only a few physical cores while the GPU can have hundreds of them (Fig. 1, source [15])[2][6][11].

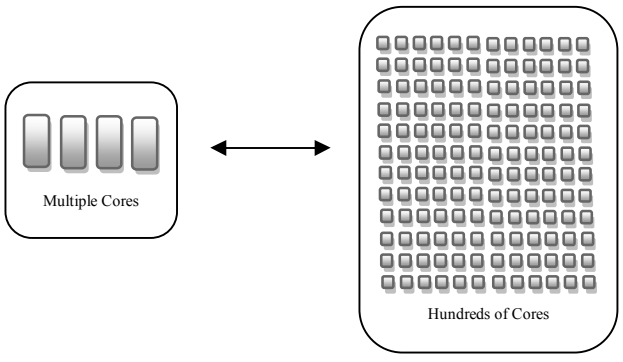

Fig. 1. Architecture comparison of CPU and GPU

It is worth mentioning that GPU computing does not use only the GPU to process data. The model assumes that GPU and CPU are cooperating. The part of application which is computationally-intensive is processed by GPU in order to boost performance. The sequential instructions from those applications run on the CPU. As a result, the computation is accelerated [15].

IV. Market competition

Two companies which are producing the GPGPU cards exist on the market: ATI(Stream) and nVidia(CUDA). Although the general concept of GPGPU is the same, the cards structure and design are not the same. Differences in the construction of cards impose the application of different main boards and other computer subassemblies. Therefore the software made for one graphic card is not compatible to other GPGPU. Even the programming languages used by software developers are dependent on a graphic card producer.

Although the technology of GPGPU seems to be 'the same', differences between products of ATI and nVidia are huge. Switching from one solution to another is impossible. Moreover, the software written for nVidia is useless for ATI and vice versa. In spite of the fact that all GPGPU cards can boost computing performance, the architectures should be used to different types of the calculations. One architecture is better for financial data modelling, other for bioinformatics computation. In order to get the best performance, one should choose the graphic card with the architecture suitable to the problem. It is not so easy at it seems to be. Despite the fact that one architecture can ideally match the numeric problem, there can be a problem with software required to make a computation. Actually, the nVidia solutions are more popular than ATI. The common software like MATLAB is able to use CUDA to boost computing performance. Moreover, at 391 Universities all over the world there are academic subjects about CUDA. Today the market race is won by nVidia, but ATI does not stay far behind [15].

V. GPGPU usage on CUDA example

GPGPU has been developing fast. The technology is adapted to boost computing performance in diverse areas, such as [15]:

- Bio-Informatics and Life Sciences,
- Computational Electromagnetics and Electrodynamics,
- Computational Finance,
- Computational Fluid Dynamics,
- Data Mining, Analytics and Databases,
- Imaging and Computer Vision,
- MATLAB Acceleration on Tesla and Quadro GPUs,
- Medical Imaging,
- Molecular Dynamics,
- Numerical Packages,
- Artificial Intelligence (neural nets)
- Weather, Atmospheric, Ocean Modelling, and Space Sciences,
- Others.

Each year technology enables faster and faster computing. According to some data from 2010, the GPU boosted performance even 149 times in comparison to CPU (fig 2, source [13]).

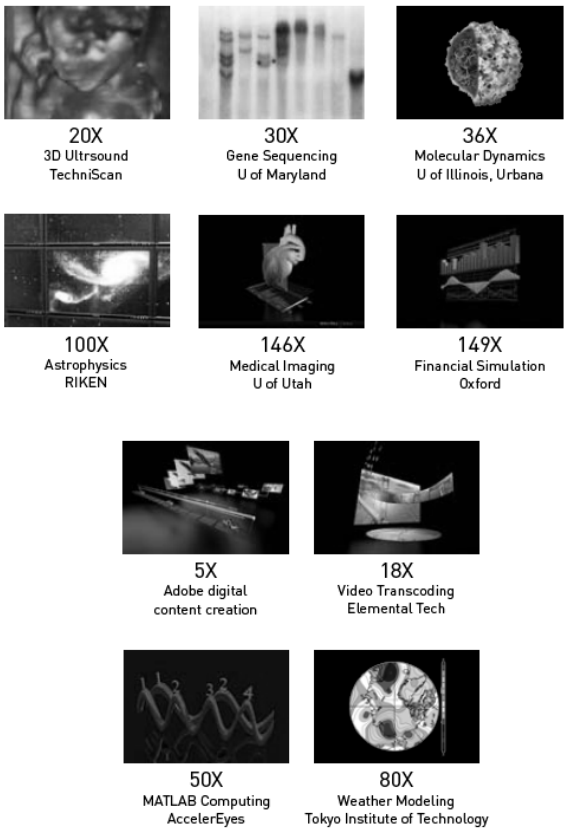

Fig 2. The computing performance boost

The decrease of time required to compute a certain problem depends on the kind of the task, model of the graphic card, the computer architecture, quality of software and quality of data. Apart from the problems mentioned above, GPU is generally efficient in computing [9]:

- Large matrix/vector operations,
- Protein Folding,
- Ray Tracing,
- Sequence Matching (Hidden Markov Models),
- Speech Recognition,
- Databases,
- Sort/Search,
- And many, many more.

Data parallelism is a key of GPGPU efficiency. The ability to use data parallelism means that huge number of data is executed during the computation in the same time. In order to achieve the parallelism the data should be independent (in each step in the computation). Usually, this approach requires redesigning of the existing algorithms or designing some new. To sum up, GPGPU ideal 'problem' is characterised by high parallelism, large data sets, no dependencies between data, high arithmetic intensity and continuous work without CPU usage [9].

VI. Business implementation

The solutions based on GPGPU technology are becoming more and more popular. Especially in the USA and Asia. The supercomputers with GPGPU are cheaper that clusters of computers with the CPU only. For example the BNP Paribas Corporate and Investment Banking replaced 500 standard CPU cores by only 2 GPGPU. A new platform brings electricity savings (the computer with GGPU consumes 2kW, the 500 CPUs were consuming 25kW). The new architecture is able to process one thousand billion calculations per second. Computing performance together with the power consumption, the GPU based solutions enable 100-fold increase in the number of calculation achieved per watt. The financial data can be processed by GPU platforms faster than on CPUs. Impressive performance boost is visible in computing the Monte Carlo algorithm and random numbers generations (those mathematical operations are useful in options pricing and risk analysis like VAR)[3][4][5][16]. The random number generation can be faster up to 50x (Fig. 3, source [14]) [11]. Another example of implementation of GPGPU into the finance sector is Hanweck Associates, LLC. The company specializes in investment and risk management. The mail offer of the company is recalculating options in real time. A few years ago, this task was done by 60 traditional servers (with CPU only). Now, the company is using only 12 CUDA graphic cards. It allows to process Volera analyses for the entire USA options market in real time[8]. Changing the technology to the GPGPU based brings cost savings on electricity, the hardware costs and data center- real estate[13][14].

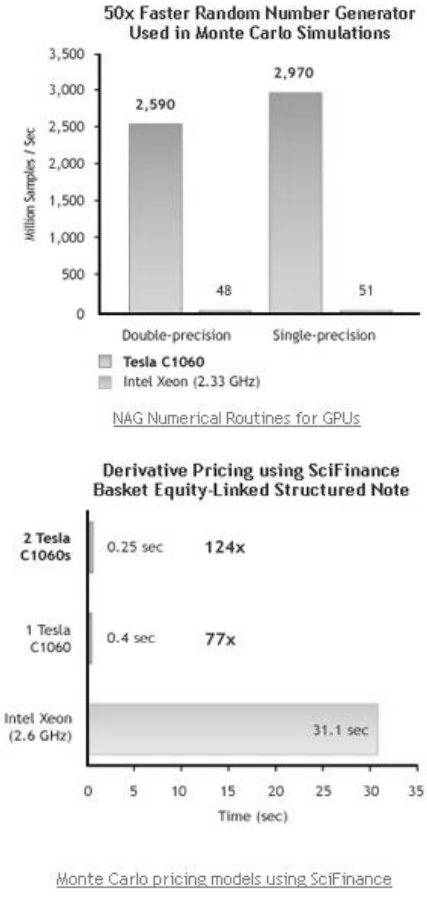

Fig 3. Financial data processing performance on GPGPU

VII. European Commission for High Performance Computing (HPC)

European Commission observes the problem of non-efficient development of HPC in Europe. The HPC technology is strongly explored in the USA and Asia even by small and medium enterprises [7]. The fear was confirmed by the research made by International Data Corporation. IDC published the report which focused on the meaning of HPC development to scientific and business competiveness. Unfortunately, Europe stays far behind the world HPC race. According to the report, from 2007 to 2009, share of worldwide investment in HPC declined from 34 to 25 percent. It is not a good result if we take into consideration the fact that the HPC market rockets each year. In order to improve the EU situation IDC recommends a long- term strategy which is supposed to[12]:

- Expand the number, size and access to HPC resources across the EU,
- Create a set of HPC exascale development lab/tested centres,
- Attract more students info scientific, engineering and HPC fields, and attract more experts from around the world to join EU scientific collaborations,
- Invest in development next-generation exascale software,
- Target a few strategic application areas for global leadership.

As a result of the activities mentioned above Europe should become one of the leaders of the HPC technology by 2020 year. Moreover, the plan will cause the following [12]:

- Europe would be recognized as the hotbed for new science and engineering research and innovation,
- The plan would preserve existing jobs and create many new jobs in both science and industry, and make national economies grow faster.

Moreover, the EU authorities were advised to improve HPC accessibility to small and medium companies. Without this, the competition on global stage is lost. The USA small and medium companies have already done that e.g: BMI, Intelligent Light, L&L Products or Swift Engineering. The companies implemented the innovative HPC technology which helped to adjust to the changing business surroundings. Moreover, the USA government supports the companies which invest in HPC solutions. Europe has to make serious movements in order to smooth the gap.

VIII. Conclusions

The GPGPU based solutions are the future of IT. The technology is significant not only for business but also for researchers. The innovation is a must. Without the up- to-date solutions European enterprises cannot stay competitive on the market. The GPGPU based technology is not only an innovation it is also way to increase competitiveness on the market. If correctly used, the HPC technology can bring huge profits for a company and for the European economy.

References

[1] David A. Patterson,John L. Hennessy, Computer organization and design: the hardware/software interface, Elseviewer, Canada 2009

[2] GPUs for General-Purpose Computing, Maximum PC, December 2008

[3] Greg N. Gregoriou, The VaR implementation handbook, The McGraw Hill Companies, USA 2009

[4] Naveen Singla, Michael Hall,Berkley Shands, Roger D. Chamberlain, Financial Monte Carlo Simulation on Architecturally Diverse Systems, http://sbs.wustl.edu/pubs/ shsc08.pdf

[5] Podlozhnyuk V., Harris M., Monte Carlo Option Pricing, nVidia http://developer.download.nvidia.com/compute/cuda/1_1/Website/projects/MonteCarlo/doc/MonteCarlo.pdf

Internet references

[6] Conway S., How Goes the Democratization of Supercomputing? http://www.scientificcomputing.com/articles-HPC-How-Goes-the-Democratization-of-Supercomputing-102810.aspx, date: 14.06.2011

[7] GPGPU implementation in finance section in Japan http://www.numtech.co.jp/files/documents/seminar/20101019/20101019E.pdf, date: 14.06.2011

[8] Hanwecassoc, LLC http://www.hanweckassoc.com/

[9] Houston M., Advanced Programming (GPGPU), graphics.stanford.edu/~mhoutson/public_talks/cs448-gpgpu.pdf, date: 12.06.2011

[10] HPC in Europe http://www.hpcuserforum.com/EU/ http://www.numtech.co.jp/files/documents/seminar/20101019/20101019E.pdf, date: 06.06.2011

[11] http://wallstreetandtech.com/it-infrastructure/showArticle.jhtml?articleID=215800756&cid=RSSfeed_WST_All, date: 06.06.2011

[12] IDC report http://www.hpcuserforum.com/EU/downloads/SR03S10.15.2010.pdf, date: 01.06.2011

[13] nVidia, finance http://www.nvidia.com/object/computational_finance.html, date: 06.06.2011

[14] nVidia, finance CUDA http://www.nvidia.com/object/cuda_finance.html

[15] nVidia, GPU computing http://www.nvidia.com/object/GPU_Computing.htmldate: 06.06.2011

[16] nVidia, Monte Carlo method http://developer.download.nvidia.com/compute/cuda/1_1/Website/projects/MonteCarlo/doc/MonteCarlo.pdf, date: 06.06.2011

[17] Standish group, http://www.standishgroup.com/, dnia 15.06.2011