

ГРИД ИНФРАСТРУКТУРИ. ТЕСТОВЕ НА INTEL XEON QUAD-CORE ПРОЦЕСОР В ГРИД СРЕДА

Ст. н.с. II, д-р Емануил Атанасов и ст.н.с. II, д-р Тодор Гюров
Институт за паралелна обработка на информацията (ИПОИ),
Българска академия на науките (БАН),
Акад. Г. Бончев, 25 А, София 1113, България
E-mails: {emanouil, gurov}@parallel.bas.bg

1. Какво е Грид

Необходимостта от изграждане на Грид инфраструктура е предизвикана от усилията на обществото, чрез използване на мощни изчислителни ресурси, да изследва:

- Фундаменталните свойства на материята;
- Геномите;
- Изменението на климата;
- Предсказване или анализ на природни бедствия (наводнения, земетресения) и промишлени аварии;
- Медицинска диагностика.

Все по-сложни сензори генерират огромни количества от данни, които служат на учени от целия свят. Техните изследванията са свързани с постоянно увеличаване на обема от данни и количество изчисления. Всичко това обуславя изграждането на Грид инфраструктурата за нуждите на науката.

За разлика от Глобалната Internet мрежа (World Wide Web), която осигурява пълен достъп до информация от източници с различно географско положение, Гридът е нова изчислителна инфраструктура, която осигурява достъп до изчислителни ресурси и данни, разпределени по цял свят. За рождена дата на системата, наречена Грид, се възприема септември 1997 г., когато се провежда семинар на тема "Изграждане на изчислителен Грид" в Аргонската Национална Лаборатория, Илинойс, САЩ. През 1998 година излиза и книга със заглавие "The Grid: Blueprint for a New Computing Infrastructure" [1] с автори Ян Фостър от Аргонската Национална Лаборатория и Карл Кеселман от Университета на Южна Калифорния. Тази книга често се нарича Грид библия.

Понятието «Грид» навлезе широко в компютърните науки. И все пак, какво е «Грид»? Изчислителна инфраструктура? Виртуална машина? Киберпространство? Ново поколение интернет? Първата широко известна дефиниция, бе дадена от Ян Фостер и Карл Кеселман в книгата им: *"Изчислителният грид представлява хардуерна и софтуерна инфраструктура, която осигурява зависим, съгласуван, разпределен и не-скъп достъп до компютърни ресурси."* В следващите си статии [2,3], те прецизират дефиницията, като подчертават, че предназначението на Грид пресмятането е *"осигуряване на съгласуван достъп до ресурси и решаване на задачи в*

динамични, мултиинституционални виртуални организации". Три са основните характеристики, които трябва да притежава една система, за да бъде наречена Грид, според Ян Фостър:

- ✓ да координира ресурси, които не са обект на централизиран контрол;
- ✓ да използва стандартни, отворени, универсални протоколи и интерфейси;
- ✓ да осигурява високо качество на услугите.

Съществуването на грид инфраструктурата се осмисля от приложенията. Естественният въпрос, който възниква, е кои научни области могат да се възползват от изчислителния грид. Краткият отговор на този въпрос е: всяка дисциплина, която понастоящем използва съвременни ресурси за научни изчисления.

2. EGEE Грид инфраструктура и приложения

Върху Грид инфраструктурата, поддържана от проекта EGEE (Enabling Grids for E-sciencE) [4], стартирал в 6-тата рамкова програма, работят грид приложения от областта на високите енергии, биомедицината, изчислителната химия, науките за земята, космическите изследвания и др.

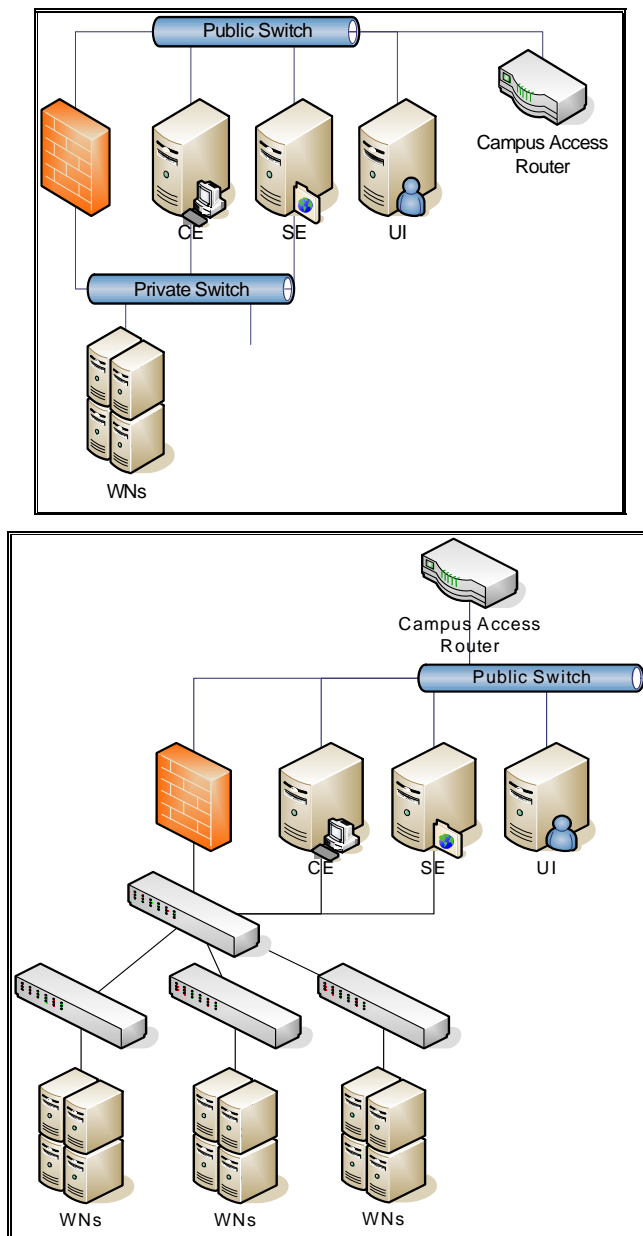
Все пак основните области на приложения са две, които направляват началната работа и удостоверяват функционалността на развиващата се Грид инфраструктура. Едната от тях е Large Hadron Collider (LHC) Computing Grid [5], която използва Грид инфраструктура, за да съхранява и анализира петабайти (10^{15} байта) реални и симулирани данни от експерименти в CERN (European Organization for Nuclear Research), в областта на физиката на високите енергии. Другата е Biomedical Grids, където няколко общности са изправени пред непосилни предизвикателства – например, издирването на данни в геномните бази данни или индексването на медицинските бази данни в болниците, които възлизат годишно на няколко терабайта данни за дадена болница.

За пълното натоварване на Грид инфраструктурата се използват така наречените "Data Challenges (DC)" [6,7], при които се изпращат огромно количество изчислителни задачи (jobs) от даден(и) потребител(и) в продължение на определено време (например един месец). Тези задачи изискват огромно количество изчислителен ресурс (CPU power) и/или произвеждат огромни количества данни, а също така се нуждаят и от големи обеми RAM памет. В крайна сметка се решава даден проблем на обществото. Ще отбележим, че последните DC по EGEE грид инфраструктурата бяха в биомедицината и бяха свързани с решаването на проблема с маларията в Африка и птичия грип. С пускането в действие на ускорителите в CERN се очаква EGEE грид инфраструктурата постоянно да бъде натоварена с обработка на данни, свързани с четирите разработени приложения в областта на високите енергии – ATLAS, CMS, LHCb и Alice [5].

Най-голямата международна Грид инфраструктурата, оперираща като 4-тото "чудо" в развитието на информационните технологии [8, 9], изградена и поддържана от проекта EGEE, се състои от около 250 грид кълстери, разположени в над 45 държави (Таблица 1).

Един грид кълстер се състои от *Computing Element* (CE), *Storage Element* (SE) и *Worker Nodes* (WNs). Основната изчислителна единица в кълстера е така-нареченият *Worker Node* (работен компютър), където се изпълняват изчислителните задачи и може да поддържа един или повече изчислителни процеса, докато SE разпределя задачите по WNs, а SE служи за съхранение на данните, (вижте Фигура 1). Има и

други услуги в грид инфраструктурата, наречени *core services*, които обслужват повече от един клъстер или цялата грид инфраструктура, но върху тяхното предназначение няма да се спираме.



Фигура 1: Примери на Грид клъстери и начин на свързаност на отделните компоненти

Таблица 2: Разпределение на Грид клъстерите в 12-те района, в които са групирани всичките 48 държави по света

РЕГИОН	Брой страни	Брой клъстери	Брой Процесори
CERN	5	12	6400
Великобритания/Ирландия	2	25	8384
Франция	2	12	7238
Германия/Швейцария	2	15	4413
Италия	1	34	4341
Северна Европа	9	30	3289
Югоизточна Европа	8	38	2727
Централна Европа	7	24	2588
Югозападна Европа	2	18	1938
Азия -Америка	8	20	1884
Русия	2	15	738
Общо	48	243	44040

Грид клъстерите в EGEE грид инфраструктурата обединяват общо над 40000 процесора с над 7 Пета-байта (10^{15}) ресурс за съхранение на данни. Тези клъстери са свързани с високопроизводителни мрежови връзки, предимно чрез европейската научно-изследователска мрежа (GEANT), като за повечето държави капацитетът достига 10 Гигабита в секунда.

България участва в този проект от самото му начало, като в момента са включени 5 български клъстера с общо около 160 процесора.

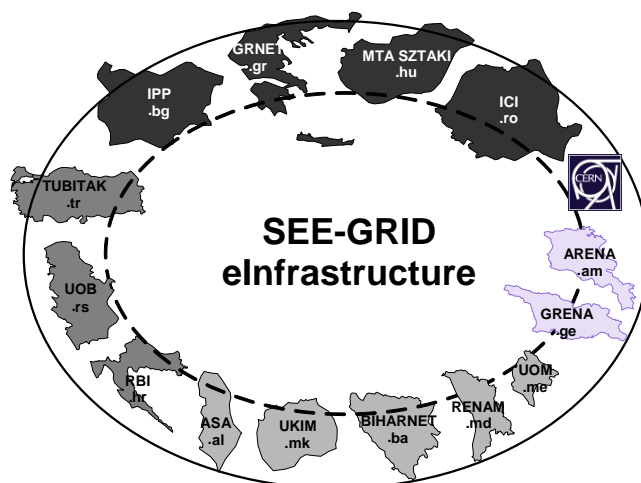
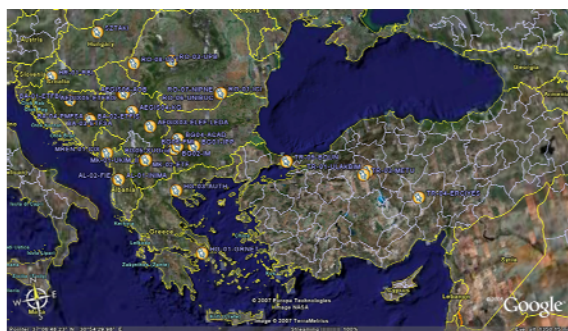
Макар че Грид софтуерът може да бъде компилиран за различни видове процесори, в EGEE стандартът е установен на процесори минимум Pentium III и съвместимите с тях, като в момента повечето от процесорите са с тактова честота над 2 GHz и поддържат 64-битовите разширения на i386 архитектурата. Като операционната система се използва Scientific Linux версия 3.0 или по-новата версия 4.5, която е безплатна и с отворен код.

За да работят грид услугите е необходимо да се инсталира съответния грид софтуер. В рамките на проекта EGEE се разработва и поддържа gLite софтуера, който също е безплатен за ползване и с отворен код. Актуалната версия на gLite в момента е 3.1.

3. SEE-GRID инфраструктура и българския опит

SEE-GRID инфраструктурата е изградена и се поддържа по проекта SEE-GRID (South Eastern European GRid-enabled eInfrastructure Development) [10]. Стартирайки в 6-тата рамкова програма, проектът преминава през две фази: през първите две години се утвърждава грид инфраструктурата в региона и се деплоят първите две грид приложения, докато през втората фаза се наблюдава разширяването ѝ, устойчивостта ѝ, и пълноценното натоварване с разнообразни грид приложения. SEE-GRID консорциумът се координира от Greek Research&Technology Network (GRNET) и включва 12 изследователски институции от всички Балкански страни, Унгария, както

и Европейския Център за Ядрена Физика (CERN), (вижте Фигура 2). Очаква се при изпълнението на проекта SEE-GRID-SCI, който ще стартира по 7-мата рамкова програма от 2008, регионалната Грид инфраструктура да се разшири и в Армения и Грузия.



Фигура 2: SEE-GRID инфраструктурата и страните, които я поддържат

В края на 2007 година SEE-GRID инфраструктурата разполага с 35 грид клъстери, над 1150 CPUs, 42 TB диково пространство за съхранение на данни [11]. Софтуерът, който се използва е същият, който е инсталиран и върху EGEE грид инфраструктурата. В консорциума, България се представлява от института по паралелна обработка на информацията от Българската академия на науките (ИПОИ-БАН), като колективът предоставя и поддържа грид ресурси: над 150 WNs, основни грид сървизи (AMGA, WMS, FTS, BDII), мониториране на работещите клъстери и отчетността на инфраструктурата, чрез Nagios и RGMA в on-line режим. Разработват се и се тестват над 10 български грид приложения, като водещо от тях е: SALUTE (Stochastic ALgorithms for Ultra-fast Transport in sEmiconductors) [12]. Това приложение обединява Монте-Карло и квази-Монте Карло алгоритми, разработени от групата по Монте Карло методи в ИПОИ-БАН и изследва различни квантови

ефекти по време на процеса на релаксация, дължащ се при електроно-фононното взаимодействие в полупроводници при наличие на електрично поле [13,14]. Освен това колектива продължава да тества и разгръща допълнителни софтуерни разработки върху нов хардуер с цел подобряване на грид инфраструктурата.

4. Тестване на Quad-Core Intel® Xeon® 5300 Processor в Грид среда

4.1. Хардуер

В рамките на нашите тестове ни беше предоставена от фирма ICOMM компютърна конфигурация, състояща се от:

- Intel® Server Board S5000PAL ;
- Quad-Core Intel® Xeon® 5300 Processor 2.4 GHz / s771 – (2 броя)
- FAN-RK_0108 2U + Actice Head Sink-LGA771 Dual Core Xeon - (2 броя)
- Memory 4 GB – Kingston 1 GB KVR533D2D8F4 FB DIMM - (4 броя)
- Hard Disk – 80 GB Seagate SATA2/7200/8MB
- Intel SC5299BRP
- Rack Conversion Kit for Intel

4.2. Операционна система, Грид софтуер

- За да инсталираме грид софтуера, започнахме с инсталация на Scientific Linux версия 4.3 в 64-битов режим. Успяхме да намерим подходящо ядро, което да разпознае новия вид процесори, така че машината логически се разпозна като 8-процесорна.
- За инсталация на gLite софтуера използвахме 32-битовите пакети (rpm) за да имаме съвместимост с приложния софтуер, който предимно е 32 битов все още. Инсталацията завърши успешно, така че успяхме да включим предоставената машина като WorkerNode (с 8 процесора!) към един от нашите грид клъстери.
- Машината успешно поемаше изчислителни задачи, като опитахме изпълнението на няколко задачи от Biomedical data challenges, които работиха в 32 битов режим.
- Тъй като машината се разпознава като паралелен компютър с 8 процесора решихме да инсталираме софтуер за паралелна обработка и да проведем поредица от стандартни тестове на паралелната производителност.

4.3. Тестване:

- Една утвърдена система за тестване на производителността в паралелен режим е NASA Parallel Benchmark [15], която е разработена от Американската космическа агенция.
- Тези тестове целят да симулират реалните паралелни приложения и да тестват скалируемостта на паралелните системи при близки до реалните условия.
- За реализация на MPI комуникациите беше използван LAM MPI, който не е оптимално решение за една SMP система, но показва най-добра производителност от достъпните ни библиотеки с отворен код.

Table 2. Резултати от тестването на NASA parallel benchmark системата с описаната 2- процесорна конфигурация Quad-Core Intel® Xeon® Processor 2.4 GHz / s771. Резултатите са в Million Elementary Operations per seconds (Mops) на процес.

Име на тест/Брой ядра	1	2	4	8
EP	21.75	19.69	18.23	17.90
LU	845.36	794.33	746.62	644.52
FT	963.92	541.12	385.90	246.41
IS	46.01	33.08	23.50	15.05

- Повечето (над 90%) от изчислителните задачи, изпълнявани върху EGEE грид инфраструктурата, са последователни, т.е. грид приложенията изпълват серии от задачи за да получат необходимите резултати. Голяма част от тях са Монте Карло симулации. От горните тестове такъв е **EP** тестът, при който има много малко комуникации между процесите, и както виждаме, ефективността на паралелно изпълнение е близка до 100% ($17.90/21.75=82\%$). При последователност от задачи липсва комуникация между процесите. Това позволява едновременно да се изпълняват 8 задачи от едно или повече грид приложения върху така конфигурираната система.
- **LU** тестът също е много показателен, тъй като при него се измерва скорост на решаване на системи линейни уравнения, която е определяща при формиране на списъка с 500 най-мощни суперкомпютри (www.top500.org). Постигнатата ефективност от 76% е много добра, като се има предвид, че използваният код не е напълно оптимизиран. По този начин се получава по-добра представа за производителността в реални условия, тъй като в повечето случаи разработчиците не разполагат с време или познания, за да постигнат оптималната производителност.
- Системата премина успешно всички тестови примери, като числените резултати бяха коректни. Тестовете *FT* и *IS* са по-специфични и натоварват в по-голяма степен комуникациите между процесите, което намалява производителността.
- Тествахме подобрението, което се постига с използване на 64-битовите разширения, като компилирахме едно наше грид приложение в областта на нано-технологиите SALUTE (Stochastic Algorithms for Ultra-fast Transport in sEmiconductors), и постигнахме подобрение от около 10%. Очаква се все повече от грид приложенията да се възползват от тези възможности, поради по-добрата производителност и допълнителните възможности за работа с големи обеми памет.

5. Заключение

Новият четириядрен процесор Quad-Core Intel® Xeon® 5300 Processor предлага:

- Съвместимост с изискванията на Грид софтуера, в 64 и 32 битов режим;
- Висока производителност и добра скалируемост в Грид среда;
- Тестваната система може да се използва и за изпълнение на паралелни задачи с използване на MPI и OpenMP стандарти, тъй като един

компютър с два процесора Quad-Core Intel® Xeon® 5300 Processor може да играе ролята на 8-процесорна паралелна машина. Производителността в такъв случай ще зависи от използването на високопроизводителна мрежова връзка.

Основно предимство на новия процесор Quad-Core Intel® Xeon® 5300 Processor е събирането на същата изчислителна мощност върху по-малко пространство, при което се постига по-добра управляемост на системата и се реализират икономии. При оборудването на бъдещите грид кълъстери все по-голяма роля ще играят изискванията за компактно разполагане на компютрите и за икономия на електрическа енергия, без компромис с производителността.

Литература:

- [1] Ian Foster, Carl Kesselman, "*The Grid: Blueprint for a New Computing Infrastructure*", Morgan-Kaufman, 1998.
- [2] Ian Foster, Carl Kesselman, Steve Tuecke, "The Anatomy of the Grid", *International J. Supercomputer Applications*, 15(3), 2001.
- [3] Ian Foster, "What is the Grid? A three Point Checklist", *GridToday*, July 20, 2002, (see also: <http://www.globus.org/alliance/publications/papers.php>).
- [4] Official EGEE website: www.eu-egee.org
- [5] Large Hadron Collider: <http://lhc.web.cern.ch/lhc/>
- [6] Data Challenges: <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/DC/>
- [7] Biomed Data Challenges: <http://egee-na4.ct.infn.it/biomed/egee-na4-biomed-050905.html>
- [8] Business Technology Leadership: <http://www.cio.com/article/135700>
- [9] PC World: <http://www.pcworld.bg/?call=USE~home;&page=paper&n=4609&pn=4>
- [10] SEE_GRID project: <http://www.see-grid.eu/>
- [11] Gstat monitoring website: <http://goc.grid.sinica.edu.tw/gstat/seegrid>
- [12] E. Atanassov, T. Gurov, A. Karaivanova, and M. Nedjalkov, Monte Carlo Grid Application for Electron Transport, *LNCS 3993, Springer* (2006) 616-623.
- [13] E. Atanassov, T. Gurov, and A. Karaivanova, Salute Application for Quantum Transport. New Grid Implementation Scheme, in: *Proceedings of the Spanish conference on e-Science Grid Computing, March 1-2 2007*, Madrid, Spain, pp. 23-32.
- [14] E. Atanassov, R. Georgieva, T. Gurov, S. Ivanovska, A. Karaivanova and M. Nedjalkov, New algorithms in the Grid Application SALUTE, in: *Proceedings of the 30th International Convention, MIPRO 2007*, May 21-25, Opatija, Croatia, pp. 217-222.
- [15] NASA Parallel Benchmarks, <http://www.nas.nasa.gov/Resources/Software/npb.html>