# Máquinas mais rápidas do mundo
# Computação Paralela e Distribuída
# MAC 5742

Jorge Augusto Melegati Gonçalves
Nº USP 5696267

Junho de 2015

# Sumário

**1 Introdução** **2**

**2 Listas** **3**
2.1 *Top500* . . . . . 3
2.1.1 LINPACK *benchmark* . . . . . 4
2.2 *Graph500* . . . . . 5
2.3 *Green500* . . . . . 5

**3 Arquiteturas** **6**
3.1 Nós . . . . . 6
3.2 Inter-nós . . . . . 6

**4 Software** **8**
4.1 Sistema Operacional . . . . . 8
4.2 Programação . . . . . 8

**5 Conclusão** **9**
5.1 Leitura recomendada . . . . . 9
5.2 Curiosidade . . . . . 9

**Referências** **11**

# Capítulo 1

# Introdução

Supercomputadores são máquinas com um altíssimo nível de computação. Sua história começa na década de 1960 e tem Seymound Cray (1925-1996) como um dos seus protagonistas. Cray, ainda na CDC (*Control Data Corporation*) foi o responsável pelo desenvolvimento do CDC 6600 [8] , um dos primeiros supercomputadores, e sua empresa Cray Research, criada em 1972, dominou o mercado de supercomputação até 1990 e está presente até hoje.

Estatísticas referentes a essas máquinas é de interesse para fabricantes, usuários e potenciais usuários pois permitem a colaboração, a troca de informações e o melhor entendimento do mercado de computação de alta perfomance [3]. Nesse contexto, surgiu a lista *Top500*. Entretanto, o conceito de velocidade está totalmente relacionado ao tipo da aplicação, por isso, outras listas existem. No capítulo 2 são apresentadas as principais listas elencando os computadores mais rápidos do mundo com enfoque principal na lista *Top500*.

A seguir, será discutido as peculiridades dessas máquinas. No capítulo 3 é discutido as diferentes arquiteturas utilizadas na criação delas tanto na construção dos nós de computação quanto na interligação entre eles. No capítulo 4, são apresentados pontos relacionados ao *software* utilizado nessas máquinas desde sistema operacionais até às linguagens de programação. Enfim, no capítulo 5, é feita a conclusão do trabalho incluindo algumas curiosidades acerca do tema.

# Capítulo 2

# Listas

## 2.1 *Top500*

A lista *Top500*[1] elenca as máquinas mais rápidas do mundo disponíveis comercialmente. Isso exclui máquinas à disposição de governos ou empresas que não sejam divulgadas por motivo de segurança nacional ou segredo industrial. A definição de *rápido* utilizado nessa lista é a capacidade de execução, medida em FLOPS (*Floating-point operations per second*, usando o benchmark *LINPACK*, que será discutido em mais detalhes na subseção 2.1.1.
A lista é divulgada duas vezes ao ano (nos meses de Junho e Novembro) desde Junho de 1993. Ela é compilada pelos seguintes pesquisadores:

**Erich Strohmaier,** NERSC/Lawrence Berkeley National Laboratory;

**Jack Dongarra,** University of Tennessee, Knoxville;

**Horst Simon,** NERSC/Lawrence Berkeley National Laboratory;

**Martin Meuer,** Prometeus;

**Hans Meuer,** University of Mannheim, Alemanha (de 1993 até a sua morte em 2014).

A tabela 2.1 apresenta os cinco primeiros colocados na lista de Novembro de 2014 (última disponível até o momento da escrita deste trabalho). Nela já é possível perceber algumas características recorrentes na lista:

**Fabricantes:** é clara a predominância da IBM e da Cray Research como fabricantes dessas super-máquinas;

**Co-processadores:** é muito comum a presença de co-processadores sejam processadores gráficos (por exemplo, NVidia) ou outros como o Intel Xeon Phi;

[1] Disponível em *http://www.top500.org*

**Arquitetura:** é de um de dois tipos: MPP (Massivelly Parallel Processor) ou Cluster.

**Capacidade de processamento:** o primeiro lugar da lista (Tiahne-2) obteve cerca de 33 PFLOPS de processamento, como comparação um computador com um Intel Core i5 2500K, possui cerca de 10GFLOPS. A lista atual possui 50 máquinas com mais de 1 PFLOPS.

Tabela 2.1: Primeiras posições da lista *Top500*

| # | Nome | Fabricante | País | Total Cores | Rmax (TFLOPS) | Potência (kW) | Mflops por Watt | Arquitetura |
|---|---|---|---|---|---|---|---|---|
| 1 | Tianhe-2 (MilkyWay-2) TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P | NUDT | China | 3M | 33.826 | 17808 | 1901,54 | Cluster |
| 2 | Titan Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x | Cray Inc. | United States | 560K | 17.590 | 8209 | 2142,77 | MPP |
| 3 | Sequoia BlueGene/Q, Power BQC 16C 1.60 GHz, Custom | IBM | United States | 1,5M | 17.173 | 7890 | 2176,58 | MPP |
| 4 | K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect | Fujitsu | Japan | 705K | 10.510 | 12659,89 | 830,18 | Cluster |
| 5 | Mira BlueGene/Q, Power BQC 16C 1.60GHz, Custom | IBM | United States | 786K | 8.586 | 3945 | 2176,58 | MPP |

### 2.1.1 LINPACK *benchmark*

O LINPACK *benchmark* relata a perfomance na solução de um sistema linear de equações com matrizes densas utilizando operações de ponto-flutuante de 64-bits [7]. O tamanho das matrizes pode ser 100, 1000 ou ter tamanho arbitrário. O *benchmark* é composto por duas funções: DGEFA, que decompõe a matriz utilizando pivoteamento parcial, e DGESL, que utiliza a decomposição da etapa anterior para resolver o sistema de equações.
A decomposição realizada se baseia na decomposição LU na qual uma matriz A não-singular é decomposta em duas matrizes triangulares L e U (inferior e superior, respectivamente). Por exemplo, para matrizes 3x3, a decomposição seria da forma:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$$

Assim, o sistema linear $Ax = b$ se torna $LUx = b$ o que é bem simples computacionalmente uma vez que L e U são matrizes triangulares. Para obter a performance contida na lista *Top500*, o tamanho da matriz é escolhido

de forma a maximizar a performance da máquina [3].

## 2.2 *Graph500*

A lista *Top500* se baseia na capacidade de computação mas nem todas as aplicações podem aproveitar dessa capacidade. Muitas aplicações são *data-intensive* e a análise da capacidade de máquinas em executá-las não era bem feita com os *benchmarks* utilizados, como o *LINPACK*. Dessa crítica, surgiu a lista *Graph500*[2]. A lista altera as posições mas as máquinas no topo são praticamente as mesmas do *Top500*. A Tiahne-2, primeiro lugar das mais rápidas, fica em sexto; o primeiro lugar da *Graph500*, a *Sequoia* é o terceiro na *Top500*, a segunda colocada é o *K Computer*, quarto na outra lista.

## 2.3 *Green500*

A lista *Green500*[3] tem como objetivo elencar os supercomputadores com melhor eficiência energética. Para isso, é utilizada a métrica de FLOPS por watt, ou seja, operações por segundo por consumo de energia. Esta lista apresenta grandes diferenças em relação *Top500*, conforme tabela 2.2 (versão de Novembro de 2014).

| **Posição** | **Nome** | **Mflops per Watt** | **Posição Top500** |
|---|---|---|---|
| 1 | ASUS ESC4000 FDR/G2S<br>Intel Xeon E5-2690v2 10C 3GHz<br>Infiniband FDR, AMD FirePro S9150 | 5271,8142 | 168 |
| 2 | ExaScaler 32U256SC Cluster<br>Intel Xeon E5-2660v2 10C 2.2GHz<br>Infiniband FDR, PEZY-SC | 4945,625592 | 369 |
| 3 | LX 1U-4GPU/104Re-1G Cluster<br>Intel Xeon E5-2620v2 6C 2.100GHz<br>Infiniband FDR, NVIDIA K20x | 4447,584063 | 392 |
| 4 | Cray CS-Storm<br>Intel Xeon E5-2660v2 10C 2.2GHz<br>Infiniband FDR, Nvidia K40m | 3962,73013 | 361 |
| 5 | Dell T620 Cluster<br>Intel Xeon E5-2630v2 6C 2.600GHz<br>Infiniband FDR, NVIDIA K20 | 3631,864623 | 241 |

Tabela 2.2: Primeiras posições da lista *Green500*

[2]Disponível em *http://www.graph500.org*
[3]Disponível em *http://www.green500.org*

# Capítulo 3

# Arquiteturas

Os supercomputadores mais rápidos do mundo são, na realidade, multicomputadores utilizando nós de computação interligado através de uma rede de intercomunicação. Segundo o tipo de rede de interconexão, temos 2 arquiteturas presentes na lista: MPP (*Massively Parallel Processors*) e *Clusters*. Os nós de computação diferem por possuir ou não co-processadores e por seu tipo: seja um GPGPU (co-processador gráfico de propósito geral) ou outro como, por exemplo, um Intel Xeon Phi o que será discutido na seção seguinte.

## 3.1 Nós

A maioria das máquinas mais rápidas do mundo implementam em seus nós de computação (ou mesmo nós de outros subsistemas) arquiteturas chamadas heterogêneas. Nesse modelo, utilizam-se além dos CPUs principais, aceleradores, principalmente GPGPUs, para propiciar performances superiores em aplicações paralelas. Como apontado por Chien et al. [1], é um largo consenso que a heterogeneidade tem o potencial de aumentar a performance e diminuir o consumo de enrrgia.
Segundo Liao et al. [6], a arquitetura dos nós de computação utilizado no supercomputador Tiahne-2 (primeiro lugar na lista *Top500* de Novembro de 2014) é chamada de *neo-heterogênea* uma vez que ele é composto por processadores Intel Xeon e co-processadores Intel Xeon Phi, possuindo diferentes ULAs mas com o mesmo conjunto de instruções.

## 3.2 Inter-nós

Dongarra et al. [2] discutem a nomenclatura dada aos diferentes tipos de supercomputadores. Neste trabalho, vamos nos ater nas duas arquiteturas encontradas na lista *Top500* apesar dos nomes utilizados não serem completamente aceitos como apontado pelos próprios autores.
Assim, *commodity cluster* é "um computador paralelo composto exlusivamente por subsistemas de computação *commodity* e redes comerciais tais

que os nós de computação são desenvolvidos e empregos em configurações *stand-alone* para largo (ou até mesmo massivos) mercados comerciais, e a rede é dedicada para uso privado do cluster (não-mundial)."
A definição de MPP é um pouco mais confusa sendo que os autores concluem "poder significar um sistema de memória distribuída, um grande sistema de memória compartilhada com ou sem coerência de cache, um grande sistema vetorial e assim por diante: cobre diversas categorias e esconde muitas diferenças salientes para ser uma descrição útil."

# Capítulo 4

# Software

## 4.1 Sistema Operacional

A predominância do Linux é clara na lista das máquinas mais rápidas. Diversas distribuições são encontradas, entre elas: Kylyn Linux (distribuição chinesa), Cray Linux Environment, SUSE, Red Hat entre outros. Se levarmos em conta também os sistemas operacionais Unix, teremos, praticamente, a lista completa sendo que apenas 1 supercomputador da lista utiliza sistema Windows.

## 4.2 Programação

Supercomputadores trazem grandes desafios para a utilização de todo o seu potencial. Segundo Latsis e Levin [5], no início, imaginava-se que a arquitetura dos novos supercomputadores seria escondida por trás de compiladores desenvolvidos especialmente para essas máquinas, ou seja, a arquitetura estaria escondida do programador da aplicação. Entretanto, isso é distante da realidade: os autores citam o caso das arquitetura de paralelização massiva sem uma memória comum a todos os nós que predominou por mais de 20 anos. Apesar disso, o problema de construir programas otimizados para essa arquitetura automaticamente não foi solucionado e os programadores são obrigados a desenvolver aplicações paralelas praticamente de forma manual.
Nas descrições encontradas sobre os supercomputadores, a programação é descrita a partir do uso das bibliotecas OpenMP, para a programação paralela no nó, e o OpenMPI, para a programação distribuída inter-nós. Na descrição do Tiahne-2 [6], os autores relatam a utilização de programação própria para os aceladores. Essa prática deve ser comum nas máquinas que possuem co-processadores para obterem o máximo de sua capacidade.

# Capítulo 5

# Conclusão

Este texto apresentou uma visão geral das máquinas mais rápidas do mundo. As primeiras posições da lista mais utilizada, *Top500*, foram apresentadas além da discussão do *benchmark* utilizado. Outras listas com enfoques diferentes (aplicações *data-intensive* e perfomance energética) foram brevemente apresentadas. A seguir, algumas peculiaridades dessas máquinas foram discutidas, primeiro relativas ao *hardware* e depois *software*, incluindo como é realizada a programação de aplicações para essas máquinas.

## 5.1 Leitura recomendada

A literatura, tanto quanto científica quanto não-científica, na área é bem vasta. Durante a pesquisa deste trabalho, diversos artigos interessantes foram encontrados mas podem não ter sido aproveitados no texto. Assim, essa seção apresenta algumas dessas fontes pois apresentam informações interessantes.
Para começar, é comum aparecerem artigos relacionados aos primeiros lugares das listas dos mais rápidos computadores. Em [6], tem-se a descrição do primeiro lugar do *Top500*, o Tiahne 2 incluindo a descrição dos seus subsistemas: computação, armazenamento, comunicação, monitoramento e diagnóstico e serviço. Diversos números são apresentados que dão uma dimensão do tamanho da máquina, notadamente que ele é composto por 125 *racks* de computação. Em [9], é descrito o K Computer, 4º lugar na lista, localizado no Japão. É interessante a preocupação com o consumo de energia. Em [4], é apresentada o *chip* IBM BlueGene Q, um *system-on-a-chip* utilizados no 3º e 5º lugares na lista.

## 5.2 Curiosidade

Uma curiosidade acerca da lista *Top500* é como estão classificados os supercomputadores brasileiros. Na lista atual, o melhor brasileiro colocado (124º)

está no SENAI/CIMATEC, foi desenvolvido pela SGI e, no LINPACH *benchmark*, chegou a 405 TFLOPS. A Petrobras possui dois computadores na lista (228º e 459º) sendo ambos desenvolvidos pela Itautec. O supercomputador do INPE completa os brasileiros na lista na 281º posição.

## Referências


[1] Chien, A. a., Snavely, A., and Gahagan, M. 10??10: A general-purpose architectural approach to heterogeneity and energy efficiency. *Procedia Computer Science 4* (2011), 1987–1996.

[2] Dongarra, J., Sterling, T., Simon, H., and Strohmaier, E. High-performance computing: Clusters, constellations, MPPs, and future directions. *Computing in Science and Engineering 7*, 2 (2005), 51–59.

[3] Dongarra, J. J., Meuer, H. W., Strohmaier, E., and Others. TOP500 supercomputer sites. *Supercomputer 13* (1997), 89–111.

[4] Haring, R., Ohmacht, M., Fox, T., Gschwind, M., Satterfield, D., Sugavanam, K., Coteus, P., Heidelberger, P., Blumrich, M., Wisniewski, R., Gara, A., Chiu, G., Boyle, P., Chist, N., and Kim, C. The IBM blue gene/Q compute chip. *IEEE Micro 32*, 2 (2012), 48–60.

[5] Latsis, a. O., and Levin, V. K. Prospects of supercomputer engineering development (based on lecture materials from MPAMCS-2012 in Dubna, August 27, 2012.). *Mathematical Models and Computer Simulations 6*, 3 (2014), 256–261.

[6] Liao, X., Xiao, L., Yang, C., and Lu, Y. MilkyWay-2 supercomputer: System and application. *Frontiers of Computer Science 8*, 3 (2014), 345–356.

[7] Padua, D. *Encyclopedia of parallel computing*, vol. 4. Springer Science & Business Media, 2011.

[8] Thornton, J. E. The CDC 6600 Project. *Annals of the History of Computing 2*, 4 (1980).

[9] Yamamoto, K., Uno, A., Murai, H., Tsukamoto, T., Shoji, F., Matsui, S., Sekizawa, R., Sueyasu, F., Uchiyama, H., Okamoto, M., Ohgushi, N., Takashina, K., Wakabayashi, D., Taguchi, Y., and Yokokawa, M. The K computer operations: Experiences and statistics. *Procedia Computer Science 29* (2014), 576–585.