

**Mestrado em Engenharia Informática**

*Engenharia dos Sistemas de Computação - 2014/2015*

Testes de Desempenho E/S (IOZONE)

5 de Maio de 2015

**Resumo**

Este trabalho demonstra o resultado do Benchmark iozone em vários nós do SeARCH no Departamento de Informática da Universidade do Minho. Desta forma será possível analisar os recursos de armazenamento com os testes intensivos que o IOZONE providencia.

**Fábio Gomes** pg27752

# **Índice**

[**Índice** 2](#_Toc415268099)

[Introdução 3](#_Toc415268100)

[NAS Parallel Benchmarks (NPB) 4](#_Toc415268101)

[Caracterização do Sistema 5](#_Toc415268102)

[Compilação, Versões e Parâmetros 6](#_Toc415268103)

[gcc / mpicc / gfortran 6](#_Toc415268104)

[Icc / ifort 6](#_Toc415268105)

[Serial 6](#_Toc415268106)

[OpenMP 6](#_Toc415268107)

[MPI 6](#_Toc415268108)

[Benchmark EP 7](#_Toc415268109)

[*Análise da Duração* 7](#_Toc415268110)

[*Análise da Memória* 10](#_Toc415268111)

[*Análise da Rede* 11](#_Toc415268112)

[*Análise de Acessos ao Disco* 11](#_Toc415268113)

[*Conclusão* 11](#_Toc415268114)

[Benchmark FT 12](#_Toc415268115)

[*Análise da Duração* 12](#_Toc415268116)

[*Análise da Memória* 14](#_Toc415268117)

[*Conclusão* 14](#_Toc415268118)

[Benchmark IS 15](#_Toc415268119)

[*Análise da Duração* 15](#_Toc415268120)

[*Análise da Memória* 17](#_Toc415268121)

[*Conclusão* 17](#_Toc415268122)

[Benchmark SP 18](#_Toc415268123)

[*Análise da Duração* 18](#_Toc415268124)

[*Análise da Memória* 20](#_Toc415268125)

[*Conclusão* 20](#_Toc415268126)

[Scripts, Diretorias e Gráficos 21](#_Toc415268127)

[Análise Final de Resultados e Conclusão 23](#_Toc415268128)

# Introdução

O problema que nos foi apresentado consiste em analisar a performance de um sistema de forma a poder saber as capacidades e limites operacionais para as aplicações que irão ser desenvolvidas para aquela infraestrutura.

Foi proposto que os testes incidissem em diferentes classes com diferentes compiladores, como o *icc* e *gcc* (versão 4.4.6 e 4.9.0), número de processos (*MPI*) e threads (*OpenMP*) e ainda para as versões SERIAL, OpenMP e MPI para cada Benchmark.

Os dados que o NPB devolve não são suficientes e para tal foi necessário recorrer a ferramentas externas para obter informações extras ao longo do tempo de execução. Especificamente o derivado *dstat* para a alocação de memória, acessos ao disco e dados enviados/recebidos na rede; o *mpstat* para a utilização dos diferentes processadores.

Com estes dados recolhidos foi possível gerar gráficos para comparar as diferentes combinações de compiladores e número de processos e threads.

# NAS Parallel Benchmarks (NPB)

São um pequeno conjunto de programas destinados a ajudar a avaliar o desempenho dos supercomputadores paralelos. Os *Benchmarks* são derivadas de aplicações da dinâmica de fluido computacional (CFD) e consistem em cinco *kernels* e três *pseudo-aplicações* (NPB 1). O pacote de Benchmark foi estendido para incluir novos pontos de referência para malha adaptativa não-estruturada, I/O, aplicações multi-zona, e grelhas computacionais paralelas. Os tamanhos dos problemas no NPB são predefinidos e indicados com diferentes classes (A,B,C,D,E,F ou S,W). Implementações de referência de NPB estão disponíveis em modelos de programação mais usadas como MPI e OpenMP (NPB 2 e NPB 3), sendo estas as versões que mais nos interessam.

Há diferentes especificações de Benchmark:

* 5 kernels
  + IS - Integer Sort, random memory access
  + EP - Embarrassingly Parallel
  + CG - Conjugate Gradient, irregular memory access and communication
  + MG - Multi-Grid on a sequence of meshes, long- and short-distance communication, memory intensive
  + FT - discrete 3D fast Fourier Transform, all-to-all communication
* 3 pseudo-aplicações
  + BT - Block Tri-diagonal solver
  + SP - Scalar Penta-diagonal solver
  + LU - Lower-Upper Gauss-Seidel solver

E classes de tamanho:

* S: small for quick test purposes
* W: workstation size (a 90's workstation; now likely too small)
* A, B, C: standard test problems; ~4X size increase going from one class to the next
* D, E, F: large test problems; ~16X size increase from each of the previous classes

Optei pela *EP*, *FT*, *IS* e *SP* como foco para os testes com classes A e B em cada.

# Caracterização do Sistema

Cada nó tem um disco associado, na tabela seguinte faço a sua associação. Para obter os Discos de cada nó recorri ao programa tentakel, para dispersão, e udevadm, para listagem, com a seguinte parametrização:

tentakel -g compute\_linux "/sbin/udevadm info -a -p /sys/class/block/sda/sda5 -q env | grep MODEL"

Para referências futuras apenas será mencionado o nome do Disco como identificação.

|  |  |
| --- | --- |
| **Nó do Cluster** | **Disco** |
| 431-3 | MB0500EBNCR |
| 431-5 | SAMSUNG\_HD502HI |
| 431-6 | SAMSUNG\_HD502HJ |
| 432-1 | MM0500EBKAE |
| 541-1 | WDC\_WD10EZRX-00A8LB0 |
| 641-8 | INTEL\_SSDSC2BW240A3F |
| 641-19 | INTEL\_SSDSC2BW240A4 |
| 652-1 | WDC\_WD20NPVT-00Z2TT0 |
| 662-6 | INTEL\_SSDSC2BW120A4 |

# Parâmetros de teste

Para este trabalho utilizei o iozone versão 3.397, escolhi output para ficheiro binário *excel* e nome específico do ficheiro temporário para não coincidir com outros testes concorrentes. Cada ficheiro output e temporário tem o nome do nó associado, neste caso o *431-3*.

/opt/iozone/bin/iozone -Ra -b 431\_3.xls -f 431\_3.tmp

# Testes utilizados

Realizei 13 testes para 512 MB de tamanho máximo de ficheiro. Sendo os testes os seguintes:

1. Write
2. Rewrite
3. Read
4. Reread
5. Random read
6. Random write
7. Backward read
8. Record rewrite
9. Stride read
10. Fwrite
11. Frewrite
12. Fread
13. Freread.

# Write

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Para 16 MB e 128 MB o **WDC\_WD20NPVT-00Z2TT0** foi o melhor, ficando em segundo lugar no teste de 512 MB quando perde para o **SAMSUNG\_HD502HJ**.

No geral o **WDC\_WD20NPVT-00Z2TT0** obteve os melhores resultados.

# Re-Write

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Para 16 MB o **WDC\_WD20NPVT-00Z2TT0** foi o melhor, depois para 128 MB e 512 MB há de novo pouca diferença entre o **WDC\_WD20NPVT-00Z2TT0** e o **SAMSUNG\_HD502HJ**.

No geral o **WDC\_WD20NPVT-00Z2TT0** obteve os melhores resultados.

# Reader

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Há medida que iam aumentando os valores do ficheiro de teste as oscilações foram aumentando.

Para 16 MB e 128 MB o **WDC\_WD20NPVT-00Z2TT0** foi o melhor e mais consistente. Para 512 MB apenas há um disco que se revelou constante, o **MM0500EBKAE**, que também no de 126 MB também se manteve.

No geral o **WDC\_WD20NPVT-00Z2TT0** obteve os melhores resultados.

# Re-Reader

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Nos 3 tamanhos o **WDC\_WD20NPVT-00Z2TT0** foi no geral o melhor.

# Random read

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Nos 3 tamanhos o **WDC\_WD20NPVT-00Z2TT0** foi no geral o melhor. Sendo que nos 128 MB o **INTEL\_SSDSC2BW240A4** esteve perto do topo.

# Random Write

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Nos 2 tamanhos iniciais o **WDC\_WD20NPVT-00Z2TT0** foi no geral o melhor mas para 512 MB o **SAMSUNG\_HD502HJ** leva uma ligeira vantagem.

# Random Write

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Para 16 MB o **INTEL\_SSDSC2BW240A3F** apresenta melhores resultados, mas o **WDC\_WD20NPVT-00Z2TT0** foi o que apresentou valores mais altos e constantes. Para 128MB o **WDC\_WD10EZRX-00A8LB0** conseguiu valores altos e constantes sem oscilações bruscas. Já para 512MB o **INTEL\_SSDSC2BW240A4** foi o que no total obteve maio *throughput*.

# Random Write

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

Por mais uma vez o **WDC\_WD20NPVT-00Z2TT0** foi o que apresentou melhores resultados nos 3 tamanhos, resultados esses bastante constantes.

# Stride Read

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

O disco **WDC\_WD10EZRX-00A8LB0** na totalidade obteve maiores resultados em 16MB e 512MB.

Para 128 MB o **WDC\_WD20NPVT-00Z2TT0** foi o melhor.

# Fwrite

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

O disco **WDC\_WD20NPVT-00Z2TT0** na totalidade obteve maiores resultados em 16MB.

Para os restantes tamanhos de teste o **SAMSUNG\_HD502HJ** aparece como pioneiro em resultados.

# Re-Fwrite

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

O disco **WDC\_WD20NPVT-00Z2TT0** na totalidade obteve melhores resultados em 16MB e 128 MB.

Já para 512MB o **SAMSUNG\_HD502HJ** foi ligeiramente melhor que o vencedor dos 2 tamanhos anteriores.

# Fread

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

O disco **WDC\_WD20NPVT-00Z2TT0** analiticamente foi superior em 16MB e 128 MB, o

**SAMSUNG\_HD502HJ** esteve com resultados muito próximos mas só foi efetivamente superior ao WDC para 512 MB.

# Re-Fread

Com os valores obtidos retirei 3 gráficos para 16 MB, 128 MB e 512 MB.

## *Conclusão*

O disco **WDC\_WD20NPVT-00Z2TT0** foi superior em 16MB e 128 MB, já o **WDC\_WD10EZRX-00A8LB0** apenas o superou nos 512 MB.

# Scripts, Diretorias e Gráficos

# Análise Final de Resultados e Conclusão

O trabalho necessário para realizar este projeto foi imenso, sendo talvez o que mais tempo precisou incluindo os de PCP do semestre passado. Tivemos que aplicar conhecimentos de Linux para mais rapidamente o realizarmos. Não consegui realizar tudo o que me propus, como realizar o teste em várias máquinas para fazer o comparar e dar sentido ao Benchmark e utilizar diferentes otimizações.

Os dados que deram origem a todo o trabalho estão divididos em 5 ficheiros *xlsx*: *grafico.xlsx* contém o resumo dos tempos obtidos, *A\_cpu.xlsx* com a utilização por processador e *A\_io.xlsx* com os dados de memória, disco e rede; e os respetivos para B, *B\_cpu.xlsx* e *B\_io.xlsx*. Devem ser consultados pois contêm dados individuais para cada teste. Apenas alguns deles foram colocados no Relatório.

Sempre utilizei o top para uma rápida análise da utilização dos processos no sistema mas com este projeto descobri novas ferramentas que serão muito úteis no futuro e que dão mais informações.

Com tudo concluído, é tempo de fazer as comparações das diferentes métricas e analisar os resultados. Nos 4 testes realizados o icc venceu três e o gcc 4.9.0 outro. Analisando ainda os dados obtidos o icc foi o compilador que no geral melhores resultados conseguiu.