DEFINIÇÃO DE AMBIENTE COMPUTACIONAL DE ALTO DESEMPENHO PARA MINERAÇÃO DE INFORMAÇÃO EM BANCO DE DADOS ASTRONÔMICOS

Murilo Moritz Parize[1]; Marcelo Massocco Cendron[2]

RESUMO

Com grandes avanços na observação espacial, a instrumentação astronômica tem gerado uma grande quantidade e variedade de dados. A heterogeneidade dos dados vem da particularidade construtiva de cada sistema de observação, que se tornam sua observação muitas vezes restrita a um determinado conjunto de características para todas as galáxias analisada. A união desses dados produz uma análise mais rica para o pesquisador, pois permite obter uma visão mais ampla das características do ponto que deseja estudar, tanto que essa ideia está sendo um importante mecanismo de troca de dados entre os astrônomos. Porém, há problemas na hora de integrar esses dados e principalmente no que se refere aos recursos computacionais. A grande quantidade de interação entre as diversas bases geram uma demanda de processamento e transações com os dispositivos de armazenamento consideravelmente elevadas. Por isso, necessita que sistemas computacionais sejam dimensionados e otimizados especificamente para esse tipo de atividade. De forma a definir uma infraestrutura que melhor se adeque as exigências computacionais da pesquisa, foram realizados testes para otimização de desempenho utilizando a tecnologia RAID, com o intuito de aumentar a taxa de vazão de dados do disco rígido e também foi testado o uso de memórias Flash (pendrive) na configuração de RAID com o intuito de mensurar a eficiência desse tipo de dispositivo devido a seu baixo tempo de acesso às informações.
Palavras-chave: Alto desempenho, Banco de dados, Mineração.

1 INTRODUÇÃO

Durante a última década, o salto tecnológico na instrumentação astronômica fez crescer de forma considerável o número de projetos dedicados ao mapeamento de galáxias no Universo. Esta nova era na indústria de levantamento de galáxias *surveys* cresceu juntamente com a nossa necessidade de compreender as propriedades observadas de galáxias e definir uma teoria cosmológica consistente com as observações. A avalanche de dados científicos obtidos por tais projetos deu origem a uma nova forma de pesquisa na área de Astrofísica Observacional, popularmente referida por mineração ou garimpagem de dados, fruto da necessidade da exploração e manipulação dos bancos de dados, além da criação de novos produtos científicos a partir dos dados brutos obtidos pelos *surveys.*

[1] Instituto Federal Catarinense de Educação Ciência e Tecnologia – Campus Videira - Aluno do Bacharelado em Ciência da Computação 2012-1, murilo.moritz33@gmail.com
[2] Instituto Federal Catarinense de Educação Ciência e Tecnologia – Campus Videira - Professor Orientador – marcelo.cendron@ifc-videira.edu.br

Com isso verificou-se a necessidade de desenvolvimento de um grande conjunto de ferramentas para redução, armazenamento, organização, integração, análise e exploração de dados. Esse conjunto de programas e protocolos denomina-se Observatório Virtual (VO).

Entretanto, devido a uma série de características, os registros mantidos nestes *surveys* não são necessariamente coincidentes. A diferença no espectro eletromagnético da visualização relacionada com cada *survey*, por exemplo, pode resultar em coleções de objetos distintos em cada um dos volumes de dados. Por conta disto, para que a pesquisa destes bancos de dados aconteça de forma eficiente, é de fundamental importância que aconteça a execução de uma etapa de relacionamento dos dados mantidos em *surveys* distintos. As ferramentas relacionadas a computação, fazendo uso das tecnologias de Banco de dados e Mineração de Dados oferece a possibilidade de aprofundar esta pesquisa.

Com o problema definido de como desenvolver a mineração de dados entre os bancos de dados astronômicos, percebe-se que uma elevada demanda computacional será necessária para a comparação e análise desses dados. Por exemplo, o projeto SDSS (WEINBERG e RADDICK, 2011) possui em sua base de dados mais de 930 mil galáxias e 120 mil quasares, enquanto o WISE (NASA, 2011), apresentava até o ano de 2011 dados de centenas de milhões de galáxias, estrelas e asteróides. Esses projetos servem como base, porém outros podem ser agregados.

Com a intersecção dessas bases há a necessidade de grande quantidade de processamento necessária para a realização da análise desses dados. Isso implica na utilização de computadores com alta capacidade de processamento, mas somente o processamento não resolve efetivamente o problema é necessário que o computador tenha também uma alta capacidade de transferência de dados das memórias.

Por isso o principal objetivo é definir um ambiente computacional de alto desempenho para mineração de informações em banco de dados astronômicos, delimitando técnicas de otimização.

A complexidade de certa tarefas que o computador realizará é um fator crucial para a estimativa de sua conclusão. Essa complexidade se traduz em tempo de espera, podendo chegar inclusive a impossibilidade de se executar um aplicativo em tempo hábil devido principalmente quantidade cálculos necessários.

Porém, o cálculo realizado pelo processador não é o único limitador do desempenho computacional e consequentemente alterações na parte física desse

componente bem como alterações na forma como se fazem os cálculos, nem sempre reduzirão significativamente a quantidade de tempo necessário para o processamento.

Para certos aplicativos como banco de dados, devido ao seu tamanho que não raramente ultrapassada as dezenas de gigabytes, há o agravante da limitação de desempenho dos dispositivos de armazenamento.

Segundo (STALLINGS, 2010), para as memórias, são utilizados três parâmetros para medir o desempenho: I) Tempo de acesso (latência), II) Tempo de ciclo de memória e III) Taxa de transferência. Os dois primeiros se referem ao tempo necessário para que os dados comecem a serem lidos ou gravados no dispositivo de armazenamento e o terceiro, se refere a capacidade de vazão que o dispositivo suporta.

Há ainda, fatores independente do hardware do armazenamento que interfere no desempenho do dispositivo de armazenamento, esses fatores, em nível de software, incluiriam (ATKIN, 2011) a forma como os dados são armazenados (sequencial ou aleatório), o tamanho do arquivos armazenados e inclusive posicionamento dos dados no disco.

Algumas técnicas podem ser implementadas para a melhoria do desempenho de banco de dados, várias alternativas são propostas:

- Memórias Flash: (PATTERSON, LEE, et al., 2008) Devido à baixa latência, consegue-se valores significativos de operações de acesso ao Banco de dados. Ainda possui grandes brechas de pesquisa nessa área (OU e HÄNDER, 2012).
- GPU: (CHILD, 2010) Melhor desempenho do banco de dados através do uso de aceleração do processamento pela placa de vídeo. A placa de vídeo possui quantidade de núcleo de processamento e vazão da memória, consideravelmente maiores do que o processador.
- RAID: (PATTERSON, LEE, et al., 2008) (redundant array of independent disks) uma solução de baixo custo e melhoria considerável do desempenho, utiliza discos rígidos convencionais para aumentar a quantidade de operações e a taxa de transferência do disco rígido.
- Sistemas multiprocessadores: permite o paralelismo de tarefas entre os núcleos do processador, porém o gargalo seriam nos dispositivos interligados que ainda apresentam restrições de desempenho e limitam o desempenho (barramento e sistemas de memória)

- Sistemas multicomputadores: através de virtualização, cluster (TAMURA, OGUCHI e KITSUREGAWA, 1997) e computação nuvem (IOSUP, OSTERMANN, et al., 2011), são boas alternativas para processar dados com alta granularidade, pois envolve o custo de transporte e alocação de tarefas que implicam em tempo perdido, se a quantidade de processamento não for significativamente elevada, esse tempo se torna penoso.

Através da análise dessas técnicas, percebemos a quantidade de possibilidades para a otimização de banco de dados. Porém nos testes das referências, percebe-se a grande variedade de ambientes de testes utilizados, por isso, há a necessidade de realizar aplicação dessas técnicas para o banco de dados.

2 PROCEDIMENTOS METODOLÓGICOS

Para o desenvolvimento do projeto foram utilizados quatro discos rígidos de 500 GB cada e quatro Pendrives como dispositivos de memórias flash para a realização de testes de desempenho.

O sistema operacional adotado para a realização dos testes foi o Debian 7 devido a sua estabilidade e segurança, onde foram criados os arranjos de RAID e as configurações dos testes. No que se refere ao uso e teste do banco de dados adotamos como sistema gerenciador de banco de dados (SGBD) o PostgreSQL por ser uma ferramenta livre e que atende as necessidades do projeto.

Inicialmente realizou-se a configuração dos arranjos RAID via software, onde utilizou-se a ferramenta *mdadm* para a execução dessa tarefa. Utilizamos a configuração de RAID 1+0 onde ocorre a primeiramente o espelhamento dos dados e depois a segmentação entre os discos do arranjo.

Para a criação do arranjo de configuração RAID 1 utilizando 2 HD's:

mdadm –C /dev/md0 –l 1 –n 2 /dev/sdb /dev/sdc

Onde o argumento *–C* representa a criação do arranjo */dev/md0*, *-l 1* representa o nível de configuração do arranjo que neste caso foi 1 o parâmetro *–n 2* refere-se a quantidade de discos que irão compor o arranjo e depois passamos os nomes dos discos */dev/sdb* e */dev/sdc*. Esse processo foi realizado duas vezes pois para a configuração de

RAID 1+0 é necessário dois arranjos em configuração de RAID 1, alterando apenas os discos conforme abaixo:

mdadm –C /dev/md1 –l 1 –n 2 /dev/sdd /dev/sde

Após a criação dos arranjos de RAID 1 é necessário a configuração do arranjo de RAID 0 onde será feito a divisão dos dados entre os dois arranjos de RAID 1, para a configuração do arranjo foi utilizado o seguinte comando:

mdadm –C /dev/md2 –l 0 –n 2/dev/md0 /dev/md1

O processo de criação dos arranjos foi realizado utilizando os HD`s e as pendrives configurados como em RAID 10, cada um em arranjo separados. Esses arranjos foram utilizados para a execução dos testes de desempenho.

Após a configuração da máquina efetuou-se os testes práticos de desempenho dos discos e memórias flash utilizamos dois Benchmarks diferentes, ***Bonnie*** **++** e o ***pgbench*** o primeiro testa exclusivamente as operações de escrita e leitura de dados aleatórios em um HD ou arranjo, e o segundo realiza testes diretamente no banco de dados onde é possível verificar o desempenho em diferentes cenários. O pgbench utiliza o teste TPC-B no qual mede o rendimento do banco de dados analisando a quantia transações realizada com o banco de dados por segundo.

Para o uso do *Bonnie++* foi necessário baixar o pacote do programa. Quando instalado o programa deve ser executado no local onde os arranjos estão montados, geralmente em sistemas operacionais Linux eles estão no diretório /media. Abaixo temos o demonstrativo do teste executando:

```
root@debian:/media# bonnie++ -d /media/md2/ -s 8g -m
galaxia -f -b -u root
Using uid:0, gid:0.
Writing intelligently...done
Rewriting...done
Reading intelligently...done
start 'em...done...done...done...done...done...
Create files in sequential order...done.
Stat files in sequential order...done.
Delete files in sequential order...done.
Create files in random order...done.
Stat files in random order...done.
Delete files in random order...done.
```

Esse teste foi realizado somente em um HD simples sem estar configurado em RAID 10 e também em um arranjo RAID10. Não foi possível a realização desse teste nos

dispositivos de memória flash devido ao fato da capacidade de armazenamento dos dispositivos ser inferior ao tamanho da memória RAM.

Quando passou-se a testar o desempenho do Banco de Dados utilizando o *pgbench* foi necessária sua instalação da base de testes no SGBD. Após instalado, para sua execução é preciso estar no diretório onde o *pgbench* foi instalado. Abaixo temos um exemplo de execução do teste:

```
postgres@debian:/usr/lib/postgresql/9.1/bin$
./pgbench -c 4 -j 2 -T 60
starting vacuum...end.
transaction type: TPC-B (sort of)
scaling factor: 1
query mode: simple
number of clients: 4
number of threads: 2
duration: 60 s
number of transactions actually processed: 1193
tps = 19.875055 (including connections establishing)
tps = 19.877058 (excluding connections establishing)
```

Os parâmetros adotados no teste são respectivamente o número de clientes que acessam o banco simultaneamente, o número de *threads* por cliente e o tempo de teste. Como resultado teremos o número de transações atuais processadas e a quantidade de transações por segundo (tps).

3 RESULTADOS E DISCUSSÕES

Com todo o ambiente preparado, os testes foram executados e os resultados coletados. Com o uso do programa Bonnie++, realizamos a comparação de desempenhos do arranjo RAID e um HD simples.

Os indicadores medidos estão tabulados na Tabela 1:

Teste HD Raid 10													
	Sequential Output							***Sequential Input***				***Random Concurrency***	
		-Per Chr-		**--Block--**		**-Rewrite-**		**-Per Chr-**		**--Block--**		**Seeks**	
Machine	Size	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*
galaxia	8G	-	-	209937	30%	85675	22%	-	-	195165	28%	137.1	4%
Latencia	-	-		378ms		369 ms		-		65947us		406ms	

Teste HD Simples													
	Sequential Output							***Sequential Input***				***Random Concurrency***	
		-Per Chr-		**--Block--**		**-Rewrite-**		**-Per Chr-**		**--Block--**		**Seeks**	
Machine	Size	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*	*K/sec*	*%CP*
galaxia	8G	-	-	127635	14%	56682	11%	-	-	140689	13%	95.1	2%
Latencia	-	-		561ms		225ms		-		20143us		774ms	

Tabela 1: Resultados Teste Bonnie++

Pode-se analisar através da Tabela 1 indicadores de leitura e escrita de dados nos dispositivos de armazenamentos testados, a unidade de medida dos testes basicamente se refere a blocos de dados por segundo, a quantidade de uso do processador e a latência medida em milissegundos. Na sub-tabela *Sequential Output* os indicadores são de escrita e reescrita.

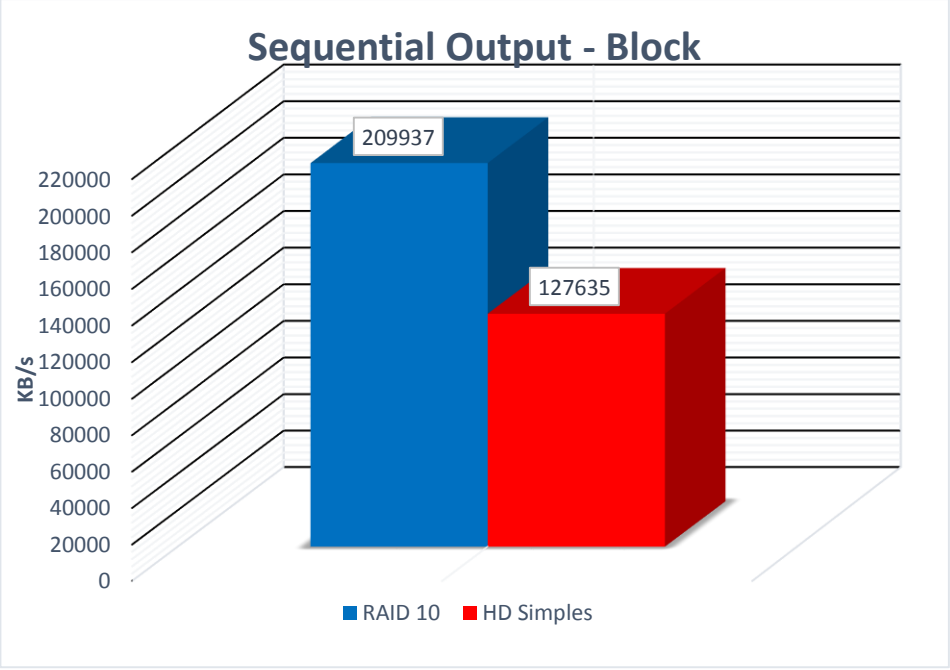

Gráfico 1: Escrita de dados por blocos

No Gráfico 1 percebe-se que a quantidade de dados escritos durante o tempo de teste é muito superior na configuração de RAID 10 do que em um HD simples. Quando falamos de reescrita de dados a configuração RAID ainda tem um elevado ganho sobre o uso de um HD simples. No quesito uso do processador o disco RAID exige uma utilização praticamente duas vezes maior que o disco normal. Pode-se observar no Gráfico 2.

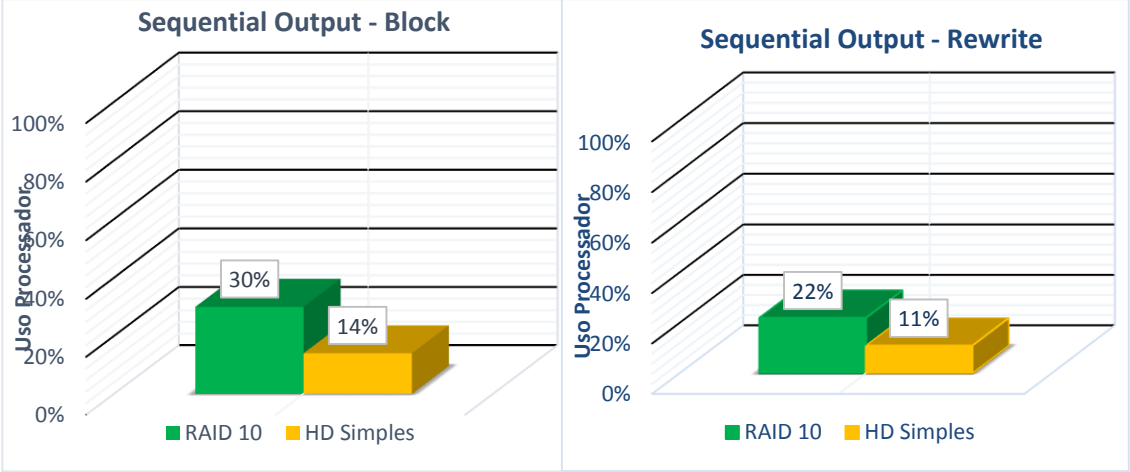

Gráfico 2: Uso Processador

Após os testes de desempenho de disco, procurou-se especificar os testes para o foco principal do projeto que é a consulta ao Banco de Dados. Para isso foi utilizado a ferramenta *pgbench.* Testou-se a atuação do banco sob diversos tipos de mídias e configurações de armazenamento, bem como a quantidade de clientes e consultas em diversos cenários.

	Cenário 1	Cenário 2	Cenário 3
Nº Clientes	4	4	8
Nº Threads	2	4	8

Tabela 2: Cenários

Na Tabela 2 observa-se os diferentes cenários escolhidos para testar o banco, abaixo na Tabela 3, tem-se o resultado dos testes, onde são analisados a quantidade de transações no banco, o número de transações por segundo incluindo conexões estáveis e o número de transações por segundo excluindo conexões estáveis.

	TESTE RAID 10 PENDRIVE		
	CENÁRIO 1	CENÁRIO 2	CENÁRIO 3
Nº Transações Processadas	1193	1548	1295
Transações por Segundo (Incluindo conexões estáveis)	19,875055	25,780689	21,514204
Transações por Segundo (Excluindo conexões estáveis)	19,877058	25,782369	21,518314
	TESTE HD RAID 10		
	CENÁRIO 1	CENÁRIO 2	CENÁRIO 3
Nº Transações Processadas	5922	6243	6040
Transações por Segundo (Incluindo conexões estáveis)	89,647735	103,98319	100,53478
Transações por Segundo (Excluindo conexões estáveis)	98,66715	103,99209	100,55094
	TESTE PENDRIVE SIMPLES		
	CENÁRIO 1	CENÁRIO 2	CENÁRIO 3
Nº Transações Processadas	8484	12564	11483
Transações por Segundo (Incluindo conexões estáveis)	141,37073	209,35936	190,21125
Transações por Segundo (Excluindo conexões estáveis)	141,38451	209,37141	190,22681
	TESTE HD SIMPLES		
	CENÁRIO 1	CENÁRIO 2	CENÁRIO 3
Nº Transações Processadas	5662	5700	5634
Transações por Segundo (Incluindo conexões estáveis)	94,316498	94,94726	93,806053
Transações por Segundo (Excluindo conexões estáveis)	94,331836	94,955544	93,82035

Tabela 3: Resultados *pgbench*

Pode-se verificar que o cenário onde obteve-se o maior número de transações processadas independentemente do tipo do dispositivo foi o cenário 2 onde 4 clientes acessam simultaneamente o banco, cada um executando 4 *threads.*

4 CONCLUSÃO

Este trabalho propõe uma abordagem para a definição de um ambiente de alto desempenho para banco de dados. Os métodos propostos utilizam de diversos recursos computacionais que propiciam essa otimização. Como são vários os métodos dedicou-se atenção especial as consultas e tempo de acesso ao disco.

Analisando os testes realizados confirmou-se o que a teoria já nos fala que a configuração RAID 10 aplicada a banco de dados é eficiente por proporcionar a leitura e escrita de dados de maneira mais ágil. De certa forma um dos resultados foi extremamente surpreendente, pois quando utilizou-se como dispositivo de armazenamento somente uma pen-drive o número de transações foi superior aos outros dispositivos, contudo sabe-se que esse dispositivo possui uma baixa estabilidade de conexão e baixa segurança de dados.

Em trabalhos futuros propõe-se definir outras técnicas de otimização de hardware para banco de dados. Uma tendência é o estudo de técnicas de processamento de dados através de GPU.

REFERÊNCIAS

ATKIN, I. **Getting the hang of IOPS**. Symantec | Connect, 25 Abril 2011. Disponível em: <http://www.symantec.com/connect/articles/getting-hang-iops>. Acesso em: 28 Maio 2012.

CHEN, P. M. et al. **RAID**: High-Performance, Reliable Secondary Storage. ACM Computing Surveys, New York, NY, v. 26, n. 2, p. 145 - 185 , 2 June 1994.

CHILD, T. **Introducing PgOpenCL A New PostgreSQL Procedural Language Unlocking the Power of the GPU**. PostgreSQL Wiki, 16 Dezembro 2010. Disponível em: <http://wiki.postgresql.org/images/6/65/Pgopencl.pdf>. Acesso em: 29 Maio 2012.

GAMA. **Galaxy And Mass Assembly**. Project overview, 2012. Disponível em: <http://www.gama-survey.org/>. Acesso em: 29 Maio 2012.

IOSUP, A. et al. **Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing.** IEEE Transactions on Parallel and Distributed Systems, Piscataway, NJ, USA , 6 June 2011. 931-945.

NASA. **WISE Delivers Millions of Galaxies, Stars, Asteroids**. WISE - Wide-field Infrared Survey Explorer, 2011. Disponível em: <http://www.jpl.nasa.gov/news/news.cfm?release=2011-117>. Acesso em: 35 Maio 2012.

OU, Y.; HÄNDER, T. **Improving Database Performance Using a Flash-Based Write Cache**, v. 7240/2012, p. 2-13, 2012. ISSN DOI: 10.1007/978-3-642-29023-7_2.

PATTERSON, D. A. et al. **A Case for Flash Memory SSD in Enterprise Database Applications**. 2008 ACM SIGMOD international conference on Management of data, New York, NY, p. 1075-1086, 2008.

STALLINGS, W. **Arquitetura e Organização de computadores**. São Paulo: Pearson Pratice Hall, v. 8, 2010.

TAMURA, T.; OGUCHI, M.; KITSUREGAWA, M. **Parallel Database Processing on a 100 Node PC Cluster: Cases for Decision Support Query Processing and Data Mining**. Supercomputing '97 Proceedings of the 1997 ACM/IEEE conference on Supercomputing, New York, NY, 1997. 1 - 16.

WEINBERG, ; RADDICK,. **The Sloan Digital Sky Survey**. Sloan Digital Sky Survey - Mapping the universe, 2011. Disponível em: <http://www.sdss.org/>. Acesso em: 31 Maio 2012.

PGBENCHTESTING. **Wiki Postgresql Pgbenchtesting.** Disponível em: wiki.postgresql.org/wiki/Pgbenchtesting. Acessado em 30 de julho de 2013