

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO SUDESTE DE MINAS GERAIS - CAMPUS RIO POMBA

Miguel Magalhães Lopes

Benchmark de desempenho entre bancos de dados em diferentes arquiteturas

Rio Pomba

20XX

Miguel Magalhães Lopes

Benchmark de desempenho entre bancos de dados em diferentes arquiteturas

Trabalho de Conclusão apresentado ao Campus Rio Pomba, do Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, como parte das exigências do curso de Bacharelado em Ciência da Computação para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Gustavo Henrique da Rocha Reis

Coorientador: CICLANO

Rio Pomba

20XX

Miguel Magalhães Lopes

Benchmark de desempenho entre bancos de dados em diferentes arquiteturas/

Miguel Magalhães Lopes. – Rio Pomba, 20XX-

15 p. : il. (algumas color.) ; 30 cm.

Orientador: Gustavo Henrique da Rocha Reis

Trabalho de Conclusão de Curso – Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas, Campus Rio Pomba, 20XX.

1. 2. I. II. III. IV.

Miguel Magalhães Lopes

Benchmark de desempenho entre bancos de dados em diferentes arquiteturas

Trabalho de Conclusão apresentado ao Campus Rio Pomba, do Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, como parte das exigências do curso de Bacharelado em Ciência da Computação para a obtenção do título de Bacharel em Ciência da Computação.

Trabalho aprovado. Rio Pomba, 00 de dezembro de 20XX.

Gustavo Henrique da Rocha Reis,
Orientador, IF Sudeste MG - Rio Pomba

CICLANO, Coorientador, IF Sudeste MG
- Rio Pomba

Dr. BELTRANO
IF Sudeste MG - Rio Pomba

Me. XXXXXXXXXXXXXXXX
IF Sudeste MG - Rio Pomba

Rio Pomba
20XX

Este trabalho é dedicado a todos
aqueles que me inspiraram, em especial
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

Agradecimientos

Resumo

Palavras-chaves: palavra1. palavra2. palavra3. palavra4.

Abstract

Key-words: word1. word2. word3. word4. word5.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

DACC	Departamento Acadêmico de Ciência da Computação
UFJF	Universidade Federal de Juiz de Fora
arm	ARM, originalmente Acorn RISC Machine, e depois Advanced RISC Machine, é uma família de arquiteturas RISC desenvolvida pela empresa britânica ARM Holdings
IDE	
x64	
x86	
aarch64	
ram	
GPU	
TPU	
CPU	
JVM	JVM (Java Virtual Machine) é uma máquina abstrata. É uma especificação que fornece um ambiente de tempo de execução no qual o bytecode do java pode ser executado. As JVMs estão disponíveis para muitas plataformas de hardware e software (ou seja, a JVM depende da plataforma).
IOT	
SBC	
BIOS	
TTL	

UART

WINE

Sumário

Introdução	2
1 Fundamentação Teórica	4
1.1 arquiteturas	4
1.1.1 Subseção de Exemplo	8
1.2 Seção de Exemplo 2	8
1.3 bancos de dados	8
1.4 software de benchmark	9
2 testes	10
3 Trabalhos Relacionados	11
4 Metodologia	12
4.1 Exemplo	12
5 Resultados	13
5.1 Resultados do Método	13
6 Conclusão	14
Referências	15

Introdução

Esta pesquisa foi baseada na crescente adoção de processadores `arm3`, estes processadores conseguem entregar uma eficiência energética muito superior a comumente utilizada nos computadores e servidores a arquitetura `x866` esta arquitetura possui uma versão 64 bit, que hoje em dia é praticamente a única variante utilizada hoje em dia a `x645` essa arquitetura é no geral apenas uma variante aditiva da `x866` na qual adiciona varias instruções e vários suportes adicionais, o principal deles sendo suporte a comandos 64bit,o mesmo pode ser dito a arquitetura `aarch647` ou `arm64` que é uma variante aditiva da `arm3`,essa não é ,como a `x645` uma versão única,mas sim uma "denominação"das variantes e evoluções da arquitetura `arm3` com suporte a 64bit,as arquiteturas passaram a ser denominadas assim a partir da `armv8`,entretanto existem versões do `arm3`,como o `armv7l` que consegue trabalhar com instruções de 64bit,apesar de ser uma arquitetura 32bit.

O foco da pesquisa foi feito em cima do uso de servidores, que é baixo, apesar de totalmente possível e existente no mundo corporativo,existem alguns servidores comerciais que utilizam a arquitetura `arm3` para trabalhar,foi feita uma comparação em utilização desses processadores para uma aplicação de banco de dados simulada,a aplicação simula a utilização de forma realística de uma base de dados de uma locadora.

O cenário foi escolhido a partir de um esquema de banco de dados genérico da internet e foram utilizados os bancos de dados `postgresql` e `mariadb`,visto que são os bancos de dados de propósito geral mais utilizados atualmente,foi preferido o `mariadb` sobre o `mysql` visto que não existe uma versão dele para a arquitetura `arm3` e ambos são basicamente o mesmo sistema,com a diferença de um deles ser open source.

foi criado um programa para a geração de dados realísticos baseados na biblioteca `faker` na linguagem `python`,estes dados são gerados para cada país específico em idiomas e caracteres compatíveis para o país designado,desta forma é plausível que estes dados,como nome,telefone,endereço e até mesmo usuário e senha sejam possíveis de existirem,foi escolhido dessa forma já que um preenchimento de dados totalmente randômicos e de tamanho fixo podem apresentar discrepâncias com o desempenho num ambiente real de uso,isso pode afetar tanto o tempo,quanto carga dos processadores negativamente, e visto que

os dados utilizados em cada teste são exatamente os mesmos, com a diferença apenas da quantidade, para a análise de desempenho variante em quantidade, quer dizer que esses dados poderiam sim simbolizar o uso num ambiente real e sendo assim o benchmark se torna mais aplicável a realidade.

1 Fundamentação Teórica

1.1 arquiteturas

Por definição arquitetura de computador é um conjunto de circuitos eletrônicos padronizado associado a um conjunto de instruções de forma a simplificar a programação deles para que funcionem comandos diferentes do binário para a programação de um eletrônico, os compiladores utilizam esses conjuntos de instrução para que seja convertido o código de uma linguagem de programação para um binário de um programa, a arquitetura também define/limita varias propriedades do hardware, como quantidade máxima de ram⁸, de disco, suporte ou não de saída de video, capacidades de rede e vários outros, mesmo que algumas dessas limitações possam ser contornadas utilizando variações da arquitetura chamadas microarquiteturas.

Uma micro arquitetura é quando é adicionado tanto um circuito eletrônico novo ao circuito original da arquitetura quanto apenas uma simplificação ou reorganização dos comandos originais de uma arquitetura, entretanto, numa micro arquitetura essas modificações são muito pequenas de forma a serem mais similares a arquitetura original do que uma nova arquitetura, dessa forma as micro arquiteturas podem ser consideradas updates de uma arquitetura, e quando são acumulados muitos desses updates, pode ser que seja gerada uma nova arquitetura, como foi o caso da arquitetura x86⁶ para a arquitetura x64⁵, onde a x64⁵ foi um upgrade grande o bastante da x86⁶ a ponto de ser considerado uma nova arquitetura, onde a principal e mais visível diferença entre esses dois é a mudança de 32bit (na x86⁶) para 64bit (na x64⁵).

As arquiteturas também podem ser definidas para coisas fora de CPU¹¹, elas podem estar definidas em GPU⁹, TPU¹⁰ e vários outros módulos de hardware de um computador, inclusive existindo arquiteturas especiais que são aplicadas a nível de software, que não são necessariamente arquiteturas de computador, mas sim um tipo diferente de arquitetura, onde existem maquinas abstratas que simplificam a programação de uma linguagem para que ela funcione de forma mais compatível com varias maquinas de ar-

arquitecturas de hardware diferentes, onde você faz otimizações de código na parte do código e da máquina virtual, como o caso da JVM¹² do Java, em que a máquina virtual de cada arquitectura pode sofrer otimizações e isso faz com que ela funcione de forma melhor dependendo da máquina virtual em uma arquitectura e pior em outra, mas sem alterar o código, e ao mesmo tempo, outra máquina virtual pode funcionar pior na primeira arquitectura e melhor na segunda.

As arquitecturas não são limitadas apenas a esses previamente citados, as arquitecturas podem estar presentes todos os tipos de circuitos integrados, como os processadores de roteadores e aparelhos IOT¹³ como lâmpadas e tomadas inteligentes. Isso quer dizer que uma arquitectura não necessariamente é algo que precise de um hardware potente ou que só funcione ou exista em computadores, as arquitecturas são a forma como os algoritmos são interpretados no hardware, o que quer dizer que dê de que exista um software e um hardware que se comuniquem existe uma arquitectura e provavelmente houve uma conversão da linguagem de programação para a linguagem de máquina dessa arquitectura deste dispositivo

as arquitecturas de computador são definidas para hardwares específicos, mas os softwares não necessariamente precisam de ser funcionais apenas em uma única arquitectura, por mais que ela seja diferente da arquitectura de outro computador, isso por que os conjuntos de instruções podem ser diferentes mas suas funcionalidades gerais podem ser iguais, de forma que mesmo que uma arquitectura seja totalmente diferente de outra, os softwares de uma podem funcionar na outra, por mais que sejam necessárias algumas adaptações, algumas dessas adaptações podem ser tão grandes que as vezes é muito complexo essa adaptação de código, para esses casos, ou mesmo para se testar o código de uma arquitectura em outra, para se executar esses códigos sem ser necessária essa adaptação são usados programas chamados de emuladores ou simuladores estes programas funcionam como uma camada de compatibilidade entre a arquitectura real da máquina que está rodando e a arquitectura na qual o programa foi pensado em funcionar, entretanto esse processo pode acarretar em uma perda considerável de desempenho, podendo resultar em casos onde máquinas com 516 gigaflops sejam necessárias para se emular máquinas com 230 gigaflops, como no caso de um emulador do sistema de videogame Playstation 3

utilizando o emulador `rpcs3`, e mesmo com essa ineficiência, esse emulador não tem 100% de compatibilidade com os softwares existentes na plataforma, de forma que nem todos os softwares dessa plataforma funcionam exatamente como deveriam, ou mesmo funcionam por mais que ambas as máquinas rodem o mesmo kernel de sistema, o kernel Linux no caso, ainda assim a perda de desempenho é muito grande pois apesar de em teoria serem o mesmo sistema operacional a diferença de arquiteturas possui um peso muito maior do que o que o sistema operacional utilizado, por mais que pareça que não é o caso.

Esse é um exemplo de como mesmo com tudo para ser um cenário igual de utilização ou mesmo um cenário melhor ao se trocar uma arquitetura de um computador inúmeras adaptações devem ser feitas ou, como no caso do macOS, criadas camadas de compatibilidade. Após o lançamento dos macbooks de 2020 com processador M1 que funcionam com a arquitetura `aarch64`⁷ a apple lançou uma camada de compatibilidade dos softwares com arquitetura `x64`⁵ para `aarch64`⁷ chamado de `rosetta2` esse software funciona parcialmente como um emulador, exceto que ele faz as adaptações num nível mais próximo do da máquina real e do sistema operacional nativo da máquina, resultando num desempenho muito superior a qualquer emulador existente, o `rosetta2` funciona de forma análoga ao projeto `WINE`¹⁸ do Linux que reinterpreta os programas windows para funcionarem no Linux, você tem uma pequena perda de desempenho por esse processo de reinterpretação em alguns casos, mas em outros essa perda é bem mais visível

As arquiteturas de computador podem variar em diversos fatores de uma para outra de forma que existam várias funcionalidades que não foram pensadas para uma arquitetura que existem em outras. Existem também propósitos diferentes para diferentes arquiteturas, como o caso dos processadores `arm`³ que foram pensados para entregar uma grande eficiência energética, enquanto os processadores `x86`⁶ foram pensados para apresentarem grande poder de processamento

Existem alguns computadores com processador `arm`³ que não são `SBC`¹⁴ isso faz com que eles possuam várias possibilidades de upgrade, que não são possíveis nos computadores `SBC`¹⁴, sendo assim existem algumas pequenas variações no funcionamento dos computadores mesmo dentro de uma mesma arquitetura que tenderia a seguir padrões

mais uniformes,entretanto uma peculiaridade tende a ser comum nos processadores arm³,eles costumam apresentar uma GPU⁹ integrada e algumas outras unidades de processamento especializadas que os processadores x86⁶ costumam ter que ser adicionadas com chips externos,uma dessas unidades é o TPU¹⁰ que ficou mais conhecida com o lançamento do windows 11 que o exigia na sua instalação por propósitos de segurança,o principal propósito da arquitetura arm³ entretanto não é se diferenciar tanto da arquitetura x86⁶,mas sim tornar computadores mais energeticamente eficientes,tanto que um computador doméstico comum utiliza de 200 a 300w por hora enquanto um computador raspberry pi 4,que é o computador arm³ mais potente atualmente da marca mais popular,consome coisa de 15w hora,que é uma diferença absurda,principalmente se levar em conta que ambos tem a capacidade de utilizar os mesmos programas de trabalho,se considerarmos sistema operacional Linux e programas open source,tanto editores de texto quanto navegadores de internet quanto IDE⁴s de programação estão disponíveis para ambos e para um uso comum funcionam tão bem quanto num cenário real.

Como os processadores arm³ começaram a ficar mais comuns visto que algumas fabricantes como a apple agora fabricam computadores baseados em arm³ e a gigabyte agora possui versões de servidor com essa arquitetura de processador,isso faz com que seja cada vez mais fácil de se utilizar essa arquitetura já que se existem mais consumidores também vão existir mais programas feitos para rodar nessa arquitetura,e visto o quanto um computador com processador arm³ economiza energia para entregar o mesmo poder de processamento de um outro com processador x86⁶, essa diferença pode ser muito benéfica para os vários tipos de empresa que utilizam servidores,já que isso pode significar um impacto considerável no consumo energético da empresa dependendo do quanto ele é utilizado a nível de processamento.

1.1.1 Subseção de Exemplo

1.2 Seção de Exemplo 2

1.3 bancos de dados

banco de dados é um método de armazenamento de dados de forma estruturada para que sejam mais fáceis de serem associados e de serem filtradas, elas também podem ser armazenadas de forma a economizar espaço de armazenamento dependendo da forma como foi otimizado o banco de dados, os bancos de dados ainda possuem a capacidade de realizar algumas redundâncias de segurança para o armazenamento de dados, de forma que a validação de um dado inserido possa ser feita a nível de armazenamento de dados e não a nível de programação, o que pode simplificar o desenvolvimento de um programa. Esses motivos são os principais de por que foram escolhidos os bancos de dados como alvo do benchmark realizado para essa comparação de arquiteturas.

os bancos de dados são onde a maioria dos dados de um sistema são salvos, esses dados são as informações necessárias para o funcionamento dos mais diversos tipos de programas tanto para ambientes comerciais quanto entreterimento ou mesmo comunicação, todos esses ramos de software utilizam bancos de dados de algum tipo para salvar os dados e acelerar o acesso dos mesmos, isso quer dizer que os bancos de dados são extremamente importantes para qualquer tipo de aplicação. Isso por si só já é motivo o bastante para se utilizar esse tipo de programa como base para os testes de desempenho entre as arquiteturas do ponto de vista de um servidor, enquanto as mais diversas aplicações podem rodar de um lado do servidor uma coisa é constante, as aplicações que rodam seguindo o modelo cliente-servidor utilizam algum tipo de banco de dados.

os tipos de bancos de dados analisados são os bancos de dados sql que são os mais genéricos, de forma que podem ser utilizados no máximo de aplicações diferentes possíveis, isso faz com que os bancos de dados sql sejam os melhores para serem simulados, um dos que foi analisado para ser testado foi o mongodb e o oracle, mas o mongodb não possui propósito geral e o da oracle não existe uma versão para arm³ até o momento que o projeto foi pensado.

1.4 software de benchmark

o benchmark foi feito com 2 softwares principais um software de terceiros chamado dbbench que é um programa dedicado para o benchmark de bancos de dados e o segundo foi criado para inserir de forma padronizada os dados para o dbbench realizar os testes. o dbbench é um software para benchmark e teste de estresse de bancos de dados variados que utilizam arquivos de configuração para fazer variados tipos de testes. o software desenvolvido gera esses arquivos de configuração, sendo que dentro dos arquivos é possível inserir as queries que serão testadas e o software criado faz isso, de forma a manter os mesmos dados pesquisados mas alterar a quantidade de consultas recorrentes ou de operações por segundo, o software criado gera a quantidade desejada de operações

2 testes

os testes foram realizados utilizando um orange pi pc + e um computador de mesa . o orange pi é um SBC ARM baseado no processador allwinner h3 com 3 usb 2.0 , 1 gb de memória ram ddr3 , uma porta de rede 10/100 e wifi bgn,essa é relativamente antiga e seu processador é um 4-core de 1.3ghz no clock máximo,é um processador relativamente bom mas não é bom o bastante para substituir um computador atual,mas consegue funcionar de forma satisfatória como servidor doméstico ,visto que seu processador é bom o bastante para operações simplificadas e poucos acessos,mas quando se tratam de muitos acessos ele pode não ser potente o bastante para aguentar. o computador de mesa é um dual core intel com 4 gb de memória ram ddr3, 4 usb 2.0 porta de rede gigabit e algumas portas sata2,entretanto essas portas não serão usadas já que o propósito é manter as 2 maquinas o mais próximas em relação a hardwaree possível outros metodos que serão usados para manter as maquinas mais similares serão limitar o clock de ambos para 1.2ghz,que é o clock mais estável para o orange pi ,a memória usada será limitada a 750 mb para o stack do docker e tanto o sistema operacional quanto os dados do docker serão salvos em cartões de memória microsd classe 10 com limitação de acesso de 10 MB/s de escrita e leitura , no orangepi será usado o leitor da própria placa para conectar o microsd e no pc será usado um adaptador usb. alem disso serão usadas versões do sistema debian,no orangepi o armbian e no pc o próprio debian padrão,ambos na versão buster

3 Trabalhos Relacionados

4 Metodologia

4.1 Exemplo

(Equação 4.1).

$$c = \frac{1}{n_f} \tag{4.1}$$

5 Resultados

5.1 Resultados do Método

6 Conclusão

Referências