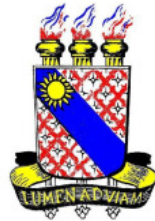


UNIVERSIDADE ESTADUAL DO CEARÁ



---

Trabalho Prático de Sistemas Distribuídos

---

Manoel Rui Pessoa de Paula  
João Gonçalves Filho

FORTALEZA  
2012

---

## Sumário

1	Introdução . . . . .	2
2	Materiais . . . . .	2
3	Configuração da Simulação . . . . .	3
4	Simulação e Resultados . . . . .	3
4.1	Cenário 1 . . . . .	3
4.2	Cenário 2 - Aplicação web hospedada em 3 data centers com 75 VM em cada. . . . .	9
4.3	Cenário 3 - Aplicação web hospedada em 3 data centers com 125,100 e 75 VM em cada. . . . .	10
4.4	Cenário 4 - Aplicação web hospedada em 3 data centers com 100 VM em cada. . . . .	11
4.5	Cenário 5 - Aplicação web hospedada em 3 data centers com 150, 125 e 100 VM em cada. . . . .	12
4.6	Sumário dos Resultados da Simulação . . . . .	13
5	Conclusão . . . . .	14
6	Bibliografia . . . . .	15

## 1 Introdução

Uma aplicação típica de grande escala na Internet que pode se beneficiar da Nuvem são aplicações de redes sociais. Esses aplicativos podem se beneficiar de nuvens porque normalmente apresentam padrões não uniformes de uso. O acesso a estes serviços varia ao longo do dia e localização geográfica das fontes de pedidos de serviço, também varia. Além disso, uma nova funcionalidade do serviço, pode causar um aumento repentino no interesse pelo serviço, conduzindo a um aumento do número de pedidos que chegam aos servidores que podem ser apenas temporários.

Infraestruturas em nuvem permite dinamicamente reagir ao aumento de pedidos, aumentando dinamicamente os recursos do aplicativo, e reduzindo os recursos disponíveis quando o número de pedidos reduz. Assim, SLAs entre provedores de nuvem e os consumidores estão satisfeitos com um custo mínimo para os consumidores.

Um bem conhecido site de redes sociais é o Facebook, que tem mais de 200 milhões de usuários registrados no mundo. Em 18/06/2009 a distribuição aproximada da base de usuários do Facebook em todo o mundo foi a seguinte: América do Norte: 80 milhões de usuários; América do Sul: 20 milhões de usuários; Europa: 60 milhões de usuários; Ásia: 27 milhões de usuários, África: 5 milhões de usuários e Oceania: 8 milhões de usuários (WICKREMASINGHE; CALHEIROS; BUYYA, 2010). Nosso trabalho busca tentar otimizar os cenários no artigo anteriormente proposto para o trabalho da disciplina, isso será feito através das alterações dos parâmetros dos cenários.

## 2 Materiais

Máquina de teste:

- Linux Ubuntu Release 12.10 (quantal) 32-bit
- Memória 2,9 GiB
- Processador Intel® Core™2 Duo CPU E7200 @ 2.53GHz x 2

Software Utilizados:

- CloudAnalyst é uma ferramenta que foi desenvolvida em uma plataforma acima do CloudSim na Universidade de Melbourne, com objetivo de auxiliar decisões de como organizar *datacenters* e máquina virtuais para melhorar o desempenho de aplicações de redes sociais.
- CloudSim é um framework para a modelagem e simulação de infraestruturas de computação em nuvem e Serviços.

### 3 Configuração da Simulação

Para serem simulados no ambiente clouddanalyzer, definimos 6 bases de usuários representando 6 regiões com os seguintes parâmetros:

User base	Region	Time Zone	Peak Hours (Local time)	Peak (GMT) Hours	Simultaneous Online Users During Peak Hrs	Simultaneous Online Users During Off-peak Hrs
UB1	0 - N. America	GMT - 6.00	7.00-9.00 pm	13:00-15:00	400,000	40,000
UB2	1 - S. America	GMT - 4.00	7.00-9.00 pm	15:00-17:00	100,000	10,000
UB3	2 - Europe	GMT + 1.00	7.00-9.00 pm	20:00-22:00	300,000	30,000
UB4	3 - Asia	GMT + 6.00	7.00-9.00 pm	01:00-03:00	150,000	15,000
UB5	4 - Africa	GMT + 2.00	7.00-9.00 pm	21:00-23:00	50,000	5,000
UB6	5 - Oceania	GMT + 10.00	7.00-9.00 pm	09:00-11:00	80,000	8,000

Fig. 1: Configuração das regiões

Em termos de custo, tomamos o seguinte plano:

Cost per VM per hour (1024Mb, 100MIPS)	\$0.10
Cost per 1Gb of data transfer (from/to Internet)	\$0.10

Fig. 2: Tabela de custos

Outros parâmetros de configuração da simulação podem ser visto na figura 3:

Assumimos que em cada datacenter configurado com os parâmetros descrito anteriormente o número de servidores é 20. As características da internet foram definidas conforme as figuras 4,5.

## 4 Simulação e Resultados

Os sete primeiros cenários foram feitos e descritos de forma semelhante em (WICKREMASINGHE et al., 2009). O cenários a seguir exibem sua configuração e os resultados obtidos na simulação.

### 4.1 Cenário 1

Nesse cenário temos dois datacenters e cada um com 50 máquinas virtuais, conforme a figura 6:

Parameter	Value Used
VM Image Size	10000
VM Memory	1024 Mb
VM Bandwidth	1000
Data Center - Architecture	X86
Data Center - OS	Linux
Data Center - VMM	Xen
Data Center - Number of Machines	20
Data Center - Memory per Machine	2048 Mb
Data Center - Storage per machine	100000 Mb
Data Center - Available BW per Machine	10000
Data Center - Number of processors per machine	4
Data Center - Processor speed	100 MIPS
Data Center - VM Policy	Time Shared
User Grouping Factor	1000
Request Grouping Factor	100
Executable Instruction Length	250

Fig. 3: Tabela de configuração das máquinas

Region/Region	0	1	2	3	4	5
0	25.0	100.0	150.0	250.0	250.0	100.0
1	100.0	25.0	250.0	500.0	350.0	200.0
2	150.0	250.0	25.0	150.0	150.0	200.0
3	250.0	500.0	150.0	25.0	500.0	500.0
4	250.0	350.0	150.0	500.0	25.0	500.0
5	100.0	200.0	200.0	500.0	500.0	25.0

Fig. 4: Tabela de latência em milisegundos

A política de agente de serviços escolhidos foi a do datacenter mais próximo, nesse caso o agente de serviços quando necessário irá encaminhar o tráfego para o datacenter mais próximo em termos de latência de transmissão.

Executando a simulação durante o período de um dia, temos os seguintes resultados:

<b>Region/Region</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>0</b>	2000.0	1000.0	1000.0	1000.0	1000.0	1000.0
<b>1</b>	1000.0	800.0	1000.0	1000.0	1000.0	1000.0
<b>2</b>	1000.0	1000.0	2500.0	1000.0	1000.0	1000.0
<b>3</b>	1000.0	1000.0	1000.0	1500.0	1000.0	1000.0
<b>4</b>	1000.0	1000.0	1000.0	1000.0	500.0	1000.0
<b>5</b>	1000.0	1000.0	1000.0	1000.0	1000.0	2000.0

Fig. 5: Tabela de largura de banda (Mbps)

Service Broker Poli... <b>Closest Data Center</b> ▼					
Data Center	# VMs	Image Size	Memory	BW	
DC1	50	10000	1024	1000	
DC2	50	10000	1024	1000	

Fig. 6: Dois datacenters com 50 VM em cada

Média global do tempo de resposta	7563,16 ms
Tempo de processamento do datacenter	7393,79 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC2	120,05	235,74	355,79
DC1	120,05	257,32	377,37

Fig. 7: Tabela de custos por datacenter (1)

Agora iremos buscar a melhor política de agente de serviços para esse cenário, agora iremos escolher a política otimiza o tempo de resposta, o agente monitora todos os datacenters, então ele busca estimar aquele que dará o melhor tempo de resposta um determinado usuário final, abaixo vemos que com o mesmo custo conseguimos um desempenho melhor tanto no tempo de resposta, como no tempo de processamento:

<b>Data Center</b>	<b>Avg (ms)</b>	<b>Min (ms)</b>	<b>Max (ms)</b>
DC1	8512,18	2,51	26306,49
DC2	6172,99	125,00	26104,60

Fig. 8: Tabela de serviço de tempo requisição (1)

Média global do tempo de reposta	6770,19 ms
Tempo de processamento do datacenter	6563,71 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

<b>Data Center</b>	<b>VM Cost \$</b>	<b>Data Transfer Cost \$</b>	<b>Total \$</b>
DC2	120,05	235,21	355,26
DC1	120,05	257,85	377,90

Fig. 9: Tabela de custos por datacenter (2)

<b>Data Center</b>	<b>Avg (ms)</b>	<b>Min (ms)</b>	<b>Max (ms)</b>
DC1	8081,73	2,51	38912,61
DC2	4899,57	2,51	17325,01

Fig. 10: Tabela de serviço de tempo requisição (2)

Nesses casos tínhamos um datacenter na região 0 que é a região que possui mais tráfego e um outro na região 1 que é a que possui o 2º maior tráfego, rodando a simulação com o mesmo custo financeiro, ou seja utilizando o número equivalente de máquinas virtuais e servidores, foi colocado um único datacenter na região 0, agora ele tem 100 máquinas virtuais e 40 servidores, conseguimos resultados bem melhores, como segue abaixo:

Média global do tempo de reposta	3968,35 ms
Tempo de processamento do datacenter	3710,15 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

Utilizando política da otimização do tempo de resposta

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC1	240,10	493,06	733,16

Fig. 11: Tabela de custos por datacenter (3)

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	3710,15	2,51	19687,91

Fig. 12: Tabela de serviço de tempo requisição (3)

Média global do tempo de reposta	3969,57 ms
Tempo de processamento do datacenter	3711,31 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

Utilizando política do datacenter mais próximo

Os resultados não mudam entre as política devido termos um ambiente centralizado. Mas pode não ser adequado ter apenas um único datacenter, devido se este cair derrubar todo o serviço, mas se dividirmos em dois datacenters na mesma região 0, temos um pequena perca de desempenho conforme os resultados abaixo:



Média global do tempo de reposta	4887,82 ms
Tempo de processamento do datacenter	4625,92 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

Utilizando política da otimização do tempo de resposta

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC2	120,05	258,31	378,36
DC1	120,05	234,75	354,80

Fig. 13: Tabela de custos por datacenter (4)

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	3828,81	2,51	15528,35
DC2	5350,33	2,51	32949,98

Fig. 14: Tabela de serviço de tempo requisição (4)

Média global do tempo de reposta	4004,09 ms
Tempo de processamento do datacenter	3746,74 ms
Custo das máquinas virtuais	240,10 \$
Custo de transferência de dados	493,06 \$
Custo Total	733,16 \$

Utilizando política do datacenter mais próximo

Em relação a política a ser adotada, destaca-se a política do mais próximo para esse cenário em particular já que mostrou ter um funcionamento melhor nesse ambiente, devido os datacenters estarem próximos. No entanto a política de otimização apresentou resultados bastante interessante. De um

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC2	120,05	246,77	366,82
DC1	120,05	246,28	366,33

Fig. 15: Tabela de custos por datacenter (5)

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	3756,96	2,51	19967,22
DC2	3736,54	2,51	19431,10

Fig. 16: Tabela de serviço de tempo requisição (5)

modo geral, a política de otimização do tempo de resposta é melhor para ambientes descentralizados o que corresponde mais a realidade. Lembrando também que o tipo de algoritmo implementado contribui para um melhor resultado.

## 4.2 Cenário 2 - Aplicação web hospedada em 3 data centers com 75 VM em cada.

Neste cenário as regiões aonde os data centers estão fixados são América do Norte( região 0 ) para o DC1, Europa(região 2) para o data center DC2 e por último a Ásia( região 3) para o DC3. Usamos as configurações para “Service Broker Policy” configurada com “Optimise Response” para conseguir um maior desempenho. Ajustamos também a configuração “Load Balancing Policy Across VM’s in a Single Data Centers” para “throttling”, visto que obteve-se melhores resultados com esse parâmetro nas simulações.

Média global do tempo de resposta	2314,05 ms
Tempo de processamento do datacenter	2149,91 ms
Custo das máquinas virtuais	540,23 \$
Custo de transferência de dados	493,06 \$
Custo Total	1033,28 \$

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC3	180,08	85,30	265,38
DC2	180,08	161,79	341,87
DC1	180,08	245,97	426,04

Fig. 17: Tabela de custos por datacenter para cenário 2

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	2911,60	2,51	15969,82
DC2	1625,11	2,51	11477,67
DC3	948,96	5,01	6597,42

Fig. 18: Tabela de serviço de tempo requisição para cenário 2

Percebemos que quanto mais VM's alocadas em cada data center consegue-se um tempo médio de resposta e tempo de processamento mais rápido em relação ao cenário que aloca 50 VM em cada data center. No entanto o custo total será bem maior visto que o número de VM's é maior.

### 4.3 Cenário 3 - Aplicação web hospedada em 3 data centers com 125,100 e 75 VM em cada.

Neste cenário apenas ajustamos o número de VM's para cada data center com 75 para DC1, 100 para o DC2 e 75 para o DC3.

Média global do tempo de reposta	2151,45 ms
Tempo de processamento do datacenter	1987,49 ms
Custo das máquinas virtuais	564,24 \$
Custo de transferência de dados	493,06 \$
Custo Total	1057,29 \$

Observe que o tempo médio de resposta foi melhor em relação ao cenário anterior. E o tempo de processamento de cada data center aparece mais próximo um do outro.

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC3	180,08	84,49	264,57
DC2	192,08	168,17	360,25
DC1	192,08	240,40	432,48

Fig. 19: Tabela de custos por datacenter para cenário 3

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	2587,71	2,51	14970,12
DC2	1676,22	2,51	11978,13
DC3	899,25	5,01	5679,61

Fig. 20: Tabela de serviço de tempo requisição para cenário 3

#### 4.4 Cenário 4 - Aplicação web hospedada em 3 data centers com 100 VM em cada.

Semelhante aos últimos cenários, as configurações permanecem as mesmas mas o número de VM em cada data center agora é 100.

Média global do tempo de resposta	2133,91 ms
Tempo de processamento do datacenter	1969,10 ms
Custo das máquinas virtuais	576,24 \$
Custo de transferência de dados	493,06 \$
Custo Total	1069,30 \$

Os resultados obtidos nesse cenário só consolida o fato de que quanto mais VM's alocadas em cada data center, consegue-se um tempo de resposta muito mais rápido. No entanto o custo total aumenta devido a quantidade dessas VM's.

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC3	192,08	84,39	276,47
DC2	192,08	172,44	364,53
DC1	192,08	236,23	428,31

Fig. 21: Tabela de custos por datacenter para cenário 4

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	2478,61	2,51	14970,12
DC2	1822,72	5,01	11978,13
DC3	841,92	7,51	6205,46

Fig. 22: Tabela de serviço de tempo requisição para cenário 4

#### 4.5 Cenário 5 - Aplicação web hospedada em 3 data centers com 150, 125 e 100 VM em cada.

Semelhante aos últimos cenários, as configurações permanecem as mesmas mas o número de VM em cada data center agora é 150 para o DC1, 125 para o DC2 e 100 para o DC3.

Média global do tempo de resposta	2158,73 ms
Tempo de processamento do datacenter	1998,34 ms
Custo das máquinas virtuais	576,24 \$
Custo de transferência de dados	493,06 \$
Custo Total	1069,30 \$

Observe que o aumento de 25 VM comparado aos dois último cenário e de forma distribuída justificado pela carga de trabalho distinta de cada região o tempo médio de resposta ainda não conseguiu ser melhor em relação ao cenário com 100 VM's para cada DC. Um fato curioso é que o custo total entre eles são semelhantes. Ou seja, pelos resultados obtidos adicionou-se 25 VM's gastando a mesma quantidade obtendo-se um tempo médio de resposta um pouco pior.

Data Center	VM Cost \$	Data Transfer Cost \$	Total \$
DC3	192,08	83,82	275,90
DC2	192,08	165,89	357,97
DC1	192,08	243,34	435,42

Fig. 23: Tabela de custos por datacenter para cenário 5

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	2636,73	2,51	14970,12
DC2	1625,84	2,51	10729,48
DC3	882,26	5,01	5927,90

Fig. 24: Tabela de serviço de tempo requisição para cenário 5

## 4.6 Sumário dos Resultados da Simulação

Os novos cenários propostos nesse trabalho podem ser visto de forma resumida na tabela abaixo:

	Média do tempo de reposta	Tempo de processamento
Dois datacenters com 50 VMs cada, datacenter mais próximo	7563,16 ms	7393,79 ms
Dois datacenters com 50 VMs cada, otimizando tempo de resposta	6770,19 ms	6563,71 ms
Um datacenter centralizado com 100 VMs, otimizando tempo de resposta	3968,35 ms	3710,15 ms
Um datacenter centralizado com 100 VMs, datacenter mais próximo	3969,57 ms	3711,31 ms
Dois datacenters centralizado com 50 VMs cada, otimizando tempo de resposta	4887,82 ms	4625,92 ms
Dois datacenters centralizado com 50 VMs cada, datacenter mais próximo	4004,09 ms	3746,74 ms

Cenário 1 e suas variações

	Média do tempo de reposta	Tempo de processamento
Cenário 2 3-75VMs	2314,05 ms	2149,91 ms
Cenário 3 3-75,100,125VMs	2151,45 ms	1987,49 ms
Cenário 4 3-100VMs	2133,91 ms	1969,10 ms
Cenário 5 3-100,125,150VMs	2158,73 ms	1998,34 ms

Demais Cenários

## 5 Conclusão

Com o objetivo principal de encontrar melhores configurações para uma simulação de grande escala da internet, como por exemplo o facebook, podemos perceber algumas características observadas. Com a proposta do primeiro

cenário com o intuito de analisar a relação das políticas adotadas pelos datacenters, vimos que o tempo de resposta das solicitações dos usuários são mais rápidas quando esses datacenters estão localizadas mais próximo dessa demanda de requisições. Mas um fato importante a ser considerado é a política da otimização do tempo de resposta que de modo geral obteve resultados bem similares com a política de execução da solicitação do usuário pelo datacenter mais próximo.

Reforçando o que já foi dito, pelos resultados do primeiro cenário percebe-se claramente que em ambientes centralizados, ou seja em uma única região, utilizando a política da otimização do tempo de resposta consegue-se um leve ganho de desempenho com praticamente o mesmo custo se utilizarmos a política do datacenter mais próximo. Mesmo em ambientes não centralizados os desempenhos são bem parecidos.

Através dos cenários restantes e justificado pelo cenário inicial a política de otimização de tempo de resposta foi adotada como base para os cenários restantes. É importante destacar que o algoritmo implementado em uma ambiente real para a otimização do tempo de resposta tem um grande impacto no desempenho do sistema. Deste modo, observamos também que o balanço da capacidade do sistema em relação da demanda de cada região pode ser uma estratégia bem interessante em se tratando de tempo de resposta e custo de alocação das máquinas virtuais.

Se revermos os últimos cenários nota-se bem essa relação da quantidade de máquina virtual que deve ser fornecida para se obter um desempenho e custo aceitável. Essa relação é bem relativo visto que em algumas aplicações é mais interessante alocar as máquinas virtuais em cada datacenter de forma mais equilibrada. A quantidade em si que deve ser disposta depende muito da demanda da região em que o datacenter está atendendo e que os seus donos estão dispostos a pagar. De maneira semelhante, muitas vezes é mais interessante alocar uma quantidade bem maior de máquina para uma determinada região justificada pelo fato da demanda ser também a maior. Essa relação deve ser estudada com bastante cuidado pois a disposição excessiva de máquinas virtuais em um determinado datacenter poderá proporcionar em um custo relativamente alto e a capacidade total do sistema ficará em uma boa parte do tempo ociosa.

## 6 Bibliografia

- WICKREMASINGHE, B.; CALHEIROS, R.N.; BUYYA, R. Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications. In: IEEE. Advanced Information



---

Networking and Applications (AINA), 2010 24th IEEE International Conference on. [S.l.], 2010. p. 446–452.

- WICKREMASINGHE, B. et al. Cloudanalyst: A cloudsim-based tool for modelling and analysis of large scale cloud computing environments. MEDC Project Report, 2009