



UNIVERSIDADE ESTADUAL DO CEARÁ
CENTRO DE CIÊNCIAS E TECNOLOGIA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Relatório do Trabalho Prático de Sistemas Distribuídos

FRANCISCO DUARTE OLIVEIRA

FORTALEZA
2012

Sumário

1	Introdução	2
2	Configuração das Simulações	3
3	Cenários e Resultados	5
3.1	Cenário 1	6
3.2	Cenário 2	7
3.3	Cenário 3	8
4	Conclusão	10

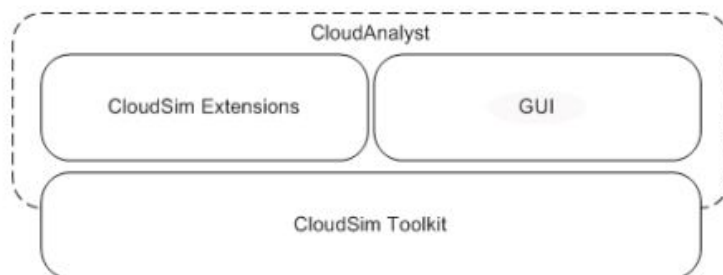
1 Introdução

Uma das principais preocupações dos desenvolvedores de aplicativos web é quanto a implantação e hospedagem de aplicações, isso pois normalmente exige a aquisição de um servidor com capacidade fixa capaz de lidar com a grande quantidade de requisições advindas do horário de pico da aplicação em questão. Ademais, também restão as preocupações quanto a manutenção da infra-estrutura, a plataforma de software de apoio a aplicação e obviamente aos prejuízos advindos da subutilização do servidor fora do horário de pico. Contudo, com os avanços na computação em nuvem surgem várias novas possibilidades para esses desenvolvedores, pois a implantação e hospedagem torna-se mais barata e mais fácil com o uso da política pay-per-use, onde o contratante paga apenas pelo que foi utilizado. Outro benefício que a nuvem fornece é possibilidade de uso de uma plataforma de infra-estrutura elástica, o que elimina a preocupação com tradeoff desempenho requerido nos horários de pico versos subutilização do servidor em horários normais.

Um dos tipos de aplicações web que podem se beneficiar com a utilização da nuvem são as chamadas redes sociais. Dessas a que mais vem se destacando nos últimos anos é o Facebook que ultrapassou em outubro desse ano, 2012, a marca de 1 bilhão de usuários ativos. Como os usuários desse tipo de aplicativo web estão distribuídos pelo globo tem-se então o problema de alocar os serviços da nuvem da melhor maneira para a aplicação. Assim, o trabalho prático aqui exposto tem por objetivo descobrir uma melhor configuração da nuvem para aplicações de larga escala, tais como o Facebook. Para tal, será utilizado o modelo de aplicação e os cenários descritos em [1], conforme pode ser visto na seção seguinte.

2 Configuração das Simulações

As simulações foram feitas utilizando o CloudSim através do CloudAnalyst que é uma ferramenta desenvolvida acima do CloudSim e que visa facilitar a utilização deste e a configuração das simulações. A figura seguinte ilustra essa relação do CloudAnalyst com o CloudSim:



Já a figura seguinte mostra as configurações da aplicação:

User base	Region	Time Zone	Peak Hours (Local time)	Peak Hours (GMT)	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 - Ocenia	GMT + 10.00	7.00-9.00 pm	09:00-11:00	80,000	8,000

Quanto as configurações de custo, foram realizados os seguintes opções:

Cost per VM \$/Hr	Memory Cost \$/s	Storage Cost \$/s	Data Transfer Cost \$/Gb
0,1	0,05	0,1	0,1

Ademais, a quantidade de máquinas virtuais por data center, a quantidade dos data centers e a localização desses varia conforme cada cenário com isso o restante das configurações comuns às simulações foram:

- Tamanho da VM: 100 MB
- Memória RAM por VM: 1 GB
- Largura de Banda por VM: 10 MB
- Arquitetura: x86
- Monitor de Máquina Virtual: Xen
- Sistema operacional: Linux
- Quantidade de Máquinas Físicas por DC: 40
- Memória RAM por Máquinas Física: 2 GB
- Processadores por Máquina Física: 4
- Velocidade por Processador: 10000 MIPS

Também foi usada a política de tempo compartilhado para o acesso das máquinas virtuais aos recursos físicos. Os usuários foram agrupados pelo fator de 1000, e as requisições foram agrupadas pelo fator de 100. Além disso, cada requisição requer que 250 instruções sejam executadas e foi escolhido a quantidade de 100KB de dados por requisição, visto que o tamanho médio de uma pagina web é esse.

Por fim, foram utilizadas as características padrões do simulador quanto as configurações da internet, essas são mostradas nas duas figuras abaixo:

Delay Matrix

The transmission delay between regions. Units in milliseconds

Region\Region	0	1	2	3	4	5
0	25	100	150	250	250	100
1	100	25	250	500	350	200
2	150	250	25	150	150	200
3	250	500	150	25	500	500
4	250	350	150	500	25	500
5	100	200	200	500	500	25

Bandwidth Matrix

The available bandwidth between regions for the simulated application. Units in Mbps

Region\Region	0	1	2	3	4	5
0	2.000	1.000	1.000	1.000	1.000	1.000
1	1.000	800	1.000	1.000	1.000	1.000
2	1.000	1.000	2.500	1.000	1.000	1.000
3	1.000	1.000	1.000	1.500	1.000	1.000
4	1.000	1.000	1.000	1.000	500	1.000
5	1.000	1.000	1.000	1.000	1.000	2.000

3 Cenários e Resultados

Nessa seção serão mostrados os parâmetros usados e resultados obtidos durante as simulações, isso visando explicar melhor o caminho percorrido até chegar na melhor configuração para a aplicação descrita na seção anterior. Em todas as subseções dessa seção é utilizada uma mesma forma como nomenclatura para expressar os parâmetros da simulação, onde os dois primeiros possuem três variações cada um e o terceiro possui seis, conforme abaixo:

ServiceBrokerPolicy_LoadBalancingPolicy_DataCentersRegion

Onde, podem variar conforme segue:

ServiceBrokerPolicy:

- CDC: Closest Data Center
- OPT: Optimize Response Time
- RecDin: Reconfigure Dinamically

LoadBalancingPolicy:

- RR: Round Robin
- ES: Equally Spread Current Execution Load
- TH: Throttled

DataCentersRegion:

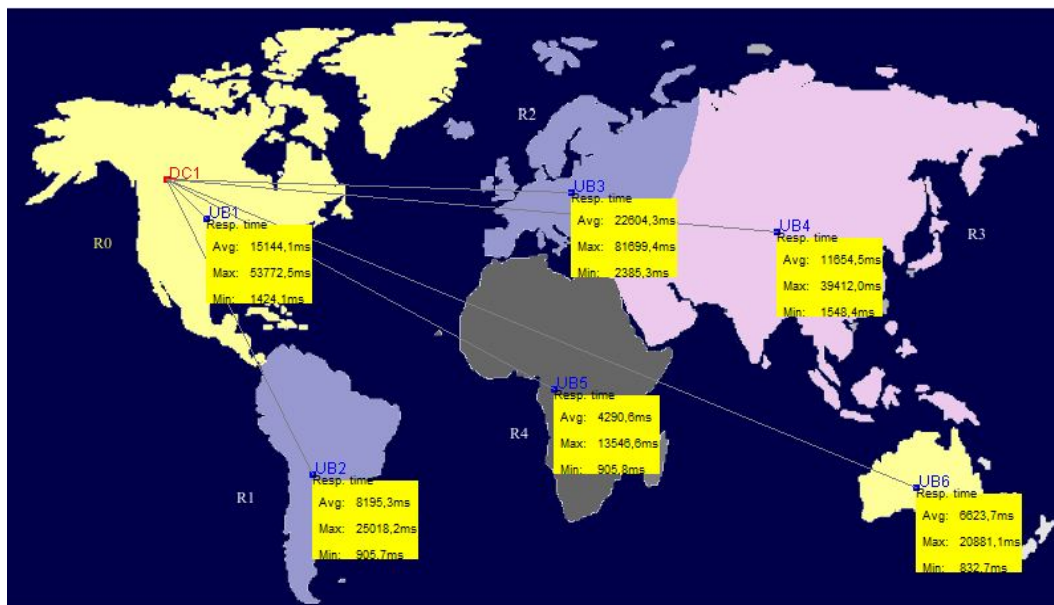
- R0: Region 0 - America do Norte e America Central
- R1: Region 0 - America do Sul
- R2: Region 0 - Europa
- R3: Region 0 - Ásia
- R4: Region 0 - África
- R5: Region 0 - Oceania

3.1 Cenário 1

Nesse primeiro cenário será feito o uso de um único data center com 50 VMs. Nele foram feitas varias simulações com duração de 1 dia. As primeiras simulações visavam descobrir qual a melhor localização para o data center segundo o perfil da aplicação e um resumo de seu resultado pode ser visto a seguir:

Parâmetros	Tempo de Resposta Médio (ms)	Tempo de Processamento Médio (ms)	Custo VM \$	Custo Transferência de Dados \$	Custo Total \$
CDC_RR_R0	14.881,12	3,00	120,05	512.737,57	512.857,62
CDC_RR_R1	20.814,53	3,00	120,05	512.737,57	512.857,62
CDC_RR_R2	16.785,97	3,00	120,05	512.737,57	512.857,62
CDC_RR_R3	20.188,05	3,00	120,05	512.737,57	512.857,62
CDC_RR_R4	20.937,82	3,00	120,05	512.737,57	512.857,62
CDC_RR_R5	20.368,89	3,01	120,05	512.737,57	512.857,62

Como se pode ver, a melhor localização para um único data center é na região 0, ou seja América do Norte. Isso ocorre devido essa região possuir em nossas simulações a maior quantidade de usuários online. Assim, esse resumo mostra que os data centers devem ser alocados próximos ao local de origem de maior carga para a aplicação. Os resultados alterando os outros parâmetros serem mostrados somente nas subseções seguintes devido suas alterações nesse cenário não terem obtido uma diferença significativa. Consequentemente, o mapa resultante da melhor configuração ficou conforme a figura seguinte:



3.2 Cenário 2

Nesse segundo cenário serão utilizados dois data centers com 25 VMs cada um. Essa configuração visa descobrir se aumentar quantidade de data centers sem aumentar a quantidade de máquinas virtuais para uma mesma aplicação é melhor. O resultado quanto a posição dos dois data centers e a melhora no tempo de resposta podem ser verificados a seguir:

Parâmetros	Tempo de Resposta Médio (ms)	Tempo de Processamento Médio (ms)	Custo VM/DC1 \$	Custo Transferência de Dados/DC1 \$	Custo VM/DC2 \$	Custo Transferência de Dados/DC2 \$	Custo Total \$
CDC_RR_R0R0	14.879,93	3,00	60,03	258.514,09	60,03	254.223,48	512.857,62
CDC_RR_R0R1	15.057,94	3,00	60,03	463.594,06	60,03	49.143,51	512.737,57
CDC_RR_R0R2	11.097,99	3,00	60,03	277.001,02	60,03	235.736,55	512.857,62
CDC_RR_R0R3	14.302,82	3,00	60,03	441.723,42	60,03	71.014,15	512.857,62
CDC_RR_R0R4	15.037,43	3,00	60,03	488.847,80	60,03	23.889,77	512.737,57
CDC_RR_R0R5	14.621,39	3,00	60,03	473.194,83	60,03	39.542,74	512.737,57

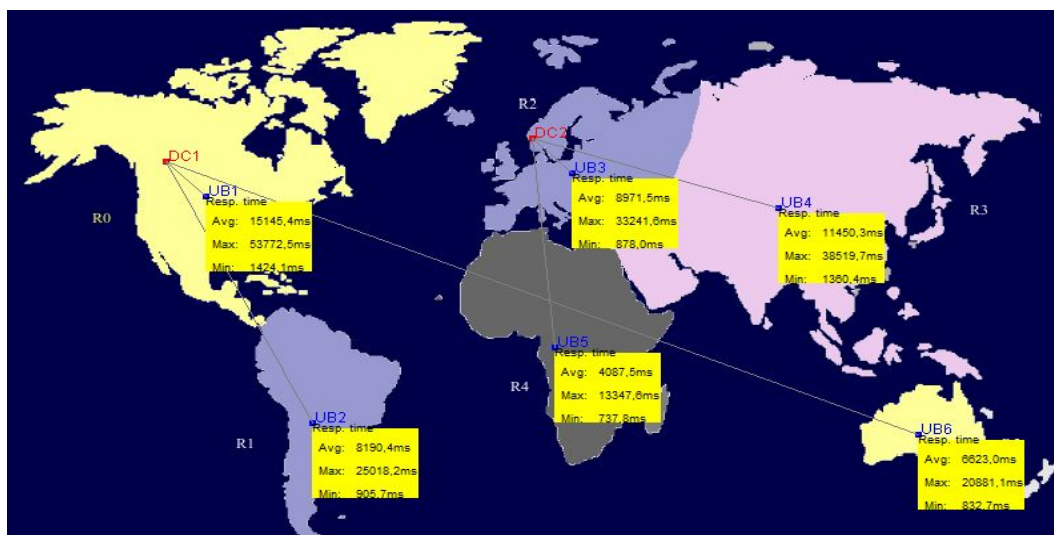
Outra vez o melhor tempo de resposta ocorreu quando as localizações dos data centers coincidiram com as maiores cargas de usuário. Conseqüentemente para essas posições foram realizadas duas novas verificações, uma modificando a política do agente de serviço e a outra modificando a política

de balanceamento de carga em cada data center o resultado é mostrado em sequencia:

Parâmetros	Tempo de Resposta Médio (ms)	Tempo de Processamento Médio (ms)	Custo VM/DC1 \$	Custo Transferência de Dados/DC1 \$	Custo VM/DC2 \$	Custo Transferência de Dados/DC2 \$	Custo Total \$
CDC_RR_R0R2	11.097,99	3,00	60,03	277.001,02	60,03	235.736,55	512.857,62
ORT_RR_R0R2	11.110,92	3,00	60,03	277.001,02	60,03	235736,55	512.857,62
RecDin_RR_R0R2	11.103,00	3,00	60,03	277.001,02	60,03	235.736,55	512.857,62

Parâmetros	Tempo de Resposta Médio (ms)	Tempo de Processamento Médio (ms)	Custo VM/DC1 \$	Custo Transferência de Dados/DC1 \$	Custo VM/DC2 \$	Custo Transferência de Dados/DC2 \$	Custo Total \$
CDC_RR_R0R2	11.097,99	3,00	60,03	277.001,02	60,03	235.736,55	512.857,62
CDC_ES_R0R2	11.098,19	3,00	60,03	277.001,02	60,03	235736,55	512.857,62
CDC_TH_R0R2	11099,74	3,00	60,03	277.001,02	60,03	235.736,55	512.857,62

Bom, vemos que o melhor resultado continua sendo o CDC_RR_R0R2. Assim o mapa desse cenário é conforme a figura a seguir:



3.3 Cenário 3

Nesse terceiro cenário será coberto o restante dos cenários propostos em [1]. Iremos aumentar o número de data centers para três e posteriormente

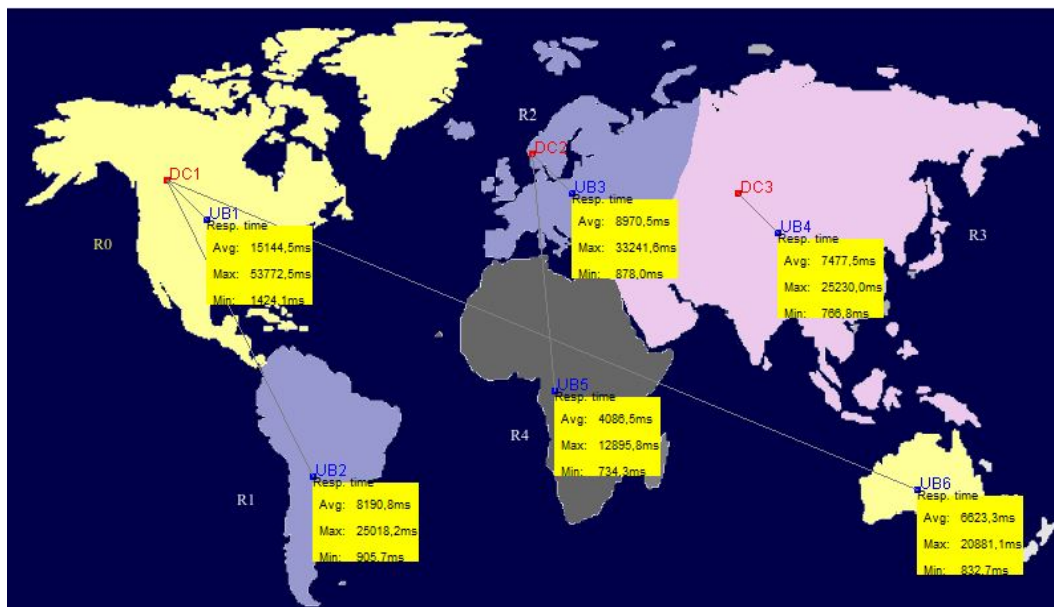
testaremos com a quantidade de VMs de 75, 50 e 25. O novo data center será colocado na região 4, conforme o resultado das simulações anteriores essa será a melhor região devido possuir a terceira maior carga de usuários da aplicação. Com isso foram feitos alterando somente o numero de VMs e os resultados são mostrados abaixo:

VMs	Parâmetros	Tempo de Resposta Médio (ms)	Tempo de Processamento Médio (ms)	Custo Total VM \$	Custo Total de Transferência de Dados/DC1 \$	Custo Total \$
25	CDC_RR_R0R2R3	10.548,42	3,00	180,08	512.737,57	512.917,65
50	CDC_RR_R0R2R3	10548,58	3,00	360,15	512737,57	513097,72
75	CDC_RR_R0R2R3	10548,52	3,00	540,23	512.737,57	513.277,80

Conforme pode ser observado, o aumento da quantidade de máquinas virtuais não ofereceu melhoras quanto ao tempo de resposta. Isso se justifica pois como cada data center possui 40 máquinas físicas as máquinas virtuais não sofrem tantas restrições por hardware compartilhado e assim o que garante a maior parcela do custo total acaba sendo a transferência de dados. Isso ocorre devido a quantidade de dados por requisição foi colocada como 100 KB para esse trabalho. Como já conhecemos o resultado para essa quantidade de VMs quanto as possíveis alterações na política de balanceamento de carga e também na de agente de serviço. Concluímos que a melhor abordagem para os cenários descritos ocorre com as configurações abaixo:

- Política de Balanceamento de Carga: Round Robin
- Política de Agente de serviço: Data Center Mais Próximo
- Localização dos Data Centers: Na região que possuir maior carga de trabalho e que ainda não possua um data center
- Quantidade de Maquinas Virtuais Por DC: 25 VMs

Vale ressaltar que o grande responsável por esse resultado é o gargalo que ocorreu na transmissão de dados. E esse ocorreu devido a natureza da simulação, uma vez que se trata de simulações de requisição web e cada requisição de página web possui em média 100 KB. Assim resultados diferentes poderiam ser observados caso retirasse esse gargalo do tamanho da página web. Contudo, já não se trataria da simulação de uma rede social em larga escala. Com isso, obtemos finalmente nossa melhor configuração para os cenários descritos no artigo já citado. O mapa para essa configuração pode ser visto a seguir:



4 Conclusão

Com esse trabalho foi possível aprender a utilizar as ferramentas de simulação CloudSim e CloudAnalist. Entretanto, o grande ganho desse trabalho foram as lições aprendidas sobre boas heurísticas para configuração da nuvem visando aplicações web. Também foi visto por meio das simulações que o tempo de resposta da requisição web tem uma melhora mais significativa quando a carga de trabalho está melhor distribuída e ainda quando os data centers que irão atender as demandas encontram-se próximos as mesmas. Outro aprendizado foi que nem sempre o aumento da quantidade de máquinas virtuais melhora o desempenho da aplicação. E isso ocorre pois montar e desmontar uma máquina virtual também consome processamento. Assim, não adianta alocar mais máquinas virtuais pensando que somente isso irá melhorar o desempenho do serviço. É preciso principalmente avaliar qual o melhor plano de distribuição de carga que pode ser feito.

Referências

- [1] B. Wickremasinghe, R.N. Calheiros, and R. Buyya. Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications. In *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*, pages 446–452. IEEE, 2010.