



# ARQUITETURA DE COMPUTADORES EM ESCALA WAREHOUSE

Lecio Charlles B. Alves  
Breno de Castro Pimenta  
Tiago Coelho Magalhães

UF *m* G

# TÓPICOS COMPUTADORES EM ESCALA WAREHOUSE

- Modelos de programação e carga de trabalho
- Arquitetura
- Infraestrutura física e custos
- Computação em nuvem

# INTRODUÇÃO

Computador em escala warehouse (WSC, *Warehouse-Scale Computers*):

- Custo total (construção, servidores) \$150M, conectam/mantém 50k-100k servidores
- Fornece serviços de Internet
  - Pesquisa, redes sociais, mapas online, compartilhamento de vídeo, compras online, e-mail, computação em nuvem, etc.
- Diferenças com datacenters:
  - Datacenters consolidam diferentes máquinas e software em um local
  - Os datacenters enfatizam as máquinas virtuais e a heterogeneidade de hardware para atender a diversos clientes
- Diferenças com “clusters” de HPC (*High-Performance Computing*):
  - Os clusters têm processadores e rede de alto desempenho
  - Os clusters enfatizam o paralelismo no nível do thread, os WSCs enfatizam o paralelismo no nível da solicitação

# FATORES IMPORTANTES DE PROJETO

Requisitos **compartilhados** com servidores:

- Custo-desempenho: trabalho realizado/USD
  - Pequenas economias somam redução de 10% do custo de capital \$15M
- Eficiência energética: trabalho / joule
  - Afeta a distribuição de energia e o resfriamento. A potência de pico afeta o custo.
- Confiabilidade por meio de redundância:  $> 99,99\%$  de inatividade/ano = 1h
  - Além de “quatro noves”, vários eventos de máscara WSC que eliminam uma WSC
- Rede E/S: com público e entre várias WSC
- Cargas de trabalho de processamento interativo e em lote: pesquisa e Map-Reduce

# FATORES IMPORTANTES DE PROJETO

## Requisitos **não compartilhados** com servidores

- Amplo paralelismo computacional não é importante
  - A maioria dos empregos é totalmente independente
  - DLP aplicado ao armazenamento; (em servidores, para a memória)
  - “Paralelismo de nível de solicitação”, SaaS, pouca necessidade de comunicação/sincronização.
- Contagem de custos operacionais
  - O consumo de energia é uma restrição primária, não secundária, ao projetar o sistema. (em servidores, só preocupação do pico de potência não exceder as especificações)
  - Custos amortizados em 10+ anos. Custos de energia, potência, refrigeração > 30% total
- Escala e suas oportunidades e problemas
  - Oportunidades: pode se dar ao luxo de construir sistemas personalizados, uma vez que WSC exige compra por volume (descontos por volume)
  - Problemas: falha servidores

# FATORES IMPORTANTES DE PROJETO

**Exemplo:** calcule a disponibilidade de um serviço executando nos 2.400 servidores na Figura 6.1 (livro referência). Ao contrário de um serviço em um WSC real, neste exemplo os serviços não podem tolerar falhas de hardware ou software. Considere que o tempo para reiniciar o software seja de cinco minutos e o tempo para reparar o hardware seja de uma hora.

# FATORES IMPORTANTES DE PROJETO

Número aprox. de eventos no 1º ano	Causa	Consequência
1 ou 2	Falhas da rede elétrica	Perda de energia em todo o WSC; não derruba o WSC se o UPS e os geradores funcionarem (ou geradores funcionam cerca de 99% do tempo).
4	Atualizações de cluster	Indisponibilidade planejada para atualizar a infraestrutura, muitas vezes para melhorias de rede, como recabeamento, atualizações de firmware, e assim por diante. Há cerca de 9 indisponibilidades de cluster planejadas para cada indisponibilidade não planejada.
1.000	Falhas de disco rígido	2-10% de taxa anual de falha de disco (Pinheiro, 2007).
	Discos lento	Ainda opera, mas roda 10-20x mais lentamente.
	Memórias ruins	Um erro incorrigível de DRAM por ano (Schroeder <i>et al.</i> , 2009).
	Máquinas mal configuradas	A configuração leva a ~30% de interrupção do serviço (Barroso e Hölzle, 2009).
	Máquinas não confiáveis	1% dos servidores reiniciados mais de uma vez por semana (Barroso e Hölzle, 2009).
5.000	Crashes individuais de servidor	A máquina é reinicializada, geralmente leva cerca de 5 minutos.

**FIGURA 6.1** Lista de indisponibilidades e anomalias com as frequências aproximadas de ocorrência no primeiro ano de um novo cluster de 2.400 servidores.



# FATORES IMPORTANTES DE PROJETO

**Resposta:** Podemos estimar a disponibilidade de serviço calculando o tempo de indisponibilidades devidas a falhas de cada componente. Vamos, conservadoramente, tomar o menor número de cada categoria na Figura 6.1 (livro referência) e dividir as 1.000 indisponibilidades igualmente entre os quatro componentes. Nós ignoramos discos lentos — o quinto componente das 1.000 indisponibilidades —, já que eles prejudicam o desempenho, mas não a disponibilidade e falhas na rede elétrica, já que o sistema de fornecimento ininterrupto de energia (*Uninterrupted Power Supply* — UPS) oculta 99% delas

$$\text{Horas Indisponibilidade}_{\text{serviço}} = (4 + 250 + 250 + 250) \times 1 \text{ hora} + (250 + 5.000) \times 5 \text{ minutos}$$

$$= 754 + 438 = 1.192 \text{ horas}$$

Uma vez que existem  $365 \times 24$  ou 8.760 horas em um ano, a disponibilidade é:

$$\text{Disponibilidade}_{\text{sistema}} = \frac{(8.760 - 1.192)}{8.760} = \frac{7.568}{8.760} = 86\%$$



# CLUSTERS E HPC VS WSC

- Clusters de computadores: precursores da WSC
  - Computadores independentes, LAN, interruptores disponíveis no mercado
  - Para cargas de trabalho com baixa necessidade de comunicação, os clusters são mais econômicos do que os multiprocessadores de memória compartilhada (precursor do multicore)
- HPC (computação de alto desempenho):
  - Custo e escala = semelhante ao WSC
  - Mas: processadores e rede muito mais rápidos. Os aplicativos HPC são muito mais interdependentes e têm maior taxa de comunicação
  - Tende a usar hw personalizado (potência e custo de i7 > servidor WSC inteiro)
  - Servidores de trabalhos de longa duração totalmente ocupados por semanas (utilização do servidor WSC = 10% - 50%)

# DATACENTERS VS WSC

## Datacenters

- Foco principal: consolidação de serviços em menos máquinas isoladas
  - Proteção de virtualização de informações confidenciais cada vez mais importante
- Heterogeneidade HW e SW (WSC é homogênea)
- O maior custo é com pessoas para mantê-lo (WSC: servidor é o custo máximo, o custo com pessoas é irrelevante)
- Escala não tão grande quanto WSC: sem benefícios de custo em larga escala

# MODELOS DE PROGRAMAÇÃO E CARGAS DE TRABALHO PARA COMPUTADORES EM ESCALA WAREHOUSE

- Estrutura de processamento em lote mais popular: MapReduce
- Gêmeo de código aberto: *Hadoop*

	4 Ago	6 Mar	7 Set	9 Set
Número de jobs MapReduce	29.000	171.000	2.217.000	3.467.000
Tempo médio de conclusão (segundos)	634	874	395	475
Anos de uso de servidor	217	2.002	11.081	25.562
Dados de entrada lidos (terabytes)	3.288	52.254	403.152	544.130
Dados intermediários (terabytes)	758	6.743	34.774	90.120
Dados de saída escritos (terabytes)	193	2.970	14.018	57.520
Número médio de servidores por serviço	157	268	394	488

**FIGURA 6.2** Uso anual de MapReduce no Google ao longo do tempo.

# MODELOS DE PROGRAMAÇÃO E CARGAS DE TRABALHO PARA COMPUTADORES EM ESCALA WAREHOUSE

- Mapa: aplica uma função fornecida pelo programador a cada registro de entrada lógica
  - Funciona em milhares de computadores
  - Fornece novo conjunto de pares de valores-chave como valores intermediários
- Reduzir: reduz os valores usando outra função fornecida pelo programador
- Exemplo: cálculo de # ocorrências de cada palavra

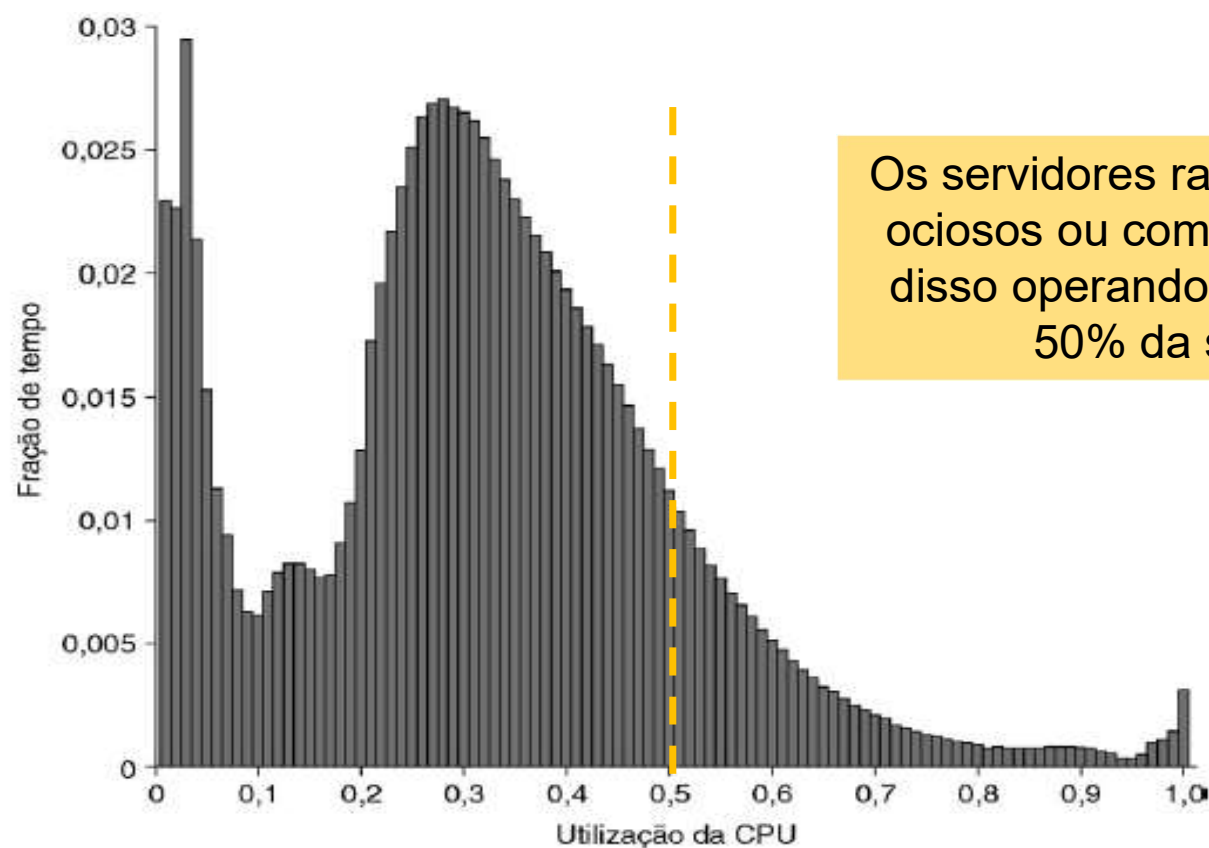
```
map(String key, String value):
    //key: nome do documento
    //value: conteúdo do documento
    para cada palavra w em value:
        EmitIntermediate(w, "1"); // Produz a lista de todas as palavras

reduce(String key, Iterator values):
    //key: uma palavra
    //values: uma lista de counts
    int result = 0;
    para cada v em values:
        result += ParseInt(v); // obtém inteiro do par key-value
    Emit(AsString(result));
```

# MODELOS DE PROGRAMAÇÃO E CARGAS DE TRABALHO PARA COMPUTADORES EM ESCALA WAREHOUSE

- MapReduce runtime environment: cronograma para mapear e reduzir tarefas para nós WSC
  - Quase no final do MapReduce, o sistema inicia as execuções de backup em nós livres e obtém os resultados do que terminar primeiro
- Disponibilidade:
  - Use réplicas de dados em diferentes servidores
  - Use consistência relaxada:
    - Não há necessidade de que todas as réplicas sempre concordem
- Demandas de carga de trabalho
  - Muitas vezes variam consideravelmente

# GOOGLE: DISTRIBUIÇÃO DE UTILIZAÇÃO DE CPU



Os servidores raramente estão completamente ociosos ou completamente utilizados, em vez disso operando a maioria do tempo entre 10-50% da sua utilização máxima.

**FIGURA 6.3** Utilização média da CPU de mais de 5.000 servidores durante um período de seis meses no Google.

# DISTRIBUIÇÃO DE UTILIZAÇÃO DE CPU

**Exemplo:** Como resultado de medições como as da Figura 6.3, o benchmark SPECPower mede o consumo de energia e o desempenho de carga de 0% a 100% em incrementos de 10%.

A única métrica geral que resume esse benchmark é a soma de todas as medidas de desempenho (operações Java por segundo do lado do servidor) dividida pela soma de todas as medições de potência em watts. Assim, cada nível é igualmente provável. Como os números resumem a mudança da métrica se os níveis forem ponderados pelas frequências de utilização na Figura 6.3?



# DISTRIBUIÇÃO DE UTILIZAÇÃO DE CPU

Carga	Desempenho	Watts	Pesos SPEC	Desempenho ponderado	Watts ponderados	Pesos da Fig. 6.3	Desempenho ponderado	Watts ponderados	
100%	2.889.020	662	9,09%	262.638	60	0,80%	22.206	5	
90%	2.611.130	617	9,09%	237.375	56	1,20%	31.756	8	
80%	2.319.900	576	9,09%	210.900	52	1,50%	35.889	9	
70%	2.031.260	533	9,09%	184.660	48	2,10%	42.491	11	
60%	1.740.980	490	9,09%	158.271	45	5,10%	88.082	25	
50%	1.448.810	451	9,09%	131.710	41	11,50%	166.335	52	
40%	1.159.760	416	9,09%	105.433	38	19,10%	221.165	79	
30%	869.077	382	9,09%	79.007	35	24,60%	213.929	94	
20%	581.126	351	9,09%	52.830	32	15,30%	88.769	54	
10%	290.762	308	9,09%	26.433	28	8,00%	23.198	25	
0%	0	181	9,09%	0	16	10,90%	0	20	
Total	15.941.825	4.967		1.449.257	452		933.820	380	
				ssj_ops/Watt	3.210			ssj_ops/Watt	2.454

**FIGURA 6.4** Resultado SPECPower da [Figura 6.17](#) usando os pesos da [Figura 6.3](#) em vez de pesos iguais.

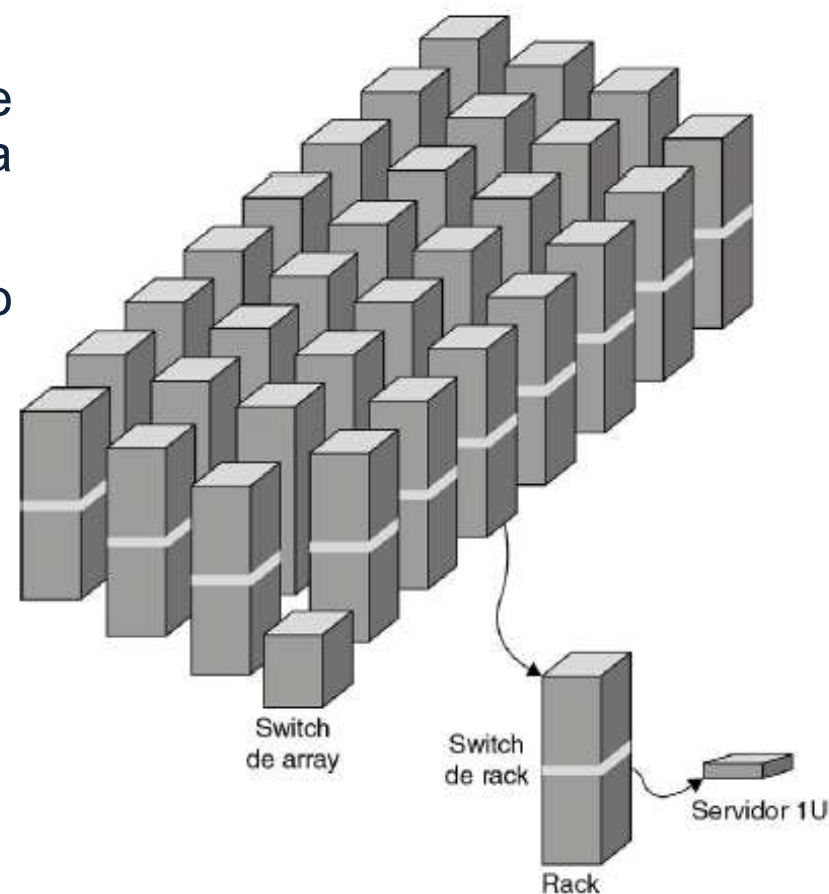
**Os pesos reduzem o resumo do desempenho em 30%**

# ARQUITETURA DE COMPUTADORES EM ESCALA WAREHOUSE

- WSC costuma usar uma hierarquia de redes para interconexão
- Estrutura padrão para manter servidores: rack de 19"
  - Servidores medidos em # unidades de rack (U) que ocupam em um rack. Um U tem 1,75 "de altura
  - Rack de 7 pés 48 U (switch Ethernet de 48 portas popular); \$ 30 / porta
- Os switches oferecem de 2 a 8 uplinks (nível de hierarquia superior)
  - O BW que sai do rack é 6-24 x menor ( $48/8 - 48/2$ ) do que o BW dentro do rack (essa proporção é chamada de "Superinscrição")
- O objetivo é maximizar a localidade de comunicação em relação ao rack
  - Penalidade de comunicação entre racks diferentes

# ARQUITETURA DE SWITCHES EM ESCALA WAREHOUSE

- Idealmente: desempenho de rede equivalente a um switch *high-end* para servidores de 50k
- Custo por porta: switch commodity projetado para 50 servidores



# ARMAZENAMENTO

- Desenho natural: preencher o rack com servidores + switch Ethernet; Armazenamento??
- Opções de armazenamento:
  - Use discos dentro dos servidores, ou
  - Armazenamento conectado à rede (servidores remotos) através do Infiniband
- WSCs geralmente dependem de discos locais
  - Google File System (GFS) usa discos locais e mantém pelo menos três réplicas, cobrindo falhas no disco local, energia, racks e clusters
- Cluster (terminologia)
  - Definição na seção 6.1: WSC = cluster muito grande
  - Barroso: próximo agrupamento de computadores, ~ 30 racks
  - Neste capítulo:
    - array: coleção de racks
    - cluster: original significando qualquer coisa, desde uma coleção de computadores em rede dentro de um rack até uma WSC inteira

# SWITCH DE ARRAYS

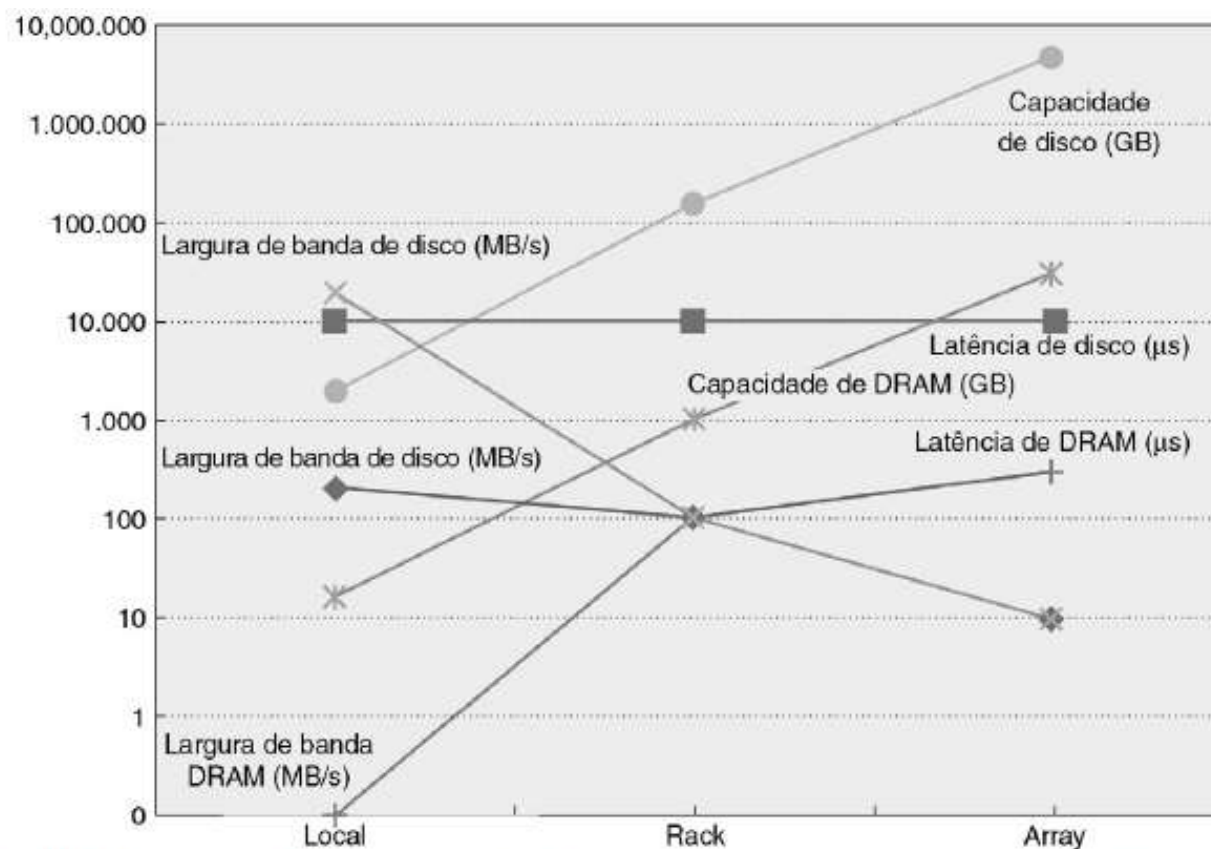
- Switch que conecta uma série de racks
- Muito mais caro do que um switch Ethernet de 48 portas
- O switch de array deve ter largura de banda maior para reduzir o problema de oversubscription.

# HIERARQUIA DE MEMÓRIA

	Local	Rack	Array
Latência de DRAM (microssegundos)	0,1	100	300
Latência de disco (microssegundos)	10.000	11.000	12.000
Largura de banda de DRAM (MB/s)	20.000	100	10
Largura de banda de disco (MB/s)	200	100	10
Capacidade de DRAM (GB)	16	1.040	31.200
Capacidade de disco (GB)	2.000	160.000	4.800.000

**FIGURA 6.6** Latência, largura de banda e capacidade da hierarquia de memória de um WSC (Barroso e Hölzle, 2009).

# HIERARQUIA DE MEMÓRIA



**FIGURA 6.7** Gráfico de latência, largura de banda e capacidade da hierarquia de memória de um WSC para os dados da [Figura 6.6](#) (Barroso e Hölzle, 2009).



# HIERARQUIA DE REDE

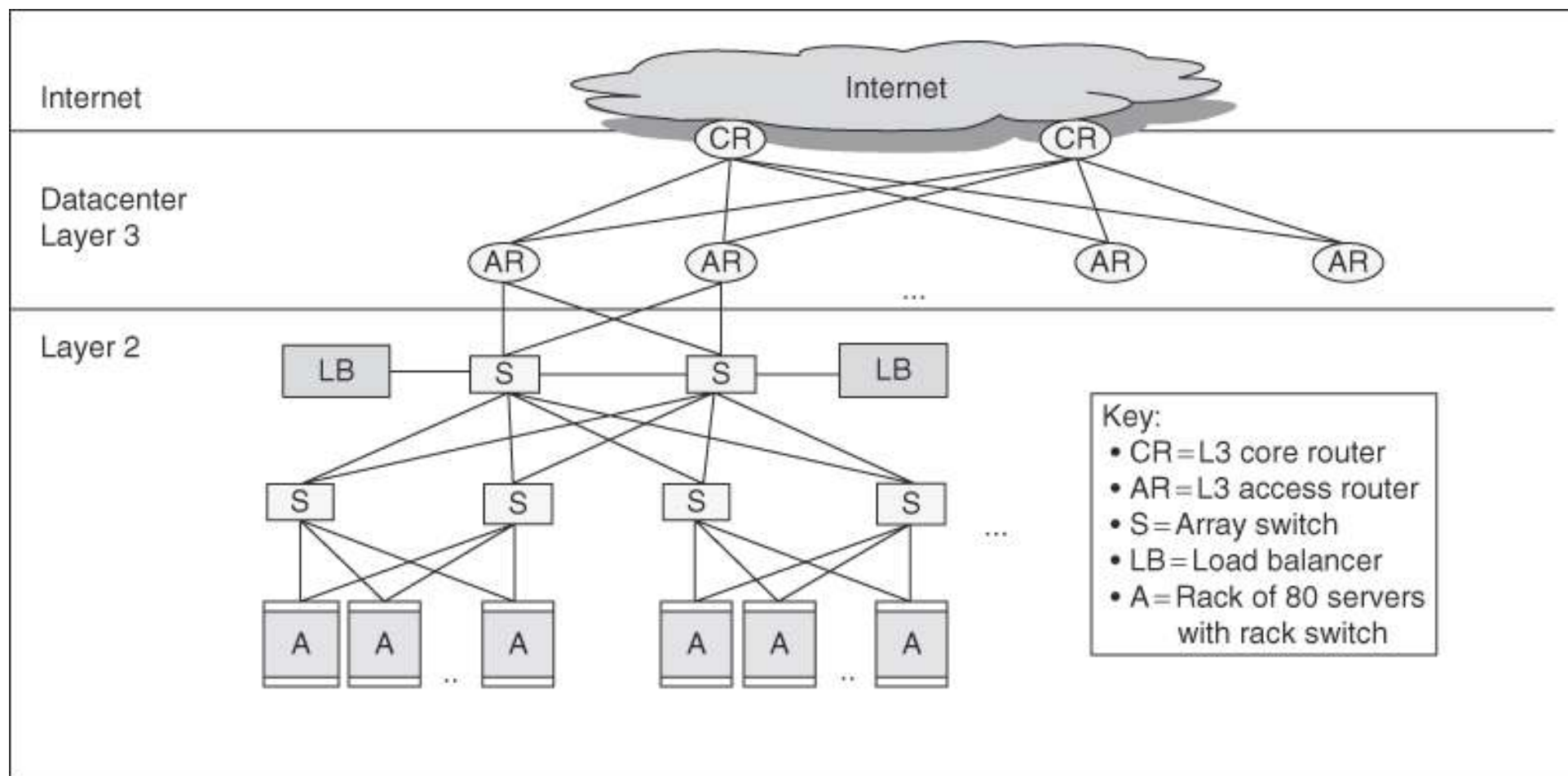


FIGURA 6.8 A rede de camada 3 usada para ligar arrays entre si e para a internet (Greenberg et al., 2009). Alguns WSCs usam um roteador de limite separado para conectar a internet aos switches de nível 3 do datacenter.<sup>1</sup>

# LATÊNCIA DE ACESSO

- Exemplo: qual é o tempo médio de acesso a memória Assumindo 90% de acessos locais a um servidor, 9% fora mas dentro do rack e 1% fora do rack mas dentro do array.
- Considerando o tempo de acesso sendo 0,1us ; 100us e 300us os acessos Local, no rack e no Array respectivamente.
- $90\% \cdot 0.1 + 9\% \cdot 100 + 1\% \cdot 300 = 12.09 \text{ us}$
- podemos ver o quão grande é a diferença entre o acesso local e o do sistema como um todo, mostrando que há uma boa margem para otimizações.

# VAZÃO DE DADOS

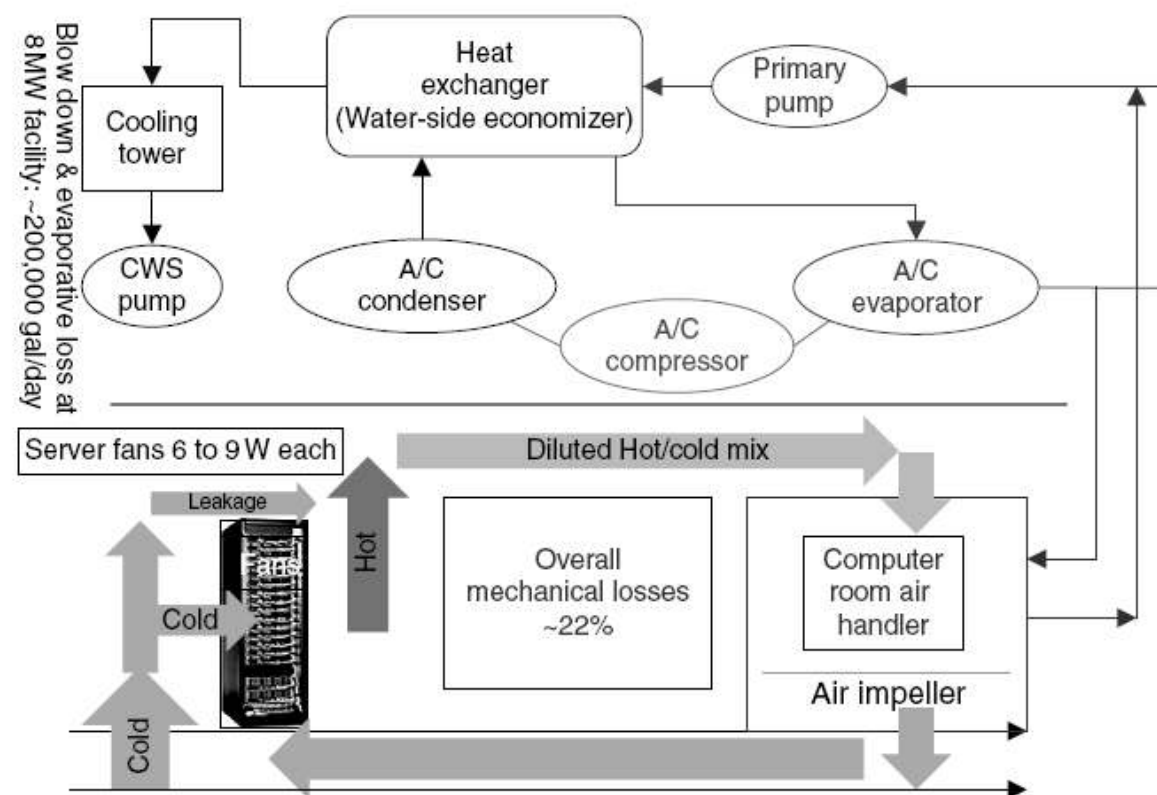
- Exemplo: Tempo maximo de transferencia de 1GB em volume de dados:

Localidade	Disco → disco	Memoria → memoria
Localmente	5 s	0.05 s
Dentro do mesmo rack	10 s	10 s
No mesmo array	100 s	100 s

- A rede é a grande limitadora da banda, o a localidade espacial tem papel importante aqui

# CUSTO DE INFRAESTRUTURA

- Deve levar em conta:
  - Proximidade a backbones de internet
  - Custo da eletricidade
  - Impostos e taxas
  - Risco de ocorrer catástrofes naturais



# CUSTO DE INFRAESTRUTURA

- Resfriamento:
  - O ambiente deve facilitar o resfriamento, as instalações devem ser resfriadas para se manterem a por volta de 22°C
  - O local das instalações deve possuir grande quantidade de água disponível para resfriamento.
  - A energia gasta com resfriamento pode chegar a mais da metade da energia gasta no resto da infraestrutura
- Distribuição de energia nos componentes:
  - Processador: 33%; RAM: 30%; discos: 10%; rede: 5%; outros: 22%

# EFICIÊNCIA DE UM WSC

A eficiência de um WSC pode ser medida através de:

## **Utilização efetiva da computação (PUE)**

- Expressa por  $(\text{energia total no WSC}) / (\text{energia gasta no processamento})$
- O ideal é que seja igual a 1: toda energia utilizada ira para o processamento.

# EFICIÊNCIA DE UM WSC

## Performance

- Uma medida de performance é a latência, percebida pelos usuários
- A tabela abaixo mostra um experimento em que foram alterados os tempos de resposta de um buscador, isso alterou o comportamento do usuário

Server delay (ms)	Increased time to next click (ms)	Queries/ user	Any clicks/ user	User satisfaction	Revenue/ user
50	--	--	--	--	--
200	500	--	-0.3%	-0.4%	--
500	1200	--	-1.0%	-0.9%	-1.2%
1000	1900	-0.7%	-1.9%	-1.6%	-2.8%
2000	3100	-1.8%	-4.4%	-3.8%	-4.3%

**Figure 6.12** Negative impact of delays at Bing search server on user behavior Schurman and Brutlag [2009].



# CUSTO DE UM WSC

Exemplo de 2010:

Tamanho da instalação (carga crítica em watts)	8.000.000
Uso médio de energia (%)	80%
Efetividade do uso de energia	1.45
Custo da energia (\$ /kWh)	\$ 0,07
% da infraestrutura energética e de refrigeração (% do custo total da instalação)	82%
<b>CAPEX para a instalação (não incluindo equipamento de TI)</b>	<b>\$ 88.000.000</b>
Número de servidores	45.978
Custo/servidor	\$ 1.450
<b>CAPEX para os servidores</b>	<b>\$ 66.700.000</b>
Número de switches de rack	1150
Custo/switch de rack	\$ 4.800
Número de switches de array	22
Custo/switch de array	\$ 300.000
Número de switches de nível 3	2
Custo/switch de nível 3	\$ 500.000
Número de roteadores de limite	2
Custo/roteador de limite	\$ 144.800
<b>CAPEX para equipamento de rede</b>	<b>\$ 12.810.000</b>
<b>CAPEX total para WSC</b>	<b>\$ 167.510.000</b>
Tempo de amortização do servidor	3 anos
Tempo de amortização de rede	4 anos
Tempo de amortização da instalação	10 anos
Custo anual do dinheiro	5%

# “UTILITY COMPUTING”

- É a computação que é produzida, mensurada e vendida por utilidade, isto é, o usuário paga por tempo de CPU, não por tempo comum

# EVOLUÇÃO DA INFRAESTRUTURA

# EVOLUÇÃO DA INFRAESTRUTURA

## Estudo de caso da Google

- Empresa que está a quase duas décadas trabalhando e implementando o estado da arte dessas infraestruturas.
- Gera uma grande quantidade de material de pesquisa para a área.

# EVOLUÇÃO DA INFRAESTRUTURA

Google DataCenter (2000)





# EVOLUÇÃO DA INFRAESTRUTURA

Servidor da Google (2009)



([http://news.cnet.com/8301-1001\\_3-10209580-92.html#ixzz0yo8bhTOH](http://news.cnet.com/8301-1001_3-10209580-92.html#ixzz0yo8bhTOH))

# EVOLUÇÃO DA INFRAESTRUTURA

## Problema:

- Com o crescimento se fez necessário a construção de DataCenters para comportar os WSCs. No entanto a construção dessas estruturas inicialmente traziam vários problemas de integração dos projetos de engenharia, do Civil ao Mecânico.



# EVOLUÇÃO DA INFRAESTRUTURA

## **Problema:**

- Com o crescimento se fez necessário a construção de DataCenters para comportar os WSCs. No entanto, a construção dessas estruturas inicialmente traziam vários problemas de integração dos projetos de engenharia, do Civil ao Mecânico.

## **Solução:**

- Em 2009 a google liberou o que seria o estado da arte de infraestrutura de WSCs em 2005: Modularização em contêineres.

# EVOLUÇÃO DA INFRAESTRUTURA

## Problema:

- Com o crescimento se fez necessário a construção de DataCenters para comportar os WSCs. No entanto a construção dessas estruturas inicialmente traziam vários problemas de integração dos projetos de engenharia, do Civil ao Mecânico.

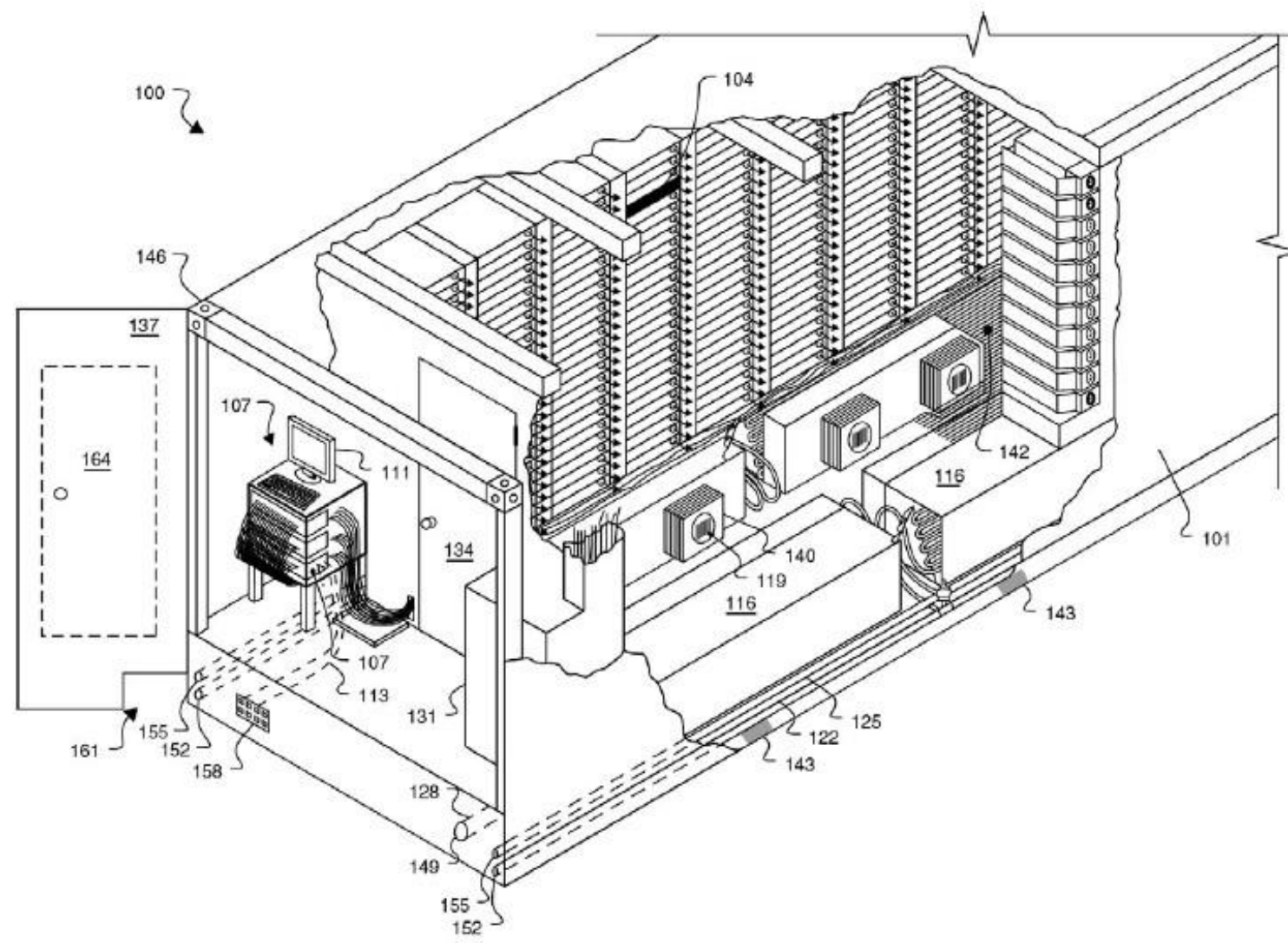
## Solução:

- Em 2009 a google liberou o que seria o estado da arte de infraestrutura de WSCs em 2005: Modularização em contêineres.
- Tanto a Google quanto a Microsoft construíram WSCs usando contêineres.

# EVOLUÇÃO DA INFRAESTRUTURA

## Google Contêiner (2005)

- Cada contêiner possui duas fileiras cada uma com 29 racks, sendo que cada rack é composto por 20 servidores empilhados.
- Cada contêiner, então, contém 1160 servidores.
- E um WSC contém em média 45 contêineres, sendo composto então por aproximadamente 50.000 servidores.
- Cada contêiner é independente, e as únicas conexões externas são rede, energia e água.



# EVOLUÇÃO DA INFRAESTRUTURA

## **Próximo passo:**

A própria Google abandonou a ideia de contêiners nos dois anos seguintes e criou formas a partir das análises de benefícios tragos pela modularização.

# EVOLUÇÃO DA INFRAESTRUTURA

## **Próximo passo:**

A própria Google abandonou a ideia de contêiners nos dois anos seguintes e criou formas a partir das análises de benefícios tragos pela modularização.

## **Exemplo de novas evoluções :**

Aproveitamento energético, na virada do século a cada 1W utilizado por um servidor, eram gastos outros 1-2W para transformação de energia e refrigeração.

Hoje em 2020, a cada 1W utilizado pelo servidor, apenas um décimo de watt é utilizado para transformação e refrigeração.



# EVOLUÇÃO DA INFRAESTRUTURA

Google (2020)  
DataCenter  
Eemshaven, Holanda

- Construção (2016)  
€ 600 milhões.
- Expansão (2018)  
€ 500 milhões
- Anunciada em junho de 2019 uma nova unidade em Agriport por aproximadamente € 1,4 bilhões



(<https://www.google.com/intl/pt-BR/about/datacenters/locations/eemshaven/>)

# EVOLUÇÃO DA INFRAESTRUTURA

“No início do século, eram prédios cheios de computadores e sistemas de refrigeração. Hoje ao olhar para um datacenter desses é o mesmo que olhar para um único computador. Pensar dessa forma holística é o que está por trás de quase todas as inovações. (...) um serviço pode ser executado em milhares de servidores e essa coleção de servidores é um computador em si.”

**Luiz André Barroso, Webinar Rede Nacional de Pesquisa, 2020.**

# COMPUTAÇÃO EM NUVEM



# COMPUTAÇÃO EM NUVEM

## Fator Econômico:

*“Com a maior escala vieram maiores economias de escala.”*

Hamilton (2010) realizou um estudo em 2006, que comparou um WSC com um datacenter tradicional com somente 1.000 servidores, reportou as seguintes vantagens:

- Redução de 5,7 vezes nos custos de armazenamento.
- Redução de 7,1 vezes nos custos administrativos.
- Redução de 7,3 vezes nos custos de rede.

# COMPUTAÇÃO EM NUVEM

## Fator Econômico:

*“Com a maior escala vieram maiores economias de escala.”*

Hamilton (2010) realizou um estudo em 2006, que comparou um WSC com um datacenter tradicional com somente 1.000 servidores, reportou as seguintes vantagens:

- Redução de 5,7 vezes nos custos de armazenamento.
- Redução de 7,1 vezes nos custos administrativos.
- Redução de 7,3 vezes nos custos de rede.

## Fator Eficiência:

- *DataCenter Tradicional* utilizava em média 10% do poder de processamento.
- WSC utilizavam, antes de 2010, em média 50% do poder de processamento.

# COMPUTAÇÃO EM NUVEM

A **Amazon**, em 2006, lança a **AWS** (Amazon Web Services) com dois serviços principais:

- S3 (Amazon Simple Storage Service).
- EC2 (Amazon Elastic Computer Cloud).

# COMPUTAÇÃO EM NUVEM

## Diferenciais:

- Máquinas virtuais:
  - Segurança;
  - Facilidade de distribuição de software;
  - Controle de uso
    - *“Virtualização significa que o desempenho oferecido não precisa ser um múltiplo inteiro do desempenho do hardware.”*
- Custo muito baixo, USD 0,10 / hora por instância.
- Nenhum contrato necessário.
- Inicialmente, em 2006, não oferecia garantias, o que não era necessário, pois era uma ótima qualidade disponível e boa parte dos softwares pré instalados eram opensource. Esses fatores tornaram a computação como serviço muito mais econômica para a Amazon e para seus clientes. (Não são mais válidos)

# COMPUTAÇÃO EM NUVEM

## Diferenciais:

- Elasticidade:
  - Possibilidade de expansão conforme a demanda;
  - **Exemplo:**
    - O **FarmVille**, da Zynga, um jogo do Facebook. Antes do FarmVille ser anunciado, o maior jogo social tinha cerca de cinco milhões de jogadores por dia. Enquanto o FarmVille após o lançamento teve:
      - Em **4 dias**, tinha 1 milhão de jogadores;
      - Em **60 dias**, tinha 10 milhões de jogadores;
      - Em **270 dias**, tinha 28 milhões de jogadores online por dia e 75 milhões por mês;

O jogo foi implementado no AWS, e conseguiram crescer de acordo com o número de usuários.

# COMPUTAÇÃO EM NUVEM

## Conclusão:

A computação em nuvem tornou os benefícios do WSC disponíveis para todos. Ela oferece associatividade de custo com a ilusão de escalabilidade infinita sem custo extra para o usuário.