



# O Papel da Rede Adaptativa nas Arquiteturas Orientadas a Serviço

## *Um White Paper do META Group*

*“Para as redes continuarem a suprir as aplicações com sucesso, não basta adicionar mais capacidade ou conectividade... É necessário um alto grau de automação, integração e planejamento arquitetônico fundados sobre a inteligência baseada em rede.”*

### **O Papel Fundamental das Redes**

As equipes de infra-estrutura lutam para acompanhar as demandas das aplicações novas e preexistentes no data center. A própria empresa a que a organização de TI (OTI) serve está sob pressão para fornecer mais com os mesmos recursos, ou até menos – novas normas de segurança aumentam a demanda de controles corporativos e de proteção da informação, e questões de conformidade, redução de custos, gerenciamento do desempenho corporativo e processos interempresariais estão na pauta dos líderes corporativos. Cada vez mais, a empresa depende da TI para aprimorar suas competências.

Dentro dos muros da OTI, o ritmo das novas aplicações está se acelerando. Impulsionados por uma demanda cada vez maior, os desenvolvedores estão adotando metodologias de desenvolvimento novas, e mais ágeis, como a RAD (desenvolvimento rápido de aplicação), ou técnicas extremas de programação, e as equipes de programação estão liberando uma inundação de novos programas e serviços executados nas redes corporativas como uma extensão externa aos clientes e parceiros comerciais. Uma empresa farmacêutica internacional tem mais de 8.000 aplicações, e esse número cresce a cada mês. Para o suporte dessas aplicações, existentes e futuras, a questão que se apresenta é como as organizações de TI irão atingir os objetivos de flexibilidade, assegurando o acesso a essas aplicações e garantindo a constante disponibilidade, escalonamento e desempenho. Ou isso se torna extremamente caro ou, em muitos casos, é impossível agregar essas funções nas aplicações – e, dessa forma, cumprir o papel essencial da rede na unificação da aplicação e sua infra-estrutura correlata.

*Sem rede, as empresas não têm aplicações e, sem as aplicações, as empresas não funcionam.*

Isso é uma grande mudança em relação ao passado, quando as redes freqüentemente eram consideradas secundárias e vistas apenas como um modo de conectividade. Mesmo assim, sem rede, as empresas não têm aplicações e, sem as aplicações, as empresas não funcionam. Para muitos, a perda da rede equivale diretamente à perda de receita (veja a figura 1). Mesmo assim, tradicionalmente, as redes se apresentam inflexíveis e incapazes de responder aos desafios das aplicações que sempre estiveram presentes, sejam eles buracos na segurança, incapacidade de escalar, problemas de desempenho na WAN ou paralisações. Isso porque falta às redes a inteligência necessária para se adaptar e fornecer de forma econômica serviços benéficos às aplicações.

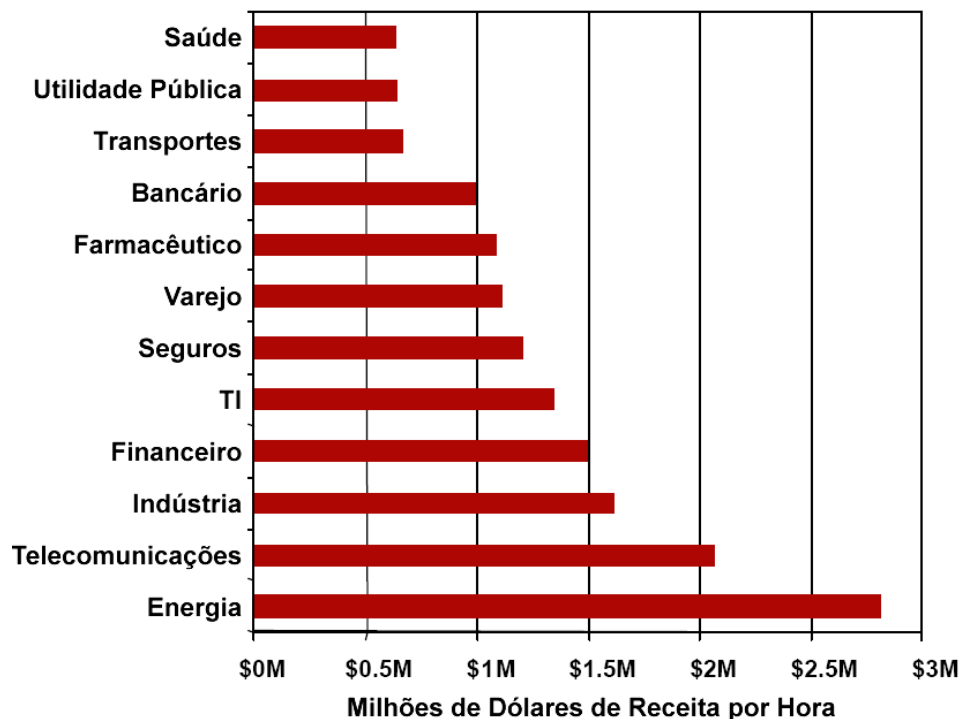
Para as redes continuarem a suprir as aplicações com sucesso não basta adicionar mais capacidade ou conectividade. Novas abordagens na entrega das aplicações, tais como as arquiteturas orientadas a serviço, ponto a ponto, e computação em grade, mudam radicalmente o fluxo do tráfego entre os pontos da rede. Os técnicos



de rede estão constantemente sob pressão, quase sem tempo suficiente para construir o necessário, muito menos entender os fluxos das aplicações e otimizar suas estruturas.

A típica abordagem reativa para superar os desafios das aplicações com a rede é aplicar soluções pontuais, como equilibrar a carga entre servidores, usar dispositivos aceleradores, de detecção e prevenção de intrusão ou soluções de transformação de conteúdo. Isso gera um alto custo de suporte da infra-estrutura, em que o domínio do problema (a aplicação) é transferido para a rede, em que vários dispositivos díspares, de diferentes fornecedores, realizam funções específicas. Exatamente o tipo de erro de design arquitetônico que as empresas deveriam evitar. É necessário um grau mais alto de automação, integração e planejamento arquitetônico fundados sobre a inteligência baseada em rede .

**Figura 1 – Perda potencial de receita por setor industrial**



Fonte: META Group

Freqüentemente, os atrasos relacionados à infra-estrutura estouram os projetos. É dada ênfase aos benefícios das novas aplicações à empresa, mas o impacto na infra-estrutura só é avaliado mais tarde, geralmente tarde demais para se fazer qualquer coisa. Dos projetos de novas aplicações, 65% têm um prazo final predeterminado, com pouco tempo alocado para as melhorias na infra-estrutura. Como resultado, as aplicações são lançadas aos usuários fornecendo um desempenho inadequado ou exigindo uma reforma da arquitetura, o que custa mais do que fazer certo na primeira vez.

Indo além, a etapa fundamental para se desenvolver serviços de infra-estrutura consistentes, confiáveis e reproduzíveis é cada vez mais crítica para as organizações de TI e para a empresa em geral. Portanto, o novo objetivo da OTI moderna é se tornar “adaptativa”. Os esforços adaptativos assumem vários nomes –



de “utilidade”, à “grade” ou “por demanda” –, mas todos possuem a mesma ênfase em alinhar as capacidades de TI, os custos e os objetivos diretamente com os da empresa. Os serviços de rede inteligentes cumprem um papel crucial: Em 2004, uma pesquisa do META Group com 308 tomadores de decisão na lista Global 2000 citaram a infra-estrutura de rede como a principal área de investimento para se tornar adaptativa (veja figura 2).

**Figura 2 – Áreas de investimento para se tornar adaptativa**



Fonte: Estudo de Organização Adaptativa, META Group 2004

#### **Adotando a Mudança: A Agilidade como Pilar do Design do Data Center**

A mudança é uma constante na OTI, e as empresas de serviços líderes a adotam ativamente. Por exemplo, o eBay adiciona até dois milhões de novos itens por dia, registra até 40 mil novos usuários por dia e muda até 30 mil linhas de código por semana – tudo isso sem parar de operar. Chizumaru, um dos maiores provedores de conteúdo do Japão, implementa automaticamente pilhas de softwares em centenas de servidores usando um modelo de atualização sem interrupção (ou *rolling upgrade*). Uma empresa mundial líder de entretenimento tem uma arquitetura orientada a serviço (SOA, *service-oriented architecture*) planejada para disponibilizar uma aplicação on-line, do servidor bruto até estar pronta para o uso, em questão de minutos – um processo que antes levava semanas. Para essas empresas, a importância da agilidade não pode ser superestimada.

#### **Expandindo Infra-Estrutura: Criando a Rede das Redes**

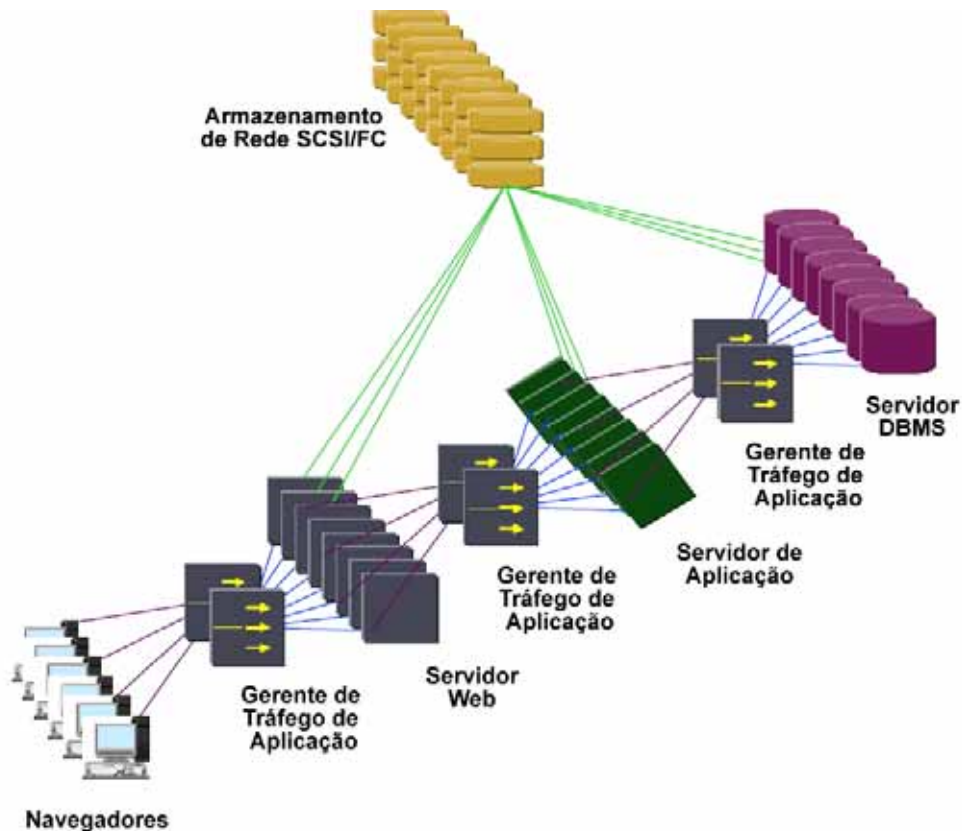
Vinte anos atrás, na era dos mainframes e minicomputadores, os usuários basicamente precisavam comprar seus servidores, terminais, armazenagens e redes de um fornecedor. O fluxo de tráfego era determinístico e facilmente modelável. Porém, a computação distribuída introduziu novas complexidades ao data center. Hoje, os usuários têm a flexibilidade de misturar e selecionar as redes de um fornecedor, os servidores de outro, a armazenagem de um terceiro e, ainda, os clientes de um outro fornecedor. Essa nova abordagem permite um crescimento



dinâmico, de modo escalonado, permitindo a cada nível se adequar à carga de trabalho. Nesse modelo, os componentes de cada domínio formam suas próprias redes de comunicação, mesmo que o sistema precise funcionar como um todo coeso (veja a figura 3).

Essa abordagem escalar aumenta a complexidade e cria a necessidade de uma arquitetura de TI específica à organização, ou uma planta tecnológica. As organizações de TI de sucesso juntam o processo de desenvolvimento arquitetônico ao processo de planejamento corporativo, criando um conjunto comum de descrições e artefatos (como modelos e diagramas arquitetônicos, documentos, normas e controles, padrões técnicos, políticas e modelos de processo). Gerenciadas corretamente, essas arquiteturas em comum resolvem os problemas de aplicações e infra-estruturas e, assim, geram um retorno mensurável do investimento e melhoram a agilidade e o controle do custo, bem como aumentam a integração com as políticas de segurança.

**Figura 3 – A arquitetura técnica é essencial**



Fonte: META Group

Essas descrições arquitetônicas não são apenas documentos abstratos. Uma arquitetura técnica de uma empresa inteira cria uma planta das aplicações de tecnologia baseada nos negócios e voltada para a solução dos problemas corporativos da empresa. As arquiteturas eficazes são sancionadas pelo seu uso contínuo e a arquitetura técnica precisa guiar e direcionar a infra-estrutura e as decisões do planejamento do crescimento. Uma parte fundamental do processo de desenvolvimento da infra-estrutura é a definição de princípios arquitetônicos abrangentes. E, para ser realmente adaptável, uma arquitetura adota a virtualização



– permitindo que a infra-estrutura física seja criada logicamente e gerenciada por um software. Quanto mais componentes uma arquitetura conseguir virtualizar com um conjunto comum de ferramentas e tecnologias, maior será a flexibilidade da organização e a capacidade de minimizar os impactos das mudanças.

### **Princípios do Design de Infra-Estruturas Adaptativas**

As seguintes orientações para descrições e avaliações técnicas dos princípios de design de infra-estruturas adaptativas são apresentadas como guia de avaliação das soluções de gerenciamento do tráfego de aplicações e determinação de como elas se associam a esses princípios específicos.

#### **Princípio da Mudança ou Adaptabilidade**

- **Descrição**  
A única constante é a mudança. Portanto, a infra-estrutura precisa ser planejada de forma que o impacto da mudança seja compreendido, permitido e eficiente.
- **Guia de Avaliação Técnica**  
Procure por soluções que possam virtualizar quase todos os recursos baseados em IP, tanto da perspectiva da WAN quanto da LAN. Um modelo do qual os componentes da arquitetura são abstraídos por uma entidade virtual torna possível a mudança ininterrupta, permite acréscimos, exclusões e atualizações.

#### **Princípio da Reutilização**

- **Descrição**  
A infra-estrutura deve ser planejada para ser reusada o quanto for possível.
- **Guia de Avaliação Técnica**  
Quando considerar as soluções que podem virtualizar os recursos, determine se elas suportam o modelo do fluxo de trabalho e o princípio de reuso baseado em componentes ou objetos. Se, por exemplo, definirmos um plano e uma política escalar comum, poderemos usar e reusar essa política em múltiplas entidades virtuais? O mesmo deve se aplicar a coisas como segurança de aplicação ou políticas de compressão de aplicações. Elas podem ser definidas uma vez e aplicadas a entidades virtualizadas similares, ou precisam ser criadas a cada vez?

#### **Princípio do Padrão Industrial**

- **Descrição**  
Os padrões de modelos, serviços técnicos, componentes e produtos da infra-estrutura devem refletir os padrões da indústria, sempre que aplicável (em especial os padrões de serviços de Internet e Web).
- **Guia de Avaliação Técnica**  
Procure atentamente a aderência estrita aos padrões, incluindo os RFCs (Request for Comments, documentos de sugestões de atualização dos padrões da Internet) e padrões emergentes vindos de entidades reconhecidas como a IEEE, WS-I ou OASIS. Não é incomum que os fornecedores tomem “atalhos” em relação aos padrões para aumentar o desempenho ou entrar no mercado mais rapidamente. Eventualmente, isso pode gerar um caos no seu ambiente. Nenhuma rede é inteiramente “limpa” e as complicações sempre surgem quando encontramos ambientes com pacotes e fragmentos fora das especificações. Sem uma obediência estrita aos padrões, seu investimento pode se perder.



### Princípio do Serviço

- **Descrição**

Os modelos e designs de serviço devem enfatizar o uso de serviços compartilhados – ou seja, especificamente soluções reutilizáveis empacotadas feitas de componentes compartilhados da infra-estrutura acessadas por meio de APIs (Interfaces de Programação de Aplicativos) e protocolos padronizados de um modo livremente integrado (como os serviços Web), com alguma ênfase em solução hospedada completa.

- **Guia de Avaliação Técnica**

O princípio do serviço orienta o conceito de ser capaz de virtualizar qualquer data center, conectividade WAN, aplicações baseadas em IP, servidor, appliance ou dispositivo de rede. Nesse paradigma, o conceito de serviço compartilhado significa aplicar as políticas que melhoram o desempenho do recurso ou da aplicação virtualizados ou retirar a carga de tarefas que fazem sentido (por exemplo, se tiver centenas de aplicações, todas implementadas e mantidas em um modelo de autenticação diferente, procure por soluções que possam centralizar essa função para essas aplicações). Pense na autenticação como um serviço que as aplicações podem subscrever, em vez de ter de implementá-la dentro de cada aplicação em si (os serviços de diretório ainda residem em outro lugar e podem ser tão heterogêneos quanto as aplicações – é a fiscalização da autenticação que está sendo centralizada). A autenticação é, então, abstraída de uma camada e pode ser implementada muito mais rapidamente, mantida com mais simplicidade e auditada para assegurar que o cumprimento das políticas corporativas se torne fisicamente possível. O mesmo princípio deveria ser aplicado em tarefas como o balanceamento de cargas, a criptografia SSL, a otimização TCP e os serviços de compressão, além das políticas de segurança de aplicativos.

### Princípio da Realidade

- **Descrição**

O planejamento pode ser implementado e/ou já é bem suportado.

- **Guia de Avaliação Técnica**

O melhor guia para se determinar isso de um fornecedor é analisar suas parcerias com outros fornecedores, avaliar os guias técnicos que documentam as implementações padrões e verificar até que ponto um fornecedor irá para garantir que o planejamento funciona e foi testado. Os estudos de casos de clientes são úteis, mas raramente cobrem bem a necessidade de avaliação do princípio da realidade.

### Princípio da Omissão

- **Descrição**

Embora um planejamento inclua todos os componentes, eles podem ser omitidos se não forem necessários em uma dada implementação.

- **Guia de Avaliação Técnica**

Isso é fundamental. Olhando novamente o princípio do serviço, se um dado planejamento e alguns de seus componentes não necessitarem de um serviço compartilhado, a solução pode ser configurada para desligar esse serviço ou parte dele? Por exemplo, alguns componentes de uma entidade virtualizada podem não se beneficiar dos serviços de compressão. De fato, o desempenho geral, seja da aplicação ou do cliente, pode ser afetado pela compressão – essa não é uma tecnologia que se ajusta a tudo. Portanto, é importante procurar soluções que tenham a granularidade de omitir certos componentes de um serviço e são



capazes de visar somente aos serviços compartilhados necessários ou úteis.

#### Princípio do Ciclo de Vida

- **Descrição**

Um planejamento deve incluir soluções para cada ciclo relevante da vida útil do fornecimento da aplicação (ou seja, planejamento, testes, disponibilização e produção).

- **Guia de Avaliação Técnica**

Quando se aplica esse princípio como um guia de tecnologia, ele também se relaciona ao princípio da reutilização. Para suportar o princípio do ciclo de vida, a solução deve ser capaz de se organizar em um fluxo de trabalho. Portanto, devem ser procuradas as soluções que permitam às organizações planejar e definir um cenário virtualizado, assim como definir políticas e etapas e testá-las. Esse tipo de solução também deve ser capaz de ser compartilhada entre múltiplas partes como uma plataforma de disponibilização para maximizar o investimento. Além disso, a solução de um fornecedor deve ter uma rede de suporte bastante especializada e abrangente capaz de ajudar no ciclo de planejamento. A assistência deve vir na forma de soluções documentadas, preferencialmente específicas para as aplicações e topologias de redes, guias de configuração, dicas tecnológicas e cenários de integração, assim como códigos, políticas ou perfis de exemplo que foram testados pelo fornecedor. Isso pode ser definido por uma organização como ponto de partida do estágio de planejamento e desenho. Procure por soluções que também tenham um bom sistema de registro de atividade – ou seja, códigos de erros descritivos e significativos baseados em eventos que possam ser depurados durante os estágios de teste e disponibilização.

#### Princípio do Custo Total de Propriedade

- **Descrição**

O planejamento da infra-estrutura deve incluir os modelos de custo total de propriedade (TCO – Total Cost of Ownership), tais como vida útil e remoção.

- **Guia de Avaliação Técnica**

Quando aplicar o princípio de TCO, é importante avaliar a definição formal de “fim de vida” do fornecedor e as políticas associadas tanto para hardware quanto para software, para assegurar que o investimento feito irá passar com segurança, no mínimo, os três anos normais do ciclo de depreciação. O local da solução e a estrutura de gerenciamento distribuída também precisam ser examinados de perto. Isso terá um impacto significativo no TCO. Por exemplo, se a pretensão é reusar, o objetivo torna-se minimizar os erros custosos e a duplicação de esforço. O gerenciamento e a interface de usuário suportam esses objetivos? Além disso, a manutenção constante e a solução dos problemas também devem feitas por medidas granulares, registros descritivos e procuráveis e uma interface integrada para assegurar um diagnóstico rápido com foco no enxugamento dos custos de manutenção. Algumas das iniciativas mais badaladas dos últimos anos, as quais estão todas relacionadas, são o provisionamento dinâmico, a computação dinâmica, grade e a computação por demanda. Apesar do tom pomposo dessas frases, há também uma dose de realidade associadas a elas. Para compreender o tão discutido potencial dessas iniciativas, a solução precisa ter uma API testada e padronizada e um kit de desenvolvimento de software (SDK, *software development kit*) que permita ser integrado à outros componentes da



arquitetura ou ecossistema. Como foi observado, como atualizar automaticamente e com elegâncias os serviços existentes em um ambiente de produção? Isso é feito permitindo que cada componente se remova automaticamente do recurso virtualizado e volte quando a atualização estiver feita. Você não deve ter de tocar fisicamente a solução que está fazendo a virtualização. Essa tarefa, geralmente trabalhosa, deve ser capaz de ser feita de modo programático. Para fazer isso, é preciso que haja uma interface padronizada com a solução que os componentes possam manipular com segurança, seja para automatizar as atualizações, seja para reagir ativamente a uma intrusão pedindo automaticamente ao dispositivo virtualizado para bloquear um endereço IP agressor.

#### Princípio do Financiamento do TCO

- **Descrição**  
O planejamento de custos deve incluir o financiamento para manter a infra-estrutura e as atualizações futuras.
- **Guia de Avaliação Técnica**  
Os preços e modelos de pacotes dos fornecedores precisam ser examinados de perto. O que parecer atrativo à primeira vista pode ocultar alguma surpresa. O fornecedor dispõe de um modelo de parcelamento ou é preciso comprar tudo antecipadamente, mesmo as funcionalidades que talvez não sejam necessárias em momento algum? Quando especificar um planejamento, tente antecipar as funcionalidades necessárias para os próximos seis meses, ou até mesmo 18 meses, assegurando que a autorização orçamentária seja criada segundo o ciclo de vida da solução ou arquitetura.

#### Princípio da Exceção

- **Descrição**  
Existirá um processo de exceção para lidar com os desvios do planejamento.
- **Guia de Avaliação Técnica**  
A única coisa que é constante é a mudança. Observe atentamente quanto uma solução é flexível e adaptativa. Muitas vezes, encontramos clientes amarrados à um design simplesmente porque a arquitetura subjacente de uma solução não pode ser adaptada. Por exemplo, para se adequar ao princípio da exceção, pergunte a si mesmo se o fornecedor da solução pode dar suporte a mais do que um único tipo de aplicação. Concentra-se apenas em aplicações Web? E voz por IP (VoIP) ou protocolo de desktop remoto (RDP) para clientes de pouco tráfego, ou algumas aplicações industriais verticais como protocolo FIX (*financial information exchange*) no setor financeiro? Tenha em mente que o objetivo principal é ser capaz de criar um modelo comum e único que possa ser alavancado por diferentes tipos de aplicação, não apenas um. Para estar condizente ao princípio da exceção, e evoluir de uma organização de TI reativa a uma viabilizadora de serviços, as soluções subjacentes na arquitetura precisam ser adaptativas.

#### Princípio do Gerenciamento da Mudança

- **Descrição**  
O planejamento suportará ao longo do tempo mudanças regulares, coordenadas e em um ritmo razoável .
- **Guia de Avaliação Técnica**  
Aplicar o guia tecnológico a esse princípio destaca a importância de funções como a auditoria. A solução tem uma auditoria robusta das



mudanças? O registro que armazena as informações é não-modificável? No melhor dos casos, uma solução tem uma interface por eventos ou mesmo uma API que pode notificar automaticamente um sistema externo de gerenciamento de mudança quando o status ou estado variar. Os atores principais do princípio de gerenciamento de mudança são a reutilização e os serviços compartilhados. Por exemplo, uma solução virtualizada usada para manter uma sessão persistente com um cliente para aplicações que levam em conta o estado pode ser escrita uma única vez e compartilhada entre várias aplicações virtualizadas? Se sim, as modificações se tornam bem mais simples, pois a mudança é feita uma vez e todos os assinantes do serviço a herdam automaticamente.

#### Princípio do Uso

- **Descrição**  
Os processos devem ser atualizados e usar o design por todo ciclo de vida da solução.
- **Guia de Avaliação Técnica**  
Para que os processos de ciclo de vida funcionem e o processo de atualização seja o mais simples possível, a interface de usuário deve ser suficientemente lógica para suportar o esforço da documentação do processo. Se a interface de usuário e gerenciamento da solução não for lógica, ela irá sobrecarregar essa tarefa. Temos visto o que deveria ser uma operação simples, como a criação de uma entidade virtual e seus nós e monitores associados, variar de três a trinta etapas, dependendo da solução. Certamente, quando mais complexo for o processo, mais custoso e mais sujeito a erros ele será.

#### Princípio da Adequação do Pacote

- **Descrição**  
Os pacotes comprados terão a conformidade avaliada e medida em relação a padrões definidos e serviços técnicos.
- **Guia de Avaliação Técnica**  
Para garantir essa avaliação técnica, o ônus de abastecer com o máximo possível de documentação técnica para melhor avaliar a conformidade é do fornecedor. É importante avaliar se um fornecedor realmente proporciona esse material e desconfiar do desempenho que o fornecedor alega. Frequentemente, o que é publicado nos sites e nos documentos de marketing não é o que se encontra na vida real, nos cenários de produção. Por exemplo, muitas vezes os números de desempenho de SSL e compressão publicados por um fornecedor foram gerados em isolamento, como se a solução só processasse SSL ou só fizesse compressão. Entretanto, no mundo real, sabemos que essas soluções estão realizando múltiplas funções e diferentes tipos de tráfego concomitantemente – não um único evento isolado. Para tirar a prova de tais alegações e ser capaz de avaliar o desempenho, peça os dados de desempenho que mostram como a solução se comporta com múltiplas funções ativas e ocorrendo ao mesmo tempo. Preferencialmente, você deveria receber a informação que se aproxima o máximo possível do que você espera de tráfego de IP.

#### Princípio da Simplicidade

- **Descrição**  
Haverá um número pequeno de planos para satisfazer a maioria das necessidades. Qualquer serviço dado, não importando sua complexidade técnica, deve definir uma interface simples e elegante que permita a reutilização. Interfaces complexas atrasarão a reutilização.



- **Guia de Avaliação Técnica**

Se a simplicidade é uma meta ou objetivo, não se esqueça de avaliar a estrutura de gerenciamento da solução. Pode-se editar, acrescentar ou remover elementos com facilidade? A criação de novos elementos é complicada ou lógica? Você pode dar nomes reais aos objetos definidos em vez de endereços IPs não descritivos?

### Princípio do Design ao Nível do Serviço

- **Descrição**

Um conjunto apropriado e variado de níveis de serviço deveria ser planejado em cada design ou configuração de infra-estrutura.

- **Guia de Avaliação Técnica**

Na avaliação das aplicações de rede que se enquadram na sua SOA, é essencial a capacidade de rastrear e relatar dados específicos relacionados aos dispositivos em si, bem como aos componentes virtualizados. As soluções precisam ter não apenas um robusto suporte ao SNMP e MIBs (bases de informação de gerenciamento) bem definidas para coletar essa informação, mas também uma API aberta que permita coletar, personalizar e mapear os dados para suas necessidades específicas de nível de serviço. Com relação aos componentes virtualizados, deve ser exigido que as soluções criem estatísticas não apenas de alto nível (por exemplo, total do tráfego de saída ou tempo de funcionamento de um serviço virtual), mas também em um nível mais fino, chegando aos componentes individuais do serviço virtual. Por exemplo, poderiam ser nós específicos ou conjuntos de nós. Esses dados geralmente não são coletados por SNMP, mas precisam ser fáceis de coletar – portanto, a necessidade de uma API aberta. Se a API for direcionada a eventos, quando ações especificamente definidas ocorrerem, uma ação automática pode ser executada. Por exemplo, quando a capacidade de um serviço ultrapassar um limite (digamos 5 Mbps), os dados podem ser automaticamente publicados em um repositório.

### **Reduzindo a Complexidade: O Objetivo-Chave**

Um objetivo fundamental tanto do planejamento arquitetônico e da infra-estrutura quanto da engenharia dos modelos de tecnologia é a redução da complexidade. Reduzir a complexidade traz vários outros benefícios, incluindo a diminuição dos custos e o aumento da velocidade com que as mudanças são feitas. Quando planejar arquiteturas mais distribuídas e cliente/servidor (e agora em n-camadas e orientada a serviço), muitos dos componentes díspares (produtos individuais e categorias de serviços) precisam ser organizados adequadamente. No catálogo de Infra-Estrutura Adaptativa do META Group, existem literalmente dezenas de componentes que abrangem uma solução completa.

Entretanto, ambas as abordagens do planejamento arquitetônico e de infra-estrutura reconhecem que uma grande lista de componentes indiferenciados não está organizada. Conseqüentemente, a maioria das empresas sugere o uso de modelos de agregação de mais alto nível para organizar melhor essa confusão de componentes. Esses componentes são organizados por domínios, como ilustra a figura 4.



Figura 4 – Domínios de tecnologia em comum

API	Plataforma (HW/OS servidor ou cliente)
Apresentação	Armazenagem
Servidor de aplicação	Rede
Servidor de integração	Segurança
Servidor de dados	Gerenciamento

Fonte: META Group

**As Implicações de uma Arquitetura Orientada a Serviços nos Data Centers**

Os modelos de SOA, incluindo os serviços Web, apresentam uma mudança significativa no modo que as organizações planejam os sistemas corporativos e as aplicações que os suportam. Essa mudança afetará não apenas as aplicações, mas também os papéis e responsabilidades na empresa e as várias ferramentas e técnicas usadas para criá-las. A SOA promete serviços altamente adaptáveis, os quais são conectados por um modelo fácil de alterar que espelha o processo da empresa e que pode ser relacionado a ele de uma maneira que um usuário corporativo entenda o sentido. A SOA enfatiza o design de textura granulada e interfaces padronizadas que ajudam permitir a criação de aplicações compostas.

Outra clara tendência arquitetônica é a geração de sistemas compostos de um número crescente de subsistemas de idêntica funcionalidade e de menor tamanho (conhecida como fusão de recursos). O objetivo do design é construir um sistema virtual altamente confiável e expansível a partir de um grande número de subsistemas moderadamente confiáveis e expansíveis. Em termos gerais, virtualizar um conjunto de diversos recursos concretos e acessá-los por meio de uma única interface uniforme que, do ponto de vista do usuário, permita que eles funcionem como um único recurso unificado que pode ser compartilhado em diferentes graus de comportamento dinâmico. Daí o papel estratégico das soluções de gerenciamento de tráfego e porque elas servem de base para essas arquiteturas.



A virtualização cria a ilusão de um sistema único gerado por duas forças complementares:

- **Miniaturização:** O tamanho e complexidade do sistema, relativos a seu desempenho, diminuem com o tempo – esteja o sistema em um chip, disco ou porta, no nível da placa ou até mesmo em um servidor
- **“Massificação”:** Essa é a combinação da multiplexação com o multiprocessamento – multiplexação de um crescente número de usuários (sejam eles pessoas ou sistemas) em um crescente número de processadores uniformes (sejam transistores, CPUs, unidades de disco, ou servidores de aplicação/base de dados). O número de subsistemas que compõem as instâncias do sistema cresce com o tempo em centenas de processadores, servidores e discos.

A complexidade exponencial dos data centers futuros não é impulsionada apenas pelo grau de integração entre todos os sistemas, mas também pela aceleração da taxa de mudança destes sistemas integrados e de suas interconexões - de mudanças na topologia da rede a alterações nas configurações de software e na integração de aplicações. Os data centers futuros serão repletos de elementos de armazenagens, servidores e redes de diversos fornecedores e definidos por interfaces cada vez mais padronizadas e integradas. Isso irá beneficiar o comprador de TI, em função da maior competição entre os fornecedores de software e hardware.

#### ***O Desafio da Segurança: “Porta 80 Sobrecarregada”***

A nova arquitetura de aplicação que usa XML e SOAP geralmente utiliza a porta 80 de TCP (porta HTTP) para a transferência de dados - normalmente a única porta aberta e confiável pelo firewall corporativo. Embora seja um modo confiável de assegurar a comunicação sem ser bloqueado, infelizmente também tem o efeito colateral de eliminar o propósito básico do firewall – proteger as aplicações. Conforme mais tráfego passa pela porta 80, torna-se mais essencial que os dispositivos examinem minuciosamente os envelopes SOAP para inspecionar as aplicações.

O próximo passo lógico, para ser realmente eficaz no longo prazo, é tornar os dispositivos intermediários, gerenciáveis, mais cientes do comportamento das aplicações. Portanto, compreender o que é um comportamento correto e positivo cria um modelo para a segurança das aplicações. Uma solução deveria permitir apenas o comportamento reconhecidamente correto. É muito mais fácil manter um sistema que sabe que uma aplicação tem permissão somente para 20 entradas de determinado tamanho e valor e deve impedir qualquer outro tipo de entrada, constantemente verificando, portanto, centenas de assinaturas negativas e possíveis ameaças.

Dessa forma, esses novos dispositivos não devem operar simplesmente nas camadas de rede, transporte e sessões como os firewall comuns, mas também devem ter ciência das aplicações XML e SOAP. Esses dispositivos são freqüentemente conhecidos no mercado como firewall, gateways ou aceleradores de aplicações XML ou Web.

O benefício de uma maior consciência das aplicações e protocolos é similar a da virtualização, pois abstrai e alivia as funções de segurança e gerenciamento da aplicação em si e centraliza a fiscalização. Para os desenvolvedores, isso é um grande alívio, pois eles ganham certa ajuda em garantir a proteção de suas



aplicações, e, para os auditores da segurança, torna-se prático uma auditoria completa. Também é muito mais fácil para os operadores analisarem as políticas de alguns dispositivos que suportam centenas de aplicações do que tentar auditar diretamente todas elas.

***Suportando Variações, Mas Gerenciando a Complexidade: O Novo Papel do “Balanceamento de Cargas” como Gerente de Tráfego de Aplicação na Alocação Dinâmica de Recursos***

*O gerente de tráfego de aplicação... pode tomar decisões inteligentes baseadas em várias informações de cada camada da arquitetura.*

A abordagem por expansão provou ser imensamente escalonável, como demonstrado em situações reais com serviços de Internet como o eBay e Expedia. Essa abordagem possui vantagens importantes na seleção de produtos e na alocação de capital para as necessidades correntes, sem precisar de altos investimentos fixos. Ainda assim, as organizações lutam para juntar o conjunto de todos os domínios e a miríade de componentes em um único sistema unificado. Os servidores de dados e de aplicações tendem a gerenciar somente a si mesmos e são cegos em relação à rede. Os dispositivos de rede simples não têm conhecimento das aplicações. Daí a necessidade de um dispositivo de rede “fluente” nas aplicações – gerente do tráfego das aplicações – que possa tomar decisões inteligentes baseadas em várias informações de cada camada da arquitetura – dos servidores de aplicações e de dados, do hardware do servidor e do armazenamento, da capacidade da rede e, até mesmo, no nível do usuário individual.

Como nomenclatura, “balanceamento de carga” descreve apenas uma pequena parte do valor que os dispositivos de rede cientes das aplicações trazem ao data center. A inteligência destes produtos na Camada 7 (ou seja, camada da aplicação nas camadas OSI) torna os dispositivos de rede flexíveis, permitindo equilibrar a carga entre vários recursos de rede (por exemplo, firewalls, links de ISP, servidores de e-mail e VoIP), seja dentro de um data center, seja entre vários data centers, enquanto atua em muitos outros papéis, além de simplesmente balancear a carga. De fato, a versatilidade desses dispositivos cria de certa forma um dilema para os fornecedores e arquitetos quando eles precisam descrever especificamente para o que eles servem. Embora haja um farto número de exemplos de uso típico, muitas organizações estão aplicando esses dispositivos de modos inusitados – tanto em aplicações web quando tradicionais e, até mesmo, em aplicações de tempo real como VoIP.

Conforme a funcionalidade central desses dispositivos amadurece, os fornecedores e as organizações de TI ponderam o rumo futuro dessas redes-para-toda-obra. Basicamente, esses dispositivos assumiram o papel de melhorar o desempenho, a eficiência e a segurança por meio do uso de múltiplas técnicas aplicadas inteligentemente no fluxo da aplicação.

O papel do gerente de tráfego da aplicação é:

- ***Alívio do processamento:*** Desafogando outros dispositivos do processamento SSL, do gerenciamento da conexão, do cache de conteúdo, do spool de aplicações e do gerenciamento dos cookies de sessões.
- ***Assegurar o acesso equitativo aos recursos:*** Usando as políticas de qualidade de serviço.
- ***Melhorar a confiabilidade:*** Permitindo que os dispositivos atrás do gerente de tráfego de aplicação sejam postos e tirados de operação sem interrupção do



serviço.

- **Melhorar o desempenho:** Por meio de várias técnicas como compressão, otimização de TCP, cache, e spool de aplicação.
- **Proteger os recursos no data center:** Usando um mecanismo de política, tendo consciência dos usuários e aplicações e se tornando um ponto de fiscalização da autenticação.
- **Gerenciamento centralizado dessas funções:** Para melhorar a confiabilidade geral dos serviços e conter custos. De fato, os recursos de gerenciamento são mais importantes do que o custo de equipamento. O arquiteto de TI de uma empresa líder de serviços on-line estima que mais de 90% do CTP dos dispositivos da rede deve-se a custos operacionais, principalmente o custo de pessoas e de serviços de suporte secundários.

#### Estudo de Caso: Redução da carga de SSL

Uma empresa multinacional criou um portal Web para atender as necessidades de mais de 90.000 funcionários. Ela precisa manter a privacidade e a segurança de todas as atividades on-line dos funcionários, portanto usou SSL em todas as suas sessões. Ao passar o SSL para um appliance, a empresa reduziu a carga total de processamento do servidor em 15%, diminuindo drasticamente sua necessidade por hardware, software e pessoas no suporte da aplicação.

Na operação real, muitos dos clientes relatam sempre descobertas de novos usos para esses appliances. Uma empresa os usará como dispositivos de balanceamento de carga para manter os cookies de segurança emitidos pela aplicação Netegrity SiteMinder. Outra está usando os dispositivos para “corrigir” aplicações mal arquitetadas. Essa empresa usa uma aplicação crítica que tem um esquema de licenciamento incomum (e, infelizmente, mal planejado): Após a autenticação de um usuário válido, a aplicação deveria enviar de volta ao cliente um endereço de IP estático do host para a sessão subsequente. Isso resultaria em uma aplicação altamente frágil – esse diálogo cliente-servidor único precisaria ser mantido ao longo de toda a sessão do usuário. Porém, ao permitir que o gerente de tráfego da aplicação atue como um proxy, a sessão pode ser mantida entre um pool de servidores, aumentando em muito a confiabilidade.

#### Estudo de Caso: Redução do Custo de Suporte ao Simplificar o Gerenciamento

Para uma grande empresa de serviços de transporte, a proliferação das aplicações Web elevou drasticamente os custos de suporte. Ela escolheu reduzir a complexidade da infra-estrutura introduzindo uma solução de gerenciamento de tráfego nos seus data centers. Após considerar outras soluções de balanceamento de carga “simples” com interface por linha de comando, ela optou por um fornecedor com uma interface gráfica completa. A empresa estima que sua escolha reduziu a força de trabalho de suporte necessária entre 1,5 a 2 vezes.



### Redução das Conexões TCP

Quando um navegador se conecta a um site, ele pode abrir várias conexões com ele, aumentando a carga do servidor *back-end* e a necessidade de largura de banda, além de introduzir uma latência excessiva. Ao transferir essa capacidade para um dispositivo de rede inteligente que use multiplexação TCP, a carga do servidor se reduz e melhora o desempenho para o usuário final. Essa capacidade fica ainda melhor quando se usa HTTP v1.1 Keep-Alives, permitindo que o dispositivo mantenha uma única conexão com o usuário. Isso permite mais do que duplicar o número de conexões que um grupo de servidores web pode suportar.

#### **Estudo de Caso: Criação da Capacidade de Sobrecarga e Ajuste do Desempenho**

Freqüentemente, os websites estão sujeitos a picos de uso, como os causados por multidões de flashes. Usando as capacidades de monitoramento do sistema das soluções de gerenciamento de tráfego de aplicação, uma empresa de serviços on-line envia automaticamente o tráfego para um conjunto de servidores auxiliares quando o conjunto de servidores ativos atinge certo limite de uso. A solução monitora os servidores Windows emitindo periodicamente chamadas à Interface de Gerenciamento do Windows requerendo informações sobre o uso da CPU de cada servidor. Ao monitorar as requisições no nível da base de dados, a empresa também pode separar as leituras das gravações de dados, redirecionando os clientes para um servidor especificamente ajustado para essa função.

#### **Estudo de Caso: Automação das Atualizações e Lançamentos de Aplicações**

Um grande fornecedor de conteúdo, publicando informações e aplicações sem fio avançadas, enfrentou o desafio de atualizar diariamente centenas de servidores de aplicação. Financeiramente, não haveria economia de escala nem seria confiável usar técnicos para gerenciar o meticuloso processo de atualizar individualmente cada servidor. Integrando os processos de gerenciamento e implementação de código de aplicação com sua rede por meio da API SOAP/XML – uma interface de serviços web padrão para manipular remotamente a configuração dos dispositivos de rede – a empresa pôde agilizar o controle e a atualização dos servidores web gerenciados, tornando as atualizações de aplicação completamente automáticas. As atividades que antes requeriam duas ou três pessoas por dia, hoje são totalmente automáticas e livres de erros. O dispositivo de rede é acessado por meio de uma API e agora oferece um ponto de controle poderoso no suporte dos esforços ininterruptos de implementação.

Todos esses casos de exemplo ilustram a criação de uma arquitetura orientada a serviços. Essas empresas estão usando a mesma solução, mas aplicando-a de diferentes maneiras, segundo o tipo de aplicação e os princípios de planejamento da infra-estrutura de aplicação usados. Nisto se apóia a força dessas soluções e porque nós as consideramos fundamentais; elas são representativas do que significa ser adaptativo.



## **Recursos dos Dispositivos de Gerenciamento de Tráfego de Aplicações**

### **Recursos Essenciais**

Os dispositivos de gerenciamento de tráfego de aplicações evoluíram bem além do simples balanceamento de carga. Esses dispositivos:

- Aceleram o desempenho da aplicação e minimizam os custos de infra-estrutura
- Usam vários dados em tempo real para as decisões de gerenciamento de tráfego, incluindo medidas do hardware do servidor, do sistema operacional, da base de dados, do servidor de aplicação (por exemplo, a utilização da CPU)
- Compreendem implicitamente as aplicações de TI mais comuns (isto é, não é necessária uma indicação explícita da aplicação)
- Podem inspecionar as cargas úteis como os envelopes SOAP e avaliar as mensagens XML, realizando uma profunda inspeção e reconhecendo as assinaturas e atividades de ataque
- Fornecem um firewall de filtragem de pacotes e isolamento das negações de serviço
- Levam em conta as particularidades do usuário nas decisões de gerenciamento de tráfego, tais como o tipo ou velocidade da conexão
- São completamente transparentes ao cliente e às aplicações (isto é, não requerem alteração no código das aplicações ou software do cliente)
- São reutilizáveis por diferentes usuários como filiais, acesso remoto de funcionários, parceiros comerciais e clientes
- São reutilizáveis entre múltiplas aplicações e tipos de aplicações
- São acessíveis remotamente como um serviço de aplicação por meio de API padronizada para formas avançadas de gerenciamento e manipulação externa, adaptando-se a processo de TI em contínua evolução e a mudanças automatizadas e em tempo real na configuração da rede (a API deve ser bem documentada e ter um kit de desenvolvimento de software para integração com a SOA)
- Aceleram tanto o tráfego de texto limpo quando o tráfego criptografado
- Não transferem os gargalos de um elemento a outro
- Resolvem mais do que um limite específico de desempenho
- Podem ser implementados de várias maneiras e por diversos fatores, de uma configuração pequena e fixa de appliances a grandes sistemas modulares

### **Atributos de Controlabilidade**

Os dispositivos de gerenciamento de tráfego de aplicações têm os seguintes atributos de controlabilidade:

- São fáceis de instalar, configurar e testar
- Fornecem relatórios que contribuem para o isolamento e resolução de problemas
- Em operação, são tolerantes a falhas e, quando inativos, fails-to-wire, ou seja, a aplicação continua a funcionar se o dispositivo acelerador for removido ou ficar off-line (opcionalmente, podem ser configurados para não ter nenhum ponto de falha)
- Podem fornecer uma grande disponibilidade e escalonamento (por exemplo, opcionalmente podem ser usados em clusters ou em outras formas tolerantes a falhas)
- Fornecem visibilidade operacional por meio do monitoramento interno da saúde do software e da integração com ferramentas de monitoração de rede



- Suportam opções flexíveis de gerenciamento, incluindo gerenciamento distribuído, hierárquico e remoto (isto é, uma interface web segura)

No geral, conforme amadurece a funcionalidade da aplicação, torna-se possível fornecer uma infra-estrutura de serviço hospedada em switches inteligentes, a qual pode ser reusada por várias aplicações. O objetivo dos novos dispositivos de rede conscientes das aplicações é então fornecer um conjunto em comum de serviços de infra-estrutura que vá bem além do serviço básico de balanceamento de carga – entrando em SSL, criptografia em massa e outras funções - com a capacidade de distribuir papéis de gerenciamento de políticas entre grupos de aplicações por meio do uso de simples APIs. Entretanto, antes que uma organização comprometa seus recursos de desenvolvimento em um conjunto específico de APIs, ela precisa certificar-se que o fornecedor da API seja um concorrente conceituado e de longo prazo no mercado de dispositivos inteligentes de rede.

### Resultado Final

- Os dispositivos inteligentes de rede cumprem um papel central no fornecimento de aplicações dos novos data centers, unificando a infra-estrutura díspar em um sistema coeso e direcionando o tráfego do usuário para o recurso mais apropriado.
- A funcionalidade desses dispositivos está indo bem além do balanceamento de carga, evoluindo para uma futura plataforma de serviços comuns de infra-estrutura, como criptografia, compressão, cache e segurança de aplicação.
- Em função da variedade de aplicações e da frequência das mudanças no data center, os usuários devem examinar dispositivos com propósitos gerais e um grande número de funcionalidades e fatores de dimensionamento do que appliances com propósitos específicos.
- Ao longo do ciclo de vida do sistema, o custo operacional dos dispositivos inteligentes de rede é muito maior do que o capital inicial desembolsado. Os compradores devem examinar os recursos de gerenciamento que reduzem a complexidade e facilitam as operações.
- Se a comutação for um elemento estratégico na infra-estrutura do data center, os usuários devem então envolver desenvolvedores para aproveitar os recursos de gerenciamento de políticas (incluindo APIs) disponíveis na plataforma.

*David Willis é vice-presidente e chefe de pesquisa de Estratégias de Infra-estrutura, um serviço de consultoria do META Group. Para informações adicionais sobre esse tópico e sobre outras ofertas do META Group, entre em contato pelo e-mail [info@metagroup.com](mailto:info@metagroup.com).*

### Sobre o META Group

#### *Return On Intelligence SM*

O META Group é um provedor líder em pesquisa em tecnologia da informação, serviços de consultoria e consultoria estratégica. Fornecendo orientações objetivas e litigáveis, os experientes analistas e consultores do META Group têm a confiança de executivos de TI e corporativos ao redor o mundo. Nossos exclusivos modelos de colaboração e serviços dedicados ao cliente ajudam os clientes a usar a TI com mais eficiência, eficácia e oportunidade para atingir seus objetivos corporativos. Visite o site [metagroup.com](http://metagroup.com) para mais detalhes sobre nossa valiosa abordagem.