Otimize o 5G com containers em bare-metal

Por que provedores de serviços de comunicação estão escolhendo essa arquitetura?

Tom Conklin, Setor de vendas de edge, América do Norte, Red Hat

Sanjay Aiyagari, Arquiteto de soluções principal, Red Hat

"A avanço dos serviços mais recentes e inovadores para containers em bare-metal possibilita uma melhor competição no mercado. O caminho da evolução vai das máquinas virtuais para os containers. É possível encaixar mais aplicações em uma infraestrutura menor e com isso reduzir custos, implantar serviços em segundos e manter o nível de segurança que os clientes exigem."

Anoop Agrawal
Diretor de engenharia de soluções para MEC e SDN, Verizon

Para conseguir os mesmos benefícios em escala dos provedores de nuvem globais, os CSPs precisam mudar sua abordagem. A paridade em desempenho, eficiência e agilidade, só será atingida por meio da evolução de máquinas virtuais para containers.

Resumo executivo

A segunda onda de migração das redes de telecomunicações para a nuvem está em curso. Na primeira, a virtualização de funções de rede (NFV) reduziu custos ao migrar de elementos de rede baseados em dispositivos para equivalentes baseados em software. Ao mesmo tempo, ofereceu suporte às necessidades legadas dos gerentes de elementos. Agora, os provedores de serviços de comunicação (CSPs) estão adotando arquiteturas e containers nativos em nuvem para aumentar eficiência, desempenho, resiliência, segurança e agilidade. A arquitetura escolhida é a de containers em bare-metal sem uma camada adicional de virtualização. Essa opção apresenta vantagens significativas para seus casos de uso específicos.

Este whitepaper destaca os benefícios da implantação de serviços 5G usando containers em servidores bare-metal em relação aos containers em máquinas virtuais.

Histórico

Quando as funções de rede começaram a migrar para a nuvem, visionários de grandes CSPs buscaram a migração direta para containers, eliminando a implantação das funções de rede em máquinas virtuais. A Red Hat foi uma das primeiras a propor essa ideia, tendo enxergado o valor do Kubernetes ao contribuir ativamente com suas versões iniciais. No entanto, naquela época, a dinâmica do setor e algumas limitações dos containers para redes de telecomunicação acabaram definindo o OpenStack® e as máquinas virtuais como a base para essa mudança. A maioria dos CSPs implantou com sucesso as funções de rede virtuais (VNFs) segundo a arquitetura de NFV do Instituto Europeu de Normas de Telecomunicações (ETSI) para oferecer serviços atuais. A curto prazo, máquinas virtuais executando aplicações legadas continuarão a existir na rede e no espaço de TI dos CSPs. Porém, a maior parte das novas aplicações e componentes 5G será implantada com o uso de containers.

Grandes mudanças tecnológicas estão por vir. Aplicações legadas estão sendo desconstruídas da antiga forma monolítica – que é mais adequada às máquinas virtuais – e se transformando em microsserviços. Atualmente, as práticas Ágil e DevOps são as mais recomendadas para desenvolver e gerenciar aplicações. Elas funcionam melhor quando criadas com o uso de componentes menores e independentes. Com essa abordagem, é possível desenvolver e implantar funcionalidades e correções para cada componente de maneira interativa, sem afetar outros elementos da aplicação. Os CSPs e seus fornecedores costumam exigir integração e entrega contínuas (CI/CD) para que possam oferecer suporte a diversos upgrades e patches todos os anos. O desenvolvimento Ágil permite que eles ofereçam um suporte muito melhor.

A lista de requisitos que resultava na adoção de máquinas virtuais diminuiu com o tempo. Por exemplo, um dos benefícios de usar as máquinas virtuais era o suporte a diversos sistemas operacionais. Hoje, esse nível de complexidade é desnecessário. O Linux® é o sistema operacional padrão para aplicações de rede. Uma nuvem com suporte a vários sistemas operacionais também aumenta a complexidade na hora de corrigir vulnerabilidades de segurança quando estas são descobertas. Além disso, agora, as funções e aplicações de rede são tipicamente stateless, e não stateful. Por isso, não dependem da afinidade com o armazenamento de rede como antes.

Redução de até 44% no custo total de propriedade (TCO)[1]

Pesquisa feita pela Red Hat e pela ACG mostra que o TCO de uma implantação 5G pode diminuir em até 44% com a implantação de uma vRAN aberta em vez de tecnologias RAN com finalidade específica. Essa redução de custo é um incetivo para os CSPs adotarem o 5G na distribuição.

Uma das mudanças mais relevantes no mundo das telecomunicações é a evolução para o 5G, tanto na rede de acesso por rádio (RAN) quanto no núcleo. [2] O 5G oferece a um CSP a capacidade de distribuir funções de rede para locais físicos a partir do lugar que fizer mais sentido para o cliente e para os negócios. Agora é possível distribuir funções de núcleo e voz em nuvens públicas, datacenters centralizados e regionais e até mesmo no local do cliente. A RAN também está evoluindo com a colaboração e padronização na OpenRAN (O-RAN) Alliance e no projeto OpenRAN para oferecer suporte a funcionalidades hospedadas em servidores prontos para uso em vez de em hardware proprietários. Os componentes podem ser distribuídos geograficamente, oferecidos por diversos fornecedores e usar uma infraestrutura compartilhada entre os elementos do núcleo de telecomunicação e as aplicações de valor adicionado para usuários finais.

Os containers são uma adequação natural a essas melhorias, e implantá-los em bare-metal traz grandes benefícios para os CSPs. Para aplicações legadas, pode haver motivos para implantar containers em máquinas virtuais. Mas, para o 5G, a implantação de containers com essa segunda camada de complexidade é desnecessária e contraproducente.

Eficiência operacional

Um dos principais benefícios da virtualização de uma rede de telecomunicações é o aumento na eficiência. Tradicionalmente, cada serviço oferecido aos clientes tinha seu segmento de tecnologia. Sua manutenção era executada por grupos independentes, que gerenciavam camadas tecnológicas, operações de rede e suporte ao cliente. Ao estabelecer funções de rede diferentes que realizam os serviços na mesma infraestrutura de nuvem, essas várias camadas organizacionais são eliminadas.

A curto prazo, as máquinas virtuais continuarão em uso na rede para as funções existentes. Porém, será difícil ignorar as vantagens significativas dos containers na hora de implantar novas funções – incluindo RAN e núcleo 5G.

A evolução das máquinas virtuais para containers não deve ser um retrocesso em termos de eficiência. O OpenStack e o Kubernetes são especialidades diferentes. Colocar containers em máquinas virtuais agrega uma camada de recursos às arquiteturas de NFV existentes em vez de apenas substituir uma delas. Expandir os recursos da rede para incluir elementos da nuvem pública e milhares de locais remotos da edge aumenta a complexidade consideravelmente. O Kubernetes para containers vem sendo usado para lidar com essa complexidade. A manutenção de máquinas virtuais nesse ambiente gera complexidade desnecessária e ineficiência operacional. Containers implantados em bare-metal têm o mesmo suporte em todas as variantes da nuvem, o que não acontece com as máquinas virtuais e suas diversas plataformas de suporte. Assim, o suporte simultâneo a containers e plataformas de máquina virtual exige muito mais pessoas, ainda mais se incluirmos as nuvens públicas e privadas. O ideal seria usar uma única plataforma de orquestração de containers para os dois tipos de nuvem.

A migração para funções de rede nativas em nuvem (CNFs) difusas pode dificultar o gerenciamento e a segurança do ciclo de vida. Os microsserviços que constituem uma aplicação nem sempre estão integrados fisicamente entre si, seja no mesmo servidor ou no mesmo local. Sem uma automação significativa, o gerenciamento do ciclo de vida se torna trabalhoso. CSPs e fornecedores usam cadeias de ferramentas DevOps, metodologia CI/CD e, agora, recursos de garantia de loop fechado para aumentar a eficiência. Porém, a automação exige habilidades específicas caras e escassas. Sempre que possível, é importante que o CSP exija um pipeline de CI/CD comum a todos os fornecedores.

É necessário criar, testar e implantar atualizações e patches de maneira automática e uniforme. Qualquer complexidade adicional na rede exige mais recursos para desenvolver e atualizar funcionalidades de automação – além da manutenção manual das funções que não podem ser automatizadas. Essa complexidade adicional prejudica os esforços dos CSPs para reduzir custos, manter a competitividade e atender às demandas do cliente.

[1] ACG Research. "Economic advantages of virtualizing the RAN in mobile operators' infrastructures," patrocinada pela Red Hat. Setembro de 2019.

[2] 5G Americas. "5G and the cloud," dezembro de 2019.

Agilidade do serviço

Com a NFV, os CSPs ganharam agilidade e podem implantar e desativar serviços muito mais rápido.[3] Além disso, com o 5G, eles pretendem abrir suas redes para permitir que terceiros solicitem e provisionem fatias dela para aplicações de atacado especializadas. Esse tipo de serviço será disponibilizado por meio de interfaces de programação de aplicações (APIs) northbound, que concedem acesso controlado para que terceiros usem os recursos de rede. A velocidade de provisionamento é crucial para a satisfação do cliente. Máquinas virtuais são muito mais lentas do que containers. Com elas, o cliente às vezes espera horas para que uma fatia solicitada com dois cliques seja instanciada na forma de várias máquinas virtuais em diferentes locais de rede. No entanto, essa experiência se torna quase instantânea com a implantação de um serviço baseado em container.

Pensando no futuro, a velocidade dos containers em bare-metal também possibilita serviços diferenciados na edge. Um usuário final pode solicitar um serviço de valor adicionado e tê-lo implantado sob demanda em um nó da edge próximo a ele. Nem mesmo as cargas de trabalho ficam permanentemente nos nós da edge; eles são liberados para hospedar as aplicações com maior demanda. Por exemplo, um cliente mobile usando uma aplicação com largura de banda alta e comunicações ultra confiáveis e de baixa latência (URLLC) ativada por 5G poderia fazer a aplicação segui-lo à medida que se aproxima ou se afasta de locais edge diferentes. As aplicações seriam instanciadas em segundos quando se aproximasse do novo local da edge. A implantação de containers em máquinas virtuais não possibilita esse benefício, já que os recursos de todas as aplicações possíveis teriam que ser reservados na edge e as máquinas precisariam ser instanciadas sob demanda, o que demoraria muito.

A agilidade do serviço com containers ajuda a acelerar a geração de receita e a liberação de recursos por serviços desativados. Ela também permite que os CSPs assumam os riscos que os tornarão mais competitivos.[4] Esses benefícios ficam ainda mais evidentes com a implantação dos containers em bare-metal.

Desempenho

Tanto nas redes de telecomunicação como na natureza, uma transformação só se torna uma evolução quando gera vantagens. Ela deve resolver problemas, tornar os negócios mais competitivos e melhorar as chances de sobrevivência. Por isso, mudanças que resultam em aumento nos custos e redução no desempenho não costumam ser adotadas. Os requisitos de desempenho específicos das redes de telecomunicação tornam a implantação de containers em máquinas virtuais um retrocesso.

Com a NFV, um dos maiores obstáculos para aplicações com núcleo mobile e subsistema multimídia de protocolo de internet (IMS) era a taxa de interrupção criada pelas máquinas virtuais. A princípio, esse efeito resultou em uma degradação de 20% do desempenho normal do servidor.[5] Para os CSPs que precisavam de um desempenho melhor, esse resultado poderia desqualificar a virtualização de certas aplicações. O foco na arquitetura e os avanços na aceleração de hardware, como a virtualização de entrada/saída de raiz única (SR-IOV) e o Data Plane Development Kit (DPDK), melhoraram bastante o desempenho. No entanto, com a combinação de containers e máquinas virtuais, a taxa de interrupção voltou a ser um problema. Containers em máquinas virtuais já não causam uma latência de processamento significativa, mas, se comparados a containers em bare-metal, a sobrecarga agregada das máquinas reduz o desempenho de entrada/saída (E/S) de um nó. Essa redução é mais expressiva nas transações de E/S pequenas do que com nas grandes. O tráfego de voz e de dados em uma rede de telecomunicações costuma ser muito fragmentado e composto de pequenas transações.

[3] Estudo de caso da Red Hat. "Turkcell cria nuvem de telecomunicações unificada com a NFV baseada no OpenStack da Red Hat," abril de 2020.

[4] Esfandiari, Shirin. "Bringing 5G-enabled services to life calls for cloud-native operations." DevOps.com, 20 de agosto de 2019.

[5] Liu, Ming, et al. "Understanding the virtualization "Tax" of scale-out pass-through GPUs in GaaS clouds: An empirical study." IEEE.org, fevereiro de 2015.

Hoje, as funções de rede que exigem o melhor desempenho são aquelas que fazem uso intensivo de operações E/S em vez do uso da unidade central de processamento (CPU). Implantá-las em máquinas virtuais por meio de containers pode prejudicar consideravelmente o desempenho. O impacto pode chegar a 30%, exigindo que mais containers desempenhem a mesma tarefa.[6] Mais containers em máquinas virtuais significa mais servidores – e mais custos.

Benefícios da infraestrutura

Antes da NFV, a utilização do hardware de rede costumava ser muito baixa em cargas normais. Mesmo assim, oferecia-se redundância – e isso gerava custos desnecessários. Além disso, era preciso armazenar peças de reposição por perto, o que aumentava ainda mais as despesas. A capacidade era modelada para períodos de pico e os gastos diários não variavam, não importando se o dia tinha sido muito lucrativo ou se teve pouca demanda e menos lucro.

Em teoria, a virtualização de várias VNFs na mesma infraestrutura melhoraria bastante a utilização. A aplicação da orquestração avançada no gerenciamento do ciclo de vida eliminaria a necessidade de hardware redundantes e peças de reposição. A nuvem ofereceria uma capacidade comum para todas as VNFs se expandirem ou falharem. Porém, na prática, as VFNs eram frequentemente implantadas em segmentos de hardware diferentes e específicos para cada fornecedor. Essa separação era necessária porque os CSPs exigiam que os fornecedores tivessem o mesmo desempenho e uptime para os produtos que os hardware legados. Sem agrupar recursos da nuvem para oferecer a disponibilidade de 99,999% exigida pelos CSPs, os segmentos continuavam sendo modelados para capacidade máxima. Além disso, costumava-se implantar segmentos de hardware redundantes para as aplicações que exigiam alta disponibilidade. Em resumo, a promessa de uma utilização aprimorada, muitas vezes, não era cumprida. Isso resultava em vários gerentes de infraestrutura virtual (VIMs) de diferentes fornecedores administrando máquinas virtuais em segmentos tecnológicos distintos na mesma rede.

Por causa dos métodos de gerenciamento do ciclo de vida e do desenvolvimento de software atuais, os componentes RAN e de núcleo 5G são, em geral, projetados como funções nativas em nuvem executadas em containers. A arquitetura especificada pela 3rd Generation Partnership Project (3GPP) permite desmembrar funções anteriormente monolíticas em microsserviço oferecidos por diferentes fornecedores e projetos open source. Os CSPs esperam que os fornecedores compartilhem a infraestrutura de nuvem e que as arquiteturas de software nativos em nuvem tornem a coexistência entre soluções de fornecedores mais compatível do que era para a NFV e outras aplicações monolíticas típicas.

Containers são ótimos no empacotamento de microsserviços. O Kubernetes orquestra aplicações formadas por microsserviços. Quando implantamos containers dentro das máquinas virtuais, os recursos de armazenamento, processamento e memória são usados para cada sistema operacional guest. Esses recursos são reservados sem o conhecimento das necessidades dos containers que serão eventualmente implantados em cada máquina virtual. Dessa forma, a utilização de recursos continua baixa como nas redes legadas, mas sua reserva é dimensionada com base no suporte a todo o complemento de containers em casos de pico de utilização – o que não corresponde à demanda de maneira consistente. Assim, como acontece com as redes legadas, quando a máquina virtual é dimensionada para tráfego máximo, o consumo regular de recursos costuma ser em torno de 20%.[7] Em bare-metal, o dimensionamento de containers para aplicações não reserva recursos dessa mesma forma: pode-se usar apenas os 20% dos recursos de processamento necessários como containers em máquinas virtuais. Portanto, é possível alocar mais containers em um servidor. Com as máquinas virtuais, baixa utilização e menor densidade de container resultam em maiores despesas operacionais e de capital com a compra e gerenciamento de hardware adicionais.

Para aplicações da edge ativadas por 5G, a otimização de recursos é crucial. Esses recursos incluem rede, RAN e cargas de trabalho de aplicações empresariais implantadas perto do cliente (para se beneficiarem da baixa latência) e largura de banda previsível. Muitos locais da edge são ambientes restritos ou limitados, então não há espaço para hardwares adicionais.

[6] Lettieri, Guiseppe, et al. "A study of I/O performance of virtual machines." The British Computer Society, Vol. 61 Nº 6, 2018.

[7] Liu, Ming, et al. "Understanding the virtualization "Tax" of scale-out pass-through GPUs in GaaS clouds: An empirical study." IEEE.org, fevereiro de 2015.

Além disso, o número de locais impõe restrições ao custo por local. Isso costuma limitá-los a um único servidor, uma vez que o custo total pode sair do controle rapidamente. Para cada servidor da edge, as máquinas virtuais geram uma despesa significativa com sistemas operacionais guest além do custo do sistema operacional host – o único necessário para containers em bare-metal. Containers em bare-metal são eficientes na viabilização de mais tráfego e mais serviços para impulsionar as margens de receita e de lucro.

Segurança

A segurança aprimorada das implantações de containers em bare-metal é um importante benefício para os CSPs implantando redes 5G e edge. Com o 4G/LTE e a NFV, um número relativamente pequeno de locais de nuvem privada era implantado e gerenciado por CSPs. O 5G possibilita arquiteturas mais distribuídas, incluindo múltiplas nuvens públicas e, possivelmente, milhares de instâncias de computação em nuvem próximas ou no local dos clientes. Além disso, aplicações que oferecem serviços de valor agregado aos clientes podem compartilhar as mesmas instâncias em nuvem com funções de rede cruciais. Essa combinação de escala e multilocação merece uma observação sobre a segurança. É um engano comum achar que é preciso uma máquina virtual para evitar que uma aplicação conteinerizada use recursos fundamentais para outra aplicação. Outro erro é pensar que essa máquinas oferecem um isolamento seguro das cargas de trabalho entre os locatários. Na verdade, a segurança é uma função do sistema operacional (Linux), e não do orquestrador (Kubernetes).

As máquinas virtuais isolam umas às outras em nós da nuvem para evitar que as cargas de trabalho "roubem" seus recursos. Quando usadas para NFV e aplicações monolíticas, elas são configuradas de acordo com os requisitos das aplicações em execução nelas. Se não forem operados corretamente, os containers implantados em máquinas virtuais podem "roubar" os recursos uns dos outros da mesma forma que no bare-metal Além disso, isolar cargas de trabalho locatárias em máquinas virtuais separadas pode reduzir a utilização. A solução para esses desafios é o Security-Enhanced Linux (SELinux), um componente incluído em várias distribuições do Linux. Ele é um módulo kernel de segurança originalmente desenvolvido pela Agência de Segurança Nacional (NSA) dos EUA que permite aos administradores controlar autorizações de acesso de alta granularidade aos componentes do sistema operacional e ao nó. Políticas de segurança são configuradas no próprio nó e, em seguida, cada container implantado no nó implementa uma política que protege seus interesses. Mesmo que os containers sejam implantados em máquinas virtuais, o SELinux deve ser acionado no sistema operacional guest para oferecer segurança. Portanto, não é preciso máquinas virtuais para proteger os containers. Elas só protegem a si mesmas de outras máquinas virtuais. Na verdade, incluir uma camada de máquinas virtuais pode ter um efeito negativo na segurança da rede do CSP e dificultar a conformidade.

Ainda sobre a segurança, é importante observar a conformidade com políticas de segurança determinadas pelo CSP, pelos locatários da rede e pelas agências regulatórias. Hardwares, sistemas operacionais, gerentes de infraestrutura virtual, aplicações, scripts de automação e outros componentes devem ser constantemente monitorados e atualizados para eliminar possíveis falhas de segurança descobertas por fornecedores, comunidades open source ou pelo setor. Por exemplo: imagine o quinto ano de operação da rede 5G. Nesse momento, haverá diversas gerações e até mesmo fornecedores de todos os componentes na rede. É preciso automatizar patches e reparos regulares para evitar desvios das expectativas de segurança. Como cada máquina virtual adiciona um sistema operacional guest para manutenção, torna-se muito mais complicado manter a conformidade.

Talvez, alguns containers ofereçam suporte a serviços que possuam SLAs específicos para segurança. A abstração de um hardware físico do container na forma de uma máquina virtual dificulta a consistência na segurança e no desempenho. No Kubernetes não existe o conceito do nó de nuvem específico no qual um container deve ser colocado. Não se pode garantir a segurança de um titular da carga de trabalho para conformidade comercial (por exemplo, o setor de cartões de pagamento) ou de saúde (por exemplo, a HIPAA) se não for possível estabelecer uma associação entre o hardware e o sistema operacional host para todos os containers. É mais fácil configurar e manter a conformidade de segurança quando os containers são implantados em bare-metal.

Conclusão

O 4G/LTE e a NFV ofereciam suporte a uma grande rede de telecomunicações usando um número relativamente pequeno de instâncias de nuvem privada. O 5G traz uma enorme mudança. Um CSP que queira se beneficiar de novas funcionalidades 5G, como latência baixa e previsível, largura de banda alta e arquitetura distribuída, alcançará o sucesso mais facilmente se implantar as novas funções de rede em containers executados em bare-metal. Eles conseguem realizar despesas de capital (CapEx) e operacionais (OpEx) mais baixas, são mais ágeis e competitivos, têm melhor desempenho em uma infraestrutura menor e oferecerem segurança de rede avançada.

As soluções open source oferecem o benefício da rapidez na inovação e na resolução de problemas, com comunidades inteiras focadas em superar os desafios do futuro. Com o open source, os CSPs não precisam se prender a apenas um fornecedor. A Red Hat oferece distribuições repletas de recursos de projetos open source com funcionalidades de valor adicionado para aprimorar a usabilidade e automatizar funções administrativas. Os CSPs podem criar redes 5G com:

Os CSPs trabalham com a Red Hat para executar suas funções de rede nativas em nuvem embare-metal e obter o máximo das distribuições 5G.

- Red Hat® OpenShift®
- Red Hat Ansible® Automation Platform
- Red Hat AMQ
- Red Hat 3Scale API Management
- Versão Red Hat do Quarkus, um ambiente de execução JavaTM nativo do Kubernetes

A Red Hat ajudou CSPs de todo mundo transformarem suas redes em NFV. Por sermos a principal colaboradora do OpenStack, do Kubernetes e de muitos outros projetos open source, temos um grande know-how sobre como transformar arquiteturas de rede de telecomunicações em nativas em nuvem. Além disso, funções de rede certificadas oferecidas por parceiros do ecossistema da Red Hat trazem mais opções para você fazer a implantação com confiança. Conheça as soluções de telecomunicações da Red Hat.

AMÉRICA LATINA
+54 11 4329 7300
latammktg@redhat.com

BRASIL
+55 11 3629 6000
marketing-br@redhat.com

Sobre a Red Hat

A Red Hat é a líder mundial em soluções empresariais de software open source, utilizando uma abordagem impulsionada pela comunidade para oferecer tecnologias confiáveis e de alto desempenho de Linux, nuvem híbrida, containers e Kubernetes. A Red Hat ajuda os clientes a integrar aplicações de TI novas e existentes, desenvolver aplicações nativas em nuvem, definir padrões por meio do nosso sistema operacional líder do setor e também automatizar, proteger e gerenciar ambientes complexos. Com serviços de consultoria, treinamento e suporte premiados, a Red Hat é a parceira de confiança das empresas listadas na Fortune 500. Como parceira estratégica de provedores de serviços em nuvem, integradores de sistemas, fornecedores de aplicações, clientes e comunidades open source, a Red Hat ajuda empresas a se prepararem para o futuro digital.

facebook.com/redhatinc
@redhatbr
linkedin.com/company/red-hat-brasil

O-F25719