

GVGO Grupo 3

Peer-to-Peer & Cloud Computing

Perguntas & Respostas

Peer to Peer

1. O BitTorrent (rede P2P desestruturada) utiliza um tracker centralizado para cada swarm. Cite três propriedades que as redes P2P estruturadas devem possuir que tornam o P2P completamente distribuído e ainda mantém um bom desempenho. Qual a solução usada para garantir essas propriedades nas redes P2P estruturadas? Cite exemplos de redes P2P estruturadas.

1. Cada nó mantém apenas uma pequena quantidade de informação sobre outros nós. Garantindo que não será caro manter esse índice atualizado.
2. Cada nó deve pesquisar entradas no índice rapidamente.
3. Cada nó pode usar o índice ao mesmo tempo, mesmo que outros nós apareçam e desapareçam. Essa propriedade significa que o desempenho do índice aumenta com o número de nós.

A solução é baseada nas DHTs (Distributed Hash Tables), que tem como funcionalidade básica o mapeamento de uma chave a um valor.

Exemplos de redes P2P estruturadas: Chord, CAN, Pastry e Tapestry.

2. Quais são as principais características dos protocolos de compartilhamento de arquivos Napster, Gnutella e BitTorrent?

- *Napster*: Servidor Centralizado, estilo híbrido, cria circuito virtual entre clientes, utilizado apenas pelo Napster.
- *Gnutella*: Servidor descentralizado, cada cliente também é servidor, estilo puro, trabalha com troca de mensagens, diversas aplicações utilizam.
- *BitTorrent*: Servidor descentralizado (utiliza Trackers), cada cliente também é servidor, utiliza sistema de fragmentação dos arquivos fornecendo maior possibilidade de compartilhamento global, utiliza DHT para realizar as interconexões, diversas aplicações utilizam.

3. Quais os principais focos na área de segurança P2P?

- Overlays Seguros e com capacidade de redundância em nível de Rede;
- Anonimato em nível de transporte;
- Resistência a "censura": publicação de arquivos de forma anônima em nível de aplicação.

4. Além do tradicional compartilhamento de arquivos, quais aplicações e porque podem ser desenvolvidas através da tecnologia p2p?

- *Streaming de vídeo*: Em contraste com o tradicional modelo cliente-servidor, com stream p2p o conteúdo é entregue ao usuário de forma totalmente distribuída. Pode ser usado em casos onde a demanda por um vídeo específico é grande, e que os mesmos dados possam ser entregues por usuários diferentes. Uma eficiente política de escolha dos peers é a chave para uma alta qualidade no serviço de streaming de vídeo.
- *Busca Distribuída*: Na busca p2p, cada indivíduo conectado a rede se comporta como uma fonte de pesquisa para o servidor. Em vez de um servidor central, cada participante da rede é um

repositório de pesquisa. Cada usuário contribuiria como uma memória cache para a base de dados. O principal problema de implementação é a dificuldade de selecionar peers que resultem em uma pesquisa com sucesso. Eg: YacY Search Engine

- *Aplicações Mobile*: Com a adoção de celulares e tablets com acesso à internet e a necessária massificação dos serviços nos aparelhos, cada vez mais sistemas que utilizam peer-to-peer migrarão para as plataformas mobile. Mensagem instantânea, compartilhamento de arquivos, stream de áudio/vídeo (o Skype atualmente usa p2p nas suas aplicações VoIP), redes privadas bluetooth para compartilhamento de arquivos e jogos.
- *E-commerce*: Compartilhar arquivos na web, via p2p, pagando direitos autorais e recebendo um royalty quando é feito um download de arquivo da sua máquina. A grande questão é como conferir segurança às transações e definir os passos necessários para uma compra num sistema p2p.

5. Quais são as aplicações p2p?

As aplicações mais conhecidas são a de compartilhamento de arquivos como Napster, Gnutella, Kazaa, BitTorrent, porém há outras aplicações com Skype, MSN, ICQ (tocar de mensagem instantânea). Também aplicações como JXTA e que se dedicam a organizar grupos de desenvolvimento, SETI@Home e Grade Computacional que fazem Computação Distribuída.

6. Quais os requisitos de uma aplicação peer-to-peer?

- *Escalabilidade Global*: Milhões de objetos acessados através de centenas de milhares de hospedeiros.
- *Otimização de Interações Locais Entre Vizinhos*: As aplicações devem colocar os recursos perto dos nós que mais os utilizam.
- *Disponibilidade*: O sistema deve detectar entrada ou saída de hospedeiros, a fim de balancear as cargas.
- *Segurança*: Uso de mecanismos de autenticação e criptografia.
- *Anonimidade*: Um hospedeiro deve ser capaz de negar plausivelmente a responsabilidade de posse e suprimento de recursos.

7. O que caracteriza um serviço P2P?

P2P é caracterizado por ser escalável. Isso se deve, porque cada nó opera ora consumindo (cliente), ora fornecendo (servidor) serviço ao sistema. Também lança mão de mecanismos fortemente descentralizados, sendo esperado que cada nó dependa o mínimo necessário de um nó central. Outra característica esperada é que cada elemento tem autonomia para ingressar ou deixar o sistema a qualquer momento e ainda assim, o sistema deve manter um bom grau de desempenho.

8. Quais as diferenças entre uma distribuição horizontal e uma distribuição vertical em P2P?

- *Distribuição vertical*: é a organização da aplicação em multi-camadas. Divide componentes logicamente diferentes em máquinas fisicamente diferentes. Com isso facilita o gerenciamento do sistema, pois cada máquina pode ser responsável por um grupo específico de funções.
- *Distribuição horizontal*: é a distribuição usada no P2P. É baseada na distribuição dos clientes e servidores. Um cliente ou servidor podem estar fisicamente divididos em partes lógicas equivalentes, cada uma operando sobre sua porção de dados, o que balanceia a carga.

Os processos que constituem o sistema são todos iguais, o que significa que as funções necessárias devem estar em todos os processos que compõem o sistema. Maioria das interações entre os processos é simétrica, com isso, cada processo atua como “cliente” e “servidor” ao mesmo tempo.

9. O que é Peer-to-peer?

Sistemas distribuídos consistindo de nós interconectados potencialmente autônomos que compartilham recursos (tal como conteúdo, ciclos de CPU, armazenamento e largura de banda) e são organizados em redes de sobreposição capazes de se adaptar a populações transitórias enquanto mantém conectividade e desempenho aceitáveis, tipicamente sem depender de uma entidade global centralizada.

10. Como podem ser classificados o serviço P2P?

Os serviços P2P podem ser classificados em dois modos: híbrido e descentralizado. O *modo descentralizado* funciona sem a ajuda de nenhum servidor ou repositório de informação intermediário. Os nós são totalmente auto-dependentes e funcionam tanto como cliente como servidor. Devido à sua dificuldade técnica e de implementação, essa solução não é muito utilizada. No *modo híbrido*, existe um servidor ou repositório responsável por realizar um intermediamento entre a conexão dos dois pares. Serviços como o Skype e BitTorrent são exemplos para este modo.

Cloud Computing

1. Como surgiu o Cloud Computing e qual sua motivação?

O conceito básico da Cloud Computing surgiu primeiramente nos anos 1960, quando o cientista da computação John McCarthy declarou que “computação pode algum dia se tornar uma utilidade pública”. Passando então de ser um bem, para ser um serviço. Em 1966 o livro “The Challenge of the Computer Utility” foi escrito por Douglas Parkhill já discutia as características básicas da Cloud Computing que conhecemos hoje, como a distribuição na forma de serviço, ser online, e passar a ilusão de capacidade infinita, comparando essa ideia às empresas provedoras de energia elétrica. Em 1999 o site Salesforce.com fez o primeiro movimento em direção ao Cloud Computing, colocando à disposição aplicações corporativas com uma interface simples. Em 2002 a Amazon lançou o Amazon Web Service.

Essas soluções ainda não eram o Cloud Computing como conhecemos hoje, com armazenamento de arquivos e computação em servidores remotos. Porém a evolução do web service para o Cloud Computing foi natural. Em 2006 o Google apresentou o Google Docs, que então de fato mostrou ao mundo que Cloud Computing já era uma realidade. Ainda em 2006 a Amazon colocou a disposição de usuários e pequenas empresas o Amazon’s Elastic Compute cloud (EC2), um serviço onde se poderia alugar computadores no qual rodariam as aplicações do cliente. Em 2008 apareceram as primeiras soluções de código aberto para a criação de nuvens privadas. A partir disso, a ideia (e realidade) do Cloud Computing se tornou cada vez mais difundida e utilizada, tanto por empresas quanto por usuários finais.

2. Apesar de possuir vantagens econômicas e de praticidade, a utilização de serviços de Cloud Computing implica assumir alguns riscos. Cite dois destes aspectos, justificando porque eles representam uma possível desvantagem na utilização dos serviços de computação na nuvem.

A utilização de serviços na nuvem focada no trânsito de dados confidenciais representa um risco para pessoas físicas/empresas pelo simples fato de não haver controle sobre onde os dados estão, já que existe a transparência para o usuário sobre o armazenamento de seus dados. Assim, dados confidenciais podem ficar expostos a terceiros (notadamente, o administrador do serviço de cloud).

Um outro risco é a perda de conexão com o serviço de cloud. A princípio, não há como garantir que os dados hospedados em servidores remotos estarão disponíveis sempre que necessário; uma simples falha de conexão do lado do usuário ou do lado do serviço de cloud pode impossibilitar o acesso aos dados quando necessário. Ainda existe o risco da perda de dados devido à destruição física das mídias de armazenamento, por exemplo, caso os administradores do serviço da nuvem não criem cópias de segurança dos dados de seus clientes.

3. O que define a filosofia do Cloud Computing, que preserva sua noção mesmo dentre implementações e tecnologias diferentes?

O conceito de Cloud se fundamenta na flexibilização do uso de poder computacional e capacidade de armazenamento, normalmente de grandes servidores, para usuários dos quais é omitido o funcionamento do sistema por baixo. Em efeito, os servidores, além de disponibilizar com eficiência e alta disponibilidade seus serviços, normalmente acessíveis por uma interface simplificada, têm a capacidade de balancear o uso de seus recursos dentre os outros servidores da mesma rede de maneira a otimizar essa disponibilidade.

Hoje em dia, independentemente do modelo de serviço empregado pela Cloud ou do tipo de aplicação ou plataforma que faz uso do serviço, o uso de Cloud Computing têm facilitado o acesso de pessoas e empresas à recursos computacionais, sejam estes pequenos ou em grande escala. Essa facilidade tem alavancado o surgimento de novas empresas, que têm tido seus custos de infraestrutura reduzidos drasticamente.

4. Qual a diferença entre GRID e CLOUD?

O Grid provê a seus usuários um certo controle sobre os recursos que ele está disponibilizando. Eventualmente, caso o Grid não possua recursos suficientes a disposição do usuário, sua aplicação ficará em uma fila aguardando para ser atendida. Já a utilização de uma Cloud parte do princípio que os recursos sempre estarão disponíveis, assim o usuário não tem um acesso aos recursos propriamente ditos, mas apenas à disponibilidade e capacidade de processamento e armazenamento que eles dispõem.

5. Quando deve-se optar por GRID Computing em detrimento à Cloud Computing?

Aplicações científicas que envolvem processamento pesado de dados são melhor executadas em ambientes Grid. Em contrapartida, aplicações mais leves voltadas a atingir um grande número de usuários são mais indicadas em um ambiente Cloud.

6. O que e quais são o modelos de implantação de Cloud Computing?

- *Privado* - O serviço é fornecido exclusivamente a uma organização. A propriedade, gerência e operação podem ser feitas pela própria organização, ser terceirizada ou uma mistura de ambos. A infraestrutura pode ser interna ou externa.
- *Comunidade* - O serviço é fornecido a uma comunidade de organizações que com interesses em comum (como repartições públicas e universidades). A propriedade, gerência e operação podem ser feitas por uma ou mais organizações da comunidade, ser terceirizada ou uma mistura de ambos. A infraestrutura pode ser interna ou externa.
- *Público* - O serviço é fornecido ao público em geral. A propriedade, gerência, operação, assim como a instalação são de responsabilidade do provedor.
- *Híbrido* - Neste modelo a nuvem é uma composição de nuvens distintas (públicas, privadas, de comunidade) que atuam como entidades únicas mas são conectadas, permitindo, por exemplo, que uma empresa mantenha dados vitais em uma nuvem privada e utilize serviços em uma nuvem pública.

7. Quais são os principais modelos de serviços de Cloud Computing e suas características?

- *IaaS (Infrastructure as a Service)* - Os provedores de Cloud Computing fornecem um conjunto de recursos que formam uma infraestrutura para seus clientes. Ou seja, computadores (físicos ou máquinas virtuais), redes, espaço de armazenamento, firewalls, etc. Neste modelo, seus usuários são os responsáveis pela manutenção e instalação dos softwares contidos na infraestrutura (inclusive sistemas operacionais), e o valor de cobrança dos serviços é medido pela quantidade de recursos desta infraestrutura o usuários está consumindo.
- *PaaS (Platform as a Service)* - Os provedores de cloud fornecem uma plataforma pronta para seus clientes, incluindo geralmente um sistema operacional, ambientes de desenvolvimento, bancos de dados, tirando a responsabilidade do usuário de ter que gerenciar os recursos de hardware e software. Os recursos de armazenamento e computação por trás do serviço são auto-escaláveis para a demanda da aplicação, sem precisar que o usuário aloque recursos manualmente.
- *SaaS (Software as a Service)* - Os provedores simplesmente fornecem aos seus usuários o acesso a um ou mais softwares armazenados na nuvem através de programas cliente (geralmente web browsers). Os usuários não gerenciam a infraestrutura nem a plataforma da nuvem, tirando a necessidade de que o software seja instalado na máquina do cliente. Isto simplifica a manutenção e o suporte do software.

8. Quais são os principais serviços de Cloud Computing disponíveis no mercado atualmente e seus modelos de serviço?

Amazon (IaaS e PaaS), VMware (PaaS), OpenStack (IaaS), Heroku (PaaS), Salesforce.com (PaaS e

SaaS), Windows Azure (PaaS e IaaS), Google (PaaS e SaaS), Hadoop (PaaS).

9. Do ponto de vista econômico, quais os principais atrativos à utilização de recursos em Cloud em detrimento a modelos clássicos de computação?

O modelo de transformação de recursos computacionais em serviços externos tem diversas vantagens econômicas, por exemplo:

- Os provedores de serviço se beneficiam de economia de escala na aquisição e manutenção de seus recursos, reduzindo os custos de implantação e operação;
- A eficiência da utilização dos recursos tende a ser muito elevada, eliminando desperdícios de recursos presentes em grande parte dos sistemas in-house. Por exemplo, servidores têm como característica serem dimensionados para sua previsão de carga de pico, sendo marginalmente utilizados quando fora desta;
- O overhead atribuído a recursos computacionais, como manutenção, é dividido entre todos os clientes do recurso.

10. Quais são os 7 princípios de segurança do Cloud Computing? Descreva brevemente cada um deles

- *Acesso privilegiado de usuários* - ter o máximo de informação sobre quem vai gerenciar os dados, ou seja, quem será o administrador dos dados..
- *Compliance com regulamentação* - As empresas são responsáveis pela segurança, integridade e a confidencialidade de seus dados. Os fornecedores de cloud computing devem estar preparados para auditorias externas e certificações de segurança.
- *Localização dos dados* - O fornecedor deve estar disposto a se comprometer a armazenar e a processar dados em jurisdições específicas, assumindo um compromisso em contrato de obedecer os requerimentos de privacidade que o país de origem da empresa pede.
- *Segregação dos dados* - Geralmente uma empresa divide um ambiente com dados de diversos clientes. Procurar entender o que é feito para a separação de dados, que tipo de criptografia é segura o suficiente para o funcionamento correto da aplicação.
- *Recuperação dos dados* - O fornecedor em cloud deve saber onde estão os dados da empresa e o que acontece para recuperação de dados em caso de erros. Importante ter um plano de recuperação completa e um tempo estimado para tal.
- *Apoio à investigação* - Obter um compromisso contratual com a empresa fornecedora do serviço e uma evidência de sucesso no passado para esse tipo de investigação.
- *Viabilidade em longo prazo* - Importante haver um plano de recuperação de dados e o formato para que possa ser utilizado em uma aplicação substituta

11. Cite os benefícios econômicos, de manutenção, de confiabilidade e de independência de dispositivo e localização da arquitetura cloud computing.

Cloud computing oferece grandes benefícios econômicos, uma vez que os recursos fornecidos pelo provedor da arquitetura podem ser gratuitos ou cobrados sob demanda. Assim, o usuário paga somente pelo que usar. Usuários dos serviços de cloud, não necessariamente, precisam se preocupar com a manutenção dos recursos, pois todo o cuidado sobre a infra-estrutura será responsabilizada ao provedor, a não ser que o usuário deseje ter um maior controle sobre os recursos. O provedor dos recursos, normalmente, também fornece um serviço de tolerância a falhas com o uso de múltiplos servidores que proverão a redundância necessária às necessidades do usuário. Como os dados e recursos dos usuários estão na nuvem, existe uma independência de dispositivo para acesso, sendo possível assim acessar os recursos na nuvem desde um desktop até um tablet. Essa independência também é verdadeira em relação a locais de acesso, sendo necessário nesse caso um simples ponto de acesso à rede para se ter acesso aos recursos contidos na nuvem.