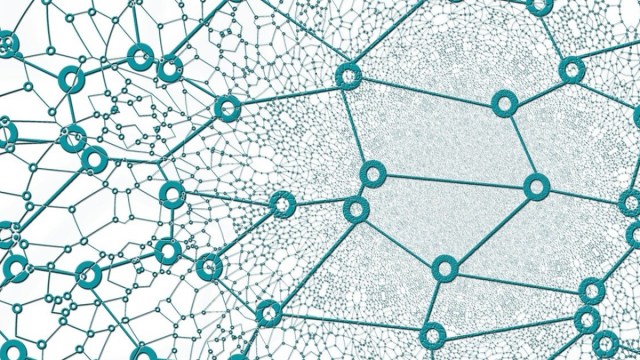


Licenciatura em Engenharia Multimédia

Computação na Nuvem

**A computação distribuída, princípios gerais do funcionamento**



Viorel Cojocari nº 21467

Vasco Sarrico nº 21459

Vanessa Brioso nº 21466

Professor Pedro Almeida

Turma Pós-Laboral

Ano letivo 2021 / 2022

Índice

[I. Introdução 3](#_Toc131615092)

[Definição de computação distribuída 4](#_Toc131615093)

[Importância da computação distribuída 5](#_Toc131615094)

[II. Arquiteturas de sistemas distribuídos 6](#_Toc131615095)

[Redes ponto a ponto 7](#_Toc131615096)

[Sistemas cliente-servidor 8](#_Toc131615097)

[Sistemas em cluster 9](#_Toc131615098)

[III. Compartilhamento de recursos em sistemas distribuídos 10](#_Toc131615099)

[Uso de vários computadores para processamento e armazenamento de dados 11](#_Toc131615100)

[Comunicação entre computadores 11](#_Toc131615101)

[IV. Segurança e confiabilidade em sistemas distribuídos 12](#_Toc131615102)

[Replicação de dados 13](#_Toc131615103)

[Deteção e tratamento de falhas 14](#_Toc131615104)

[V. Protocolos de comunicação em sistemas distribuídos 15](#_Toc131615105)

[Sincronização de tarefas 15](#_Toc131615106)

[Transmissão e recepção de dados 17](#_Toc131615107)

[VI. Conclusão 18](#_Toc131615108)

[Benefícios da computação distribuída 19](#_Toc131615109)

[Desafios e oportunidades futuras 20](#_Toc131615110)

# **I. Introdução**

# **Definição de computação distribuída**

Computação distribuída é um modelo de processamento de dados em que um sistema computacional é composto por múltiplos componentes autônomos, como computadores, dispositivos móveis, servidores, sensores, entre outros, que trabalham juntos de forma coordenada para realizar uma tarefa computacional.

Nesse modelo, cada componente do sistema é responsável por uma parte da computação, e a comunicação e coordenação entre eles é realizada por meio de uma rede de comunicação, como a internet. Essa abordagem permite que tarefas complexas sejam divididas em partes menores, que podem ser processadas em paralelo, aumentando assim a eficiência e a velocidade de processamento.

A computação distribuída é utilizada em diversas áreas, como processamento de grandes volumes de dados, inteligência artificial, sistemas de controle de tráfego aéreo, entre outras aplicações que requerem alta capacidade de processamento e coordenação entre múltiplos componentes.

**Importância da computação distribuída**

Algumas das principais razões pelas quais a computação distribuída é importante são:

1. Escalabilidade: Isto significa que é possível aumentar a capacidade de processamento de um sistema sem ter que substituí-lo completamente por um novo sistema, o que pode ser mais caro e complexo.
2. Redundância: A computação distribuída também permite que o sistema tenha redundância em caso de falhas em algum componente. Isso significa que, se um componente falhar, outros componentes do sistema podem continuar processando a tarefa, garantindo assim a continuidade do processamento.
3. Eficiência: A computação distribuída permite que tarefas sejam divididas em partes menores, que podem ser processadas em paralelo, aumentando a eficiência e a velocidade de processamento. Isso é especialmente importante em tarefas que requerem alto poder de processamento, como inteligência artificial, simulações complexas e análise de grandes volumes de dados.
4. Flexibilidade: A computação distribuída permite que diferentes componentes do sistema tenham diferentes funções e responsabilidades. Isso significa que é possível criar sistemas mais flexíveis e adaptáveis, que podem ser ajustados para atender às necessidades específicas de uma determinada tarefa ou aplicação.
5. Redução de custos: A computação distribuída pode ser uma opção mais econômica do que investir em um único sistema de alto desempenho. Isso porque, ao distribuir o processamento entre vários componentes, é possível utilizar recursos existentes e mais baratos, em vez de ter que investir em um único sistema de alto desempenho, que pode ser mais caro e difícil de manter.

# **II. Arquiteturas de sistemas distribuídos**

As arquiteturas de sistemas distribuídos podem ser classificadas de acordo com a forma como os componentes são organizados e como as comunicações entre eles são realizadas. As principais arquiteturas de sistemas distribuídos são:

1. Cliente-Servidor: Nesta arquitetura, os serviços são fornecidos por um servidor centralizado e os clientes fazem solicitações para acessar esses serviços. Essa arquitetura é comum em sistemas de banco de dados, onde os clientes acessam informações armazenadas no servidor.
2. Peer-to-Peer: Nesta arquitetura, todos os componentes são iguais e colaboram para realizar a tarefa. Cada componente é tanto um cliente quanto um servidor, e eles se comunicam uns com os outros diretamente para compartilhar recursos ou trocar informações. Essa arquitetura é comum em sistemas de compartilhamento de arquivos, como o BitTorrent.
3. Arquitetura em camadas: Nesta arquitetura, os componentes são organizados em camadas, onde cada camada tem uma função específica. A comunicação entre as camadas é realizada de forma estruturada e organizada, com cada camada fornecendo serviços para a camada superior. Essa arquitetura é comum em sistemas de rede, onde as camadas podem incluir a camada de aplicação, a camada de transporte e a camada de rede.
4. Arquitetura orientada a serviços: Nesta arquitetura, os serviços são fornecidos como unidades independentes, que podem ser combinadas para formar sistemas mais complexos. Cada serviço é autônomo e pode ser acessado por meio de uma interface bem definida. Essa arquitetura é comum em sistemas de comércio eletrônico, onde serviços como pagamentos, envio de pedidos e gerenciamento de estoque podem ser fornecidos por diferentes fornecedores.
5. Arquitetura de microsserviços: Nesta arquitetura, o sistema é dividido em microsserviços independentes, cada um executando uma tarefa específica. Cada microsserviço é autônomo e pode ser escalado independentemente. A comunicação entre os microsserviços é realizada por meio de APIs e protocolos bem definidos. Essa arquitetura é comum em sistemas modernos de computação em nuvem e em aplicativos web escaláveis.

**Redes ponto a ponto**

**Sistemas cliente-servidor**

**Sistemas em cluster**

# **III. Compartilhamento de recursos em sistemas distribuídos**

No contexto de sistemas distribuídos, a computação em nuvem é um modelo que permite o compartilhamento eficiente de recursos entre diferentes sistemas e usuários. Ao adotar a computação em nuvem, os sistemas distribuídos podem acessar recursos sob demanda, de forma flexível e escalável, sem precisar investir em infraestrutura própria.

Além disso, a computação em nuvem oferece recursos de segurança, backup e recuperação de desastres que podem ser críticos para sistemas distribuídos que lidam com dados sensíveis e operações críticas. Dessa forma, o compartilhamento de recursos em sistemas distribuídos por meio da computação em nuvem pode trazer benefícios significativos em termos de eficiência, escalabilidade e segurança.

## **Uso de vários computadores para processamento e armazenamento de dados**

O uso de vários computadores para processamento e armazenamento de dados é conhecido como computação distribuída e é uma abordagem que permite lidar com problemas computacionais complexos de forma mais eficiente e escalável. A computação distribuída é baseada na ideia de dividir um problema em tarefas menores e distribuí-las entre vários computadores que trabalham em paralelo para processar essas tarefas.

Essa abordagem tem sido amplamente adotada em áreas como inteligência artificial, aprendizado de máquina, análise de big data e simulação numérica, onde o processamento de grandes volumes de dados é comum. Ao distribuir o processamento de dados em vários computadores, é possível reduzir o tempo necessário para realizar tarefas computacionais complexas, aumentando assim a eficiência e reduzindo os custos.

Além disso, a computação distribuída também permite o armazenamento de grandes quantidades de dados em vários computadores, formando um sistema de armazenamento distribuído. Esse sistema pode oferecer maior disponibilidade, confiabilidade e escalabilidade em comparação com um sistema de armazenamento centralizado.

No entanto, a computação distribuída também apresenta desafios, como a coordenação das tarefas distribuídas, o gerenciamento da comunicação entre os computadores e a garantia de que todos os dados sejam mantidos em sincronia. Portanto, a escolha da abordagem de computação distribuída deve ser cuidadosamente avaliada para garantir que atenda às necessidades específicas de processamento e armazenamento de dados.

**Comunicação entre computadores**

A comunicação entre computadores é um aspecto essencial da computação distribuída e permite que vários computadores trabalhem juntos para realizar tarefas complexas. Existem diferentes formas de comunicação entre computadores, como a comunicação por meio de cabos de rede, ondas de rádio, comunicação por satélite e outros meios.

A comunicação entre computadores pode ser realizada por meio de diferentes protocolos de comunicação, como o TCP/IP, que é amplamente utilizado na Internet. Esse protocolo estabelece uma conexão confiável entre os computadores e garante que os dados sejam transferidos com segurança e integridade.

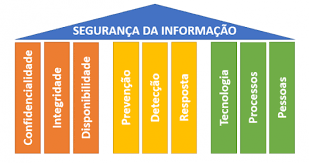
Além disso, a comunicação entre computadores pode ocorrer de diferentes formas, como a comunicação ponto a ponto, onde dois computadores se comunicam diretamente, ou a comunicação em rede, onde vários computadores são conectados a uma rede e podem se comunicar uns com os outros.

A comunicação entre computadores pode ser realizada em diferentes níveis, desde a comunicação de baixo nível entre os componentes de hardware até a comunicação de alto nível entre os aplicativos que rodam em diferentes computadores.

Por fim, a comunicação entre computadores pode apresentar desafios, como problemas de latência, perda de pacotes, congestionamento de rede e outros fatores que podem afetar o desempenho e a qualidade da comunicação. Portanto, é importante entender os diferentes aspectos da comunicação entre computadores e adotar as práticas recomendadas para garantir uma comunicação eficiente e confiável.

# **IV. Segurança e confiabilidade em sistemas distribuídos**

A segurança e confiabilidade em sistemas distribuídos são aspetos críticos para garantir o bom funcionamento dos sistemas e proteger os dados e recursos que são partilhados entre os vários elementos dos mesmos. Existem vários desafios e ameaças que afetam a segurança em sistemas distribuídos, como falhas de hardware, problemas de rede, ataques de segurança cibernética e erros de programação.

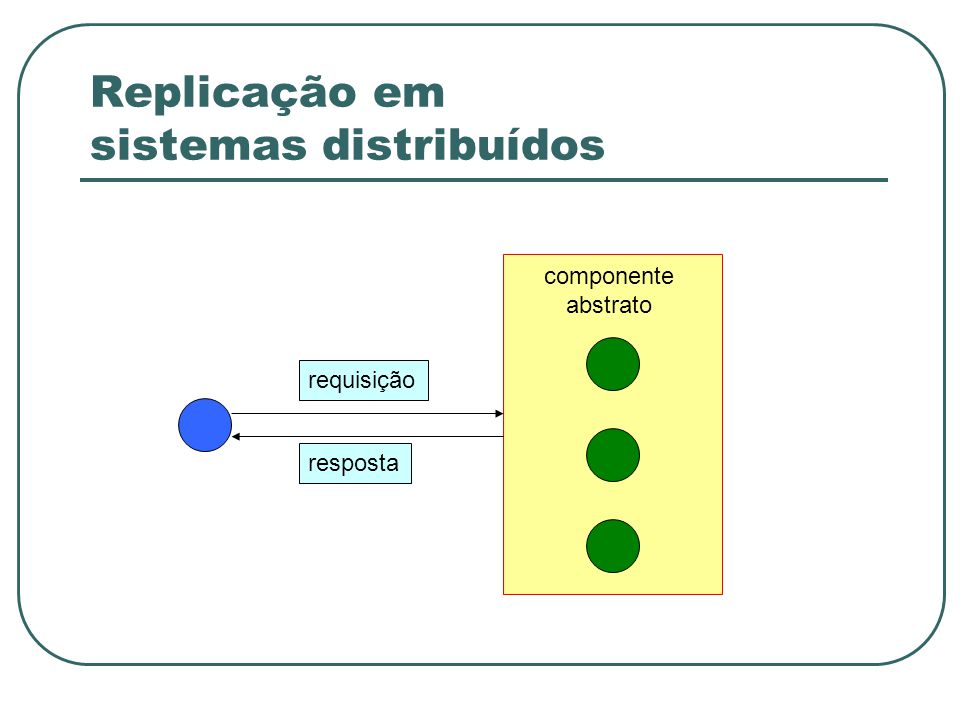


Para garantir a segurança em sistemas distribuídos, é importante adotar medidas e práticas recomendadas, como criptografia, autenticação, autorização, monitoramento de rede, backup e recuperação de desastres. Além disso, é fundamental realizar testes de segurança com frequência para identificar fragilidades e falhas no sistema.

Outro aspeto importante é a confiabilidade dos sistemas distribuídos, que está relacionada à capacidade do sistema de manter a disponibilidade, integridade e consistência dos dados e recursos partilhados. Para garantir isto em sistemas distribuídos, é necessário adotar técnicas como replicação de dados, deteção e correção de falhas, gerir transações e distribuição de carga.

**Replicação de dados**

A replicação de dados é uma técnica utilizada em sistemas distribuídos para melhorar a disponibilidade, confiabilidade e desempenho do sistema. Ela consiste em manter cópias idênticas dos dados em diferentes nós do sistema, de forma que, caso um nó falhe, os dados ainda estejam disponíveis em outros nós.



Existem diferentes tipos de replicação de dados, como a replicação ativa e a replicação passiva. Na replicação ativa, todas as cópias dos dados são atualizadas simultaneamente, de forma que qualquer nó possa ser usado para aceder ou atualizar os dados. Já na replicação passiva, uma cópia principal dos dados é mantida em um nó mestre, enquanto as cópias secundárias são mantidas em nós escravos, que são atualizadas periodicamente com as mudanças feitas na cópia principal.

A replicação de dados traz vários benefícios para os sistemas distribuídos, como melhoria da disponibilidade e desempenho do sistema, e redução do tempo de recuperação dos dados em caso de falhas. Além disso, a replicação de dados pode ser utilizada para distribuir a carga de trabalho entre os diferentes nós do sistema, melhorando o desempenho e reduzindo o tempo de resposta do sistema.

## No entanto, a replicação de dados também apresenta desafios e complexidades, como a necessidade de manter a consistência dos dados em todas as cópias e a sincronização de dados entre os diferentes nós do sistema. Portanto, é importante adotar técnicas de replicação de dados adequadas para garantir a disponibilidade, confiabilidade e desempenho do sistema distribuído.

**Deteção e tratamento de falhas**

A deteção e tratamento de falhas são essenciais em sistemas distribuídos para garantir a disponibilidade, confiabilidade e desempenho do sistema. As falhas podem ocorrer em diferentes componentes do sistema, como hardware, software, rede e aplicativos, e podem ser causadas por vários fatores, como erros de programação, sobrecarga de rede, defeitos de hardware, entre outros.

A deteção de falhas em sistemas distribuídos pode ser realizada de diferentes maneiras, como por meio do monitoramento de desempenho do sistema, testes de integridade e confiabilidade, análise de *logs* e alarmes. É importante que os sistemas distribuídos sejam projetados para detetar as falhas o mais cedo possível, para que as medidas corretivas possam ser tomadas antes que os usuários percebam qualquer impacto negativo no desempenho ou na disponibilidade do sistema.

Uma vez detetadas as falhas, é necessário que haja um plano de tratamento de falhas eficiente para corrigir o problema o mais rápido possível e minimizar o impacto no sistema. O tratamento de falhas em sistemas distribuídos pode ser realizado de diferentes maneiras, como por meio da recuperação de desastres, substituição de componentes defeituosos, redireccionamento de fluxo de dados e replicação de dados.

A replicação de dados é uma técnica importante para o tratamento de falhas, pois permite que, em caso de falha em um nó, outro nó possa assumir a carga de trabalho sem interrupções. Além disso, é importante que os sistemas distribuídos tenham mecanismos de backup e recuperação de dados para garantir a integridade dos dados em caso de falhas.

Em resumo, a deteção e tratamento de falhas são fundamentais em sistemas distribuídos para garantir a disponibilidade, confiabilidade e desempenho do sistema. É importante que os sistemas sejam projetados para detetar as falhas o mais cedo possível e que haja um plano de tratamento de falhas eficiente para corrigir o problema o mais rápido possível e minimizar o impacto no sistema.

# **V. Protocolos de comunicação em sistemas distribuídos**

Os protocolos de comunicação são conjuntos de regras e padrões que governam a comunicação entre computadores em sistemas distribuídos. Eles são essenciais para garantir que os diferentes computadores em um sistema distribuído possam se comunicar de forma eficaz e coordenada para executar uma tarefa ou compartilhar dados.

Os protocolos de comunicação em sistemas distribuídos incluem várias etapas que permitem que os computadores se comuniquem uns com os outros. Essas etapas podem incluir:

1. Estabelecimento da conexão: o protocolo define como um computador estabelece uma conexão com outro.
2. Transmissão de dados: o protocolo define como os dados são transmitidos de um computador para outro.
3. Recebimento de dados: o protocolo define como um computador recebe os dados enviados por outro.
4. Tratamento de erros: o protocolo define como os erros de comunicação são tratados.
5. Encerramento da conexão: o protocolo define como uma conexão é encerrada após a conclusão da tarefa ou transmissão de dados.

Existem vários protocolos de comunicação amplamente usados em sistemas distribuídos, incluindo TCP/IP, HTTP, FTP, SMTP, entre outros. Cada protocolo é projetado para atender a uma finalidade específica e possui diferentes níveis de confiabilidade, segurança e eficiência. São essenciais para garantir a coordenação e a integração de sistemas distribuídos, permitindo que os computadores trabalhem juntos de forma eficiente para atender aos objetivos.

## **Sincronização de tarefas**

A sincronização de tarefas é um aspeto importante dos protocolos de comunicação em sistemas distribuídos, que permite que diferentes computadores trabalhem juntos de forma coordenada para executar uma tarefa específica. Em sistemas distribuídos, as tarefas geralmente são divididas em partes menores e executadas em diferentes computadores, que precisam ser sincronizados para garantir que a tarefa seja concluída com sucesso.

Existem vários protocolos de sincronização de tarefas que podem ser usados em sistemas distribuídos, incluindo:

1. Protocolos de bloqueio: esses protocolos são usados para controlar o acesso a recursos compartilhados em sistemas distribuídos. Eles permitem que apenas um computador aceda a um recurso compartilhado de cada vez, evitando conflitos e garantindo que o recurso seja usado corretamente.
2. Protocolos de transação: esses protocolos são usados para garantir a integridade dos dados, permitindo que as transações sejam executadas em diferentes computadores de forma coordenada. Eles garantem que as transações sejam concluídas com sucesso ou que sejam totalmente desfeitas caso ocorra um erro.
3. Protocolos de eleição: esses protocolos são usados para selecionar um líder ou coordenador em um grupo de computadores em um sistema distribuído. O líder é responsável por coordenar as tarefas e garantir que todos os computadores estejam trabalhando juntos de forma coordenada.
4. Protocolos de sincronização de relógio: esses protocolos são usados para sincronizar os relógios de diferentes computadores em um sistema distribuído, permitindo que eles coordenem suas ações e eventos com precisão.

Os protocolos de sincronização de tarefas ajudam a garantir que os diferentes computadores em um sistema distribuído possam trabalhar juntos de forma coordenada para concluir uma tarefa com sucesso.

**Transmissão e receção de dados**

A troca de dados são aspetos fundamentais dos protocolos de comunicação. São projetados para permitir que os computadores se comuniquem entre si, transmitindo e recebendo dados de diferentes tipos e formatos.

A transmissão de dados envolve o envio de informações de um computador para outro em um sistema distribuído. Para que isso aconteça, os dados precisam ser codificados em um formato que possa ser entendido pelos computadores de destino e enviados por meio de uma rede de comunicação. Existem vários protocolos de transmissão de dados que podem ser usados em sistemas distribuídos, incluindo TCP/IP, HTTP, FTP, SMTP, entre outros.

A receção de dados envolve o processo de receber informações enviadas por outro computador em um sistema distribuído. Para que isso aconteça, o computador recetor precisa estar conectado à rede de comunicação e ter um protocolo de receção de dados configurado para receber e interpretar los corretamente. O protocolo de receção de dados pode incluir verificações de integridade, como *checksums* ou *hashes*, para garantir que os dados recebidos sejam precisos e completos.

Além disso, em sistemas distribuídos, a transmissão e receção de dados podem ser afetadas por vários fatores externos, como latência, perda de pacotes, congestionamento de rede e outros problemas que podem afetar a qualidade da transmissão de dados. Por esse motivo, os protocolos de comunicação em sistemas distribuídos precisam ser projetados para lidar com esses desafios e garantir que os dados sejam transmitidos e recebidos com segurança e eficácia.

# **VI. Conclusão**

A computação distribuída é uma abordagem para a construção de sistemas de computação que permite que múltiplos computadores trabalhem juntos de forma coordenada para executar tarefas complexas. Essa abordagem oferece várias vantagens em relação aos sistemas centralizados, incluindo maior escalabilidade, maior disponibilidade, melhor desempenho e maior resiliência a falhas.

No entanto, a computação distribuída também apresenta desafios significativos em termos de gerenciamento de recursos, sincronização de tarefas, transmissão e recepção de dados, segurança e outros aspectos críticos. Para lidar com esses desafios, são necessários protocolos e algoritmos específicos que permitem que os computadores em um sistema distribuído trabalhem juntos de forma coordenada e eficiente.

No geral, a computação distribuída é uma abordagem importante para a construção de sistemas de computação de grande escala e complexidade. À medida que as demandas por recursos computacionais continuam a crescer, a computação distribuída provavelmente se tornará cada vez mais importante para a construção de sistemas eficientes e escaláveis que possam lidar com as necessidades do mundo moderno.

## **Benefícios da computação distribuída**

A computação distribuída oferece uma série de benefícios em relação aos sistemas centralizados, incluindo:

1. Escalabilidade: a computação distribuída permite que os recursos computacionais sejam facilmente adicionados ou removidos do sistema, o que permite que o sistema cresça ou diminua de acordo com as necessidades.
2. Melhor desempenho: a computação distribuída pode executar tarefas mais rapidamente do que os sistemas centralizados, porque as tarefas podem ser distribuídas em vários computadores e executadas em paralelo.
3. Disponibilidade: a computação distribuída pode oferecer maior disponibilidade do sistema, porque se um computador falhar, outros computadores podem assumir as tarefas que estavam sendo executadas pelo computador que falhou.
4. Redundância: a computação distribuída pode ser usada para criar sistemas redundantes, que podem ser usados para garantir que as informações e os dados estejam sempre disponíveis.
5. Flexibilidade: a computação distribuída pode ser usada para criar sistemas flexíveis, que podem ser adaptados às necessidades específicas de diferentes usuários ou aplicações.
6. Custo-efetividade: a computação distribuída pode ser mais econômica do que os sistemas centralizados, porque permite que os recursos computacionais sejam compartilhados e utilizados de forma mais eficiente.
7. Resiliência a falhas: a computação distribuída é mais resiliente a falhas do que os sistemas centralizados, porque os dados e as tarefas podem ser distribuídos em vários computadores e redundâncias podem ser criadas para garantir que o sistema continue funcionando mesmo em caso de falhas.

No geral, a computação distribuída pode ser uma solução eficaz para lidar com os desafios computacionais modernos, oferecendo maior escalabilidade, desempenho, disponibilidade, flexibilidade e resiliência a falhas.

## **Desafios e oportunidades futuras**

Embora a computação distribuída apresente muitos benefícios, também existem desafios e oportunidades futuras a serem considerados. Alguns desses desafios e oportunidades incluem:

1. Gerenciamento de recursos: gerenciar recursos distribuídos em vários computadores pode ser complexo e desafiador. Futuras soluções de gerenciamento de recursos devem ser desenvolvidas para garantir que os recursos sejam alocados de forma eficiente e otimizada.
2. Segurança: com tantos dispositivos conectados a um sistema distribuído, a segurança é uma grande preocupação. A computação distribuída deve ser protegida contra invasões e ataques cibernéticos, o que pode ser difícil devido à natureza distribuída da computação.
3. Sincronização de tarefas: garantir que as tarefas sejam executadas em ordem correta e de forma síncrona pode ser um desafio em sistemas distribuídos. Novos protocolos e algoritmos devem ser desenvolvidos para garantir que as tarefas sejam sincronizadas de forma eficiente.
4. Transmissão e recepção de dados: o envio e o recebimento de dados em sistemas distribuídos podem ser problemáticos. As soluções futuras devem se concentrar em melhorar a eficiência da transmissão e garantir que os dados cheguem ao destino corretamente e de forma segura.
5. Análise de dados em tempo real: a computação distribuída pode ser usada para análise de dados em tempo real, mas pode ser desafiador garantir que a análise seja precisa e rápida o suficiente. Soluções futuras devem se concentrar em melhorar a eficiência da análise de dados em tempo real.
6. Oportunidades para a IoT: a computação distribuída é fundamental para a Internet das Coisas (IoT), permitindo que dispositivos se comuniquem uns com os outros e realizem tarefas complexas. A computação distribuída pode ser usada para melhorar a eficiência da IoT e permitir novas aplicações e serviços.
7. Aumento da capacidade de processamento: com a computação distribuída, é possível aumentar significativamente a capacidade de processamento disponível para executar tarefas complexas. Isso pode levar a novas aplicações e serviços que não eram possíveis anteriormente.

No geral, a computação distribuída oferece muitas oportunidades para melhorar a eficiência e a capacidade de processamento dos sistemas de computação modernos, mas também apresenta desafios que precisam ser superados. O futuro da computação distribuída dependerá da capacidade de enfrentar esses desafios e aproveitar as oportunidades apresentadas.

**Webgrafia**