

PEER DATA MANAGEMENT SYSTEM

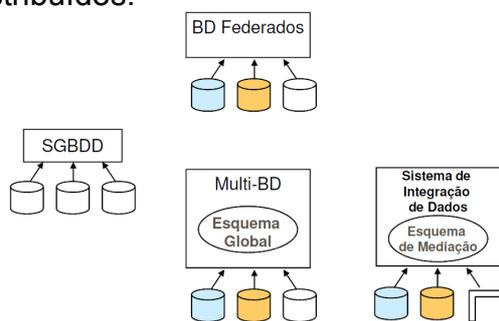
**INTRODUÇÃO, INFRA-ESTRUTURA E
MAPEAMENTO DE ESQUEMAS**

AGENDA

- Data Management System
- Peer Data Management System
- P2P
- Infra-estrutura
- Funcionamento do PDMS
- Mapeamento de Esquemas

DATA MANAGEMENT SYSTEM

- Necessidade de acesso a dados armazenados em múltiplas fontes de dados
- Inúmeros tipos de sistemas de gerenciamento de dados distribuídos.



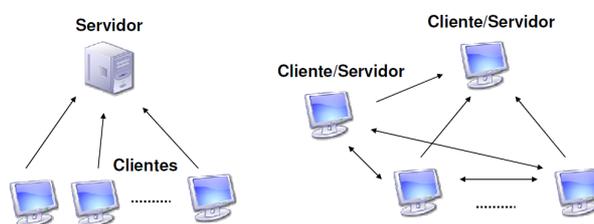
PEER DATA MANAGEMENT SYSTEM

- Ausência de controle Centralizado
- Auto-organização
- Comportamento dinâmico
- Dados não-estruturados
- Busca por palavra-chave (conteúdo)
- Disseminação dos sistemas P2P



PEER-TO-PEER

- Paradigma genérico de arquitetura de software, que contrasta com a computação cliente-servidor
- Peers que compartilham e/ou utilizam os recursos de outro(s) peer(s) para execução de tarefas de modo descentralizado



PEER-TO-PEER

- Cliente/Servidor
- Ausência de um ponto único de falha
- Escalabilidade
- Diversificação dos recursos
 - Dados
 - Espaço de armazenamento
 - Poder computacional



PEER-TO-PEER



- Vantagens
 - Aumento do poder computacional
 - Redução de custo para o compartilhamento de recursos
 - Autonomia dos pontos participantes

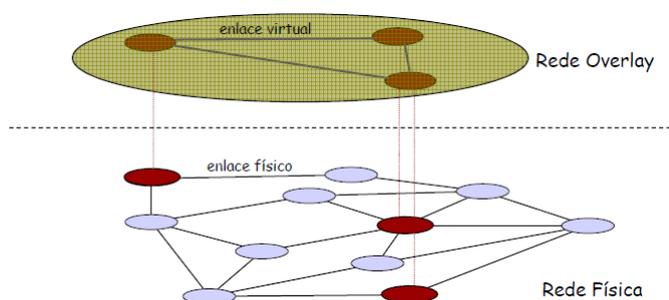
PEER-TO-PEER



- Desvantagens
 - Redundância
 - Tempo de Resposta
 - Autonomia dos pontos participantes
 - Perda de dados

PEER-TO-PEER

- Rede overlay
- Rede virtual, formada pela interconexão de pontos sobre uma infra-estrutura de uma rede física



PEER-TO-PEER

BD Distribuídos

- Pequeno número de fontes
- Nós controlados na rede
- Dados consistentes
 - Coordenação
- Dados estruturados (e.g., modelo relacional)
- Transações
- Restrições de integridade
- Consultas complexas
- Esquemas criados pelo administrador
- Topologia relativamente fixa

Sistemas P2P

- Grande número de fontes
- Nós entram e saem da rede
- Dados não confiáveis
 - Autonomia
- Dados semi-estruturados (e.g. XML)
- Sem transações
- Sem restrições de integridade
- Consultas simples
- Esquemas criados pelo usuário
- Rede imprevisível

INFRA-ESTRUTURA

INFRA-ESTRUTURA

- **Topologia geralmente diferente da rede física**
- **Algoritmos focados em otimizar a comunicação entre os pontos da rede**
- **Tipos**
 - Topologia Pura (Estruturada e Não Estruturada)
 - Topologia Híbrida
 - Topologia Super-peer

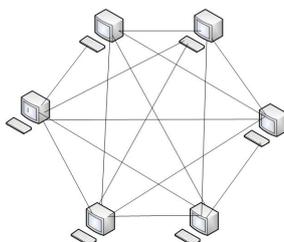
INFRA-ESTRUTURA



TOPOLOGIA PURA

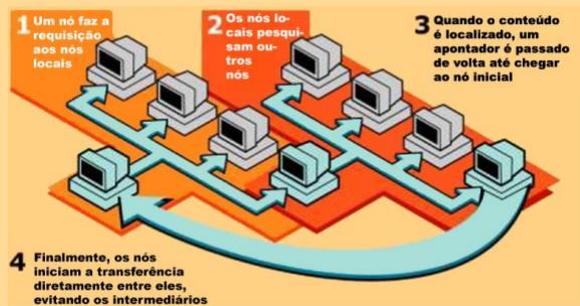
(DESCENTRALIZADO)

- Não há distinções de funções
- Cada ponto pode atuar como servidor ou cliente
- Os pontos não possuem informações sobre outros pontos da rede



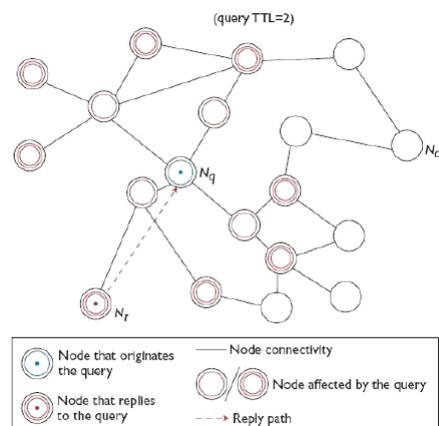
TOPOLOGIA PURA - NÃO ESTRUTURADA

- *Flooding*
- Time to Live (TTL)
- *Gossip Protocol*
- Alta escalabilidade
- Boa cobertura e alta precisão



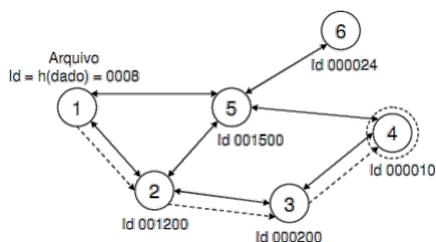
TOPOLOGIA PURA - NÃO ESTRUTURADA

GNUTELLA



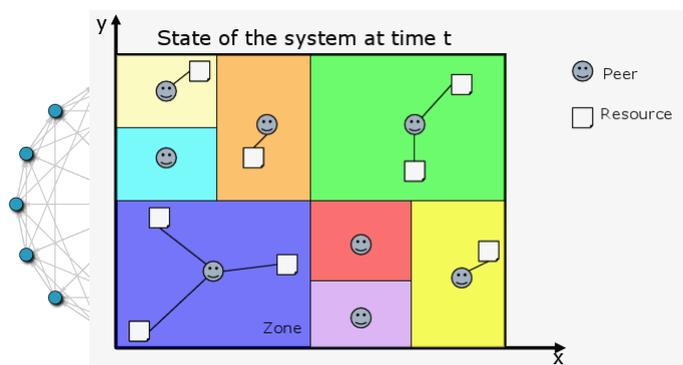
TOPOLOGIA PURA - ESTRUTURADA

- Construído de forma determinística
- DHT para indexação e localização de dados na rede
- Pontos e Dados recebem identificadores
- Tabela Hash para dividir o espaço de busca
- Reduz número de acesso aos pontos e quantidade de informações de roteamento



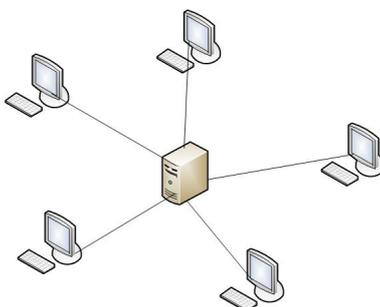
TOPOLOGIA PURA - ESTRUTURADA

CAORD



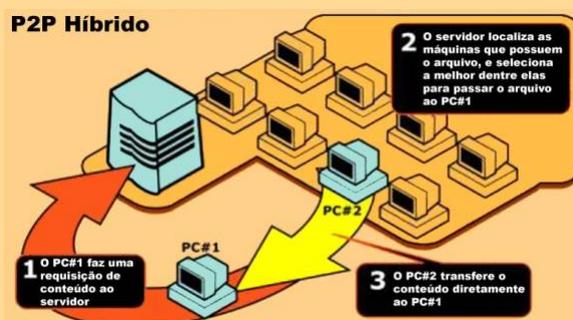
TOPOLOGIA HÍBRIDA

- Servidores armazenam informações de controle
- Os pontos devem registrar seus serviços disponíveis
- Minimiza problemas de gerenciamento
- Existência de um ponto de falha



TOPOLOGIA HÍBRIDA

NAPSTER



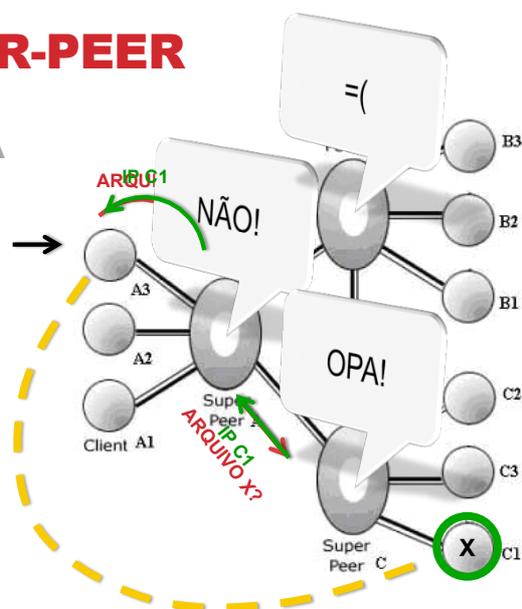
SUPER-PEER

- Super pontos assumem funções similares ao do servidor na topologia híbrida:
 - Indexação
 - Processamento de consultas
 - Controle de acesso
 - Gerenciamento de Metadados
- Geralmente agrupam pontos com semântica comum (Cluster)
- Diminuem o tempo de resposta da consulta

SUPER-PEER

KAZAA

Q(x) →



PDMS

PEER DATA MANAGEMENT SYSTEM

PDMS

- Cenário:
 - Distribuição massiva dos sistemas atuais
 - Sites geograficamente distribuídos
 - Heterogeneidade
 - Autonomia
 - Volatilidade



PDMS

- Quantidade excessiva de pontos
- Cada ponto representa uma fonte de dados e disponibiliza seus dados através de esquemas exportados
- As consultas podem ser formuladas em qualquer ponto do sistema

PDMS

- Propriedades incorporadas
 - Autonomia
 - Expressividade da consulta
 - Eficiência
 - Qualidade do serviço
 - A tolerância a falhas
 - Segurança

PDMS

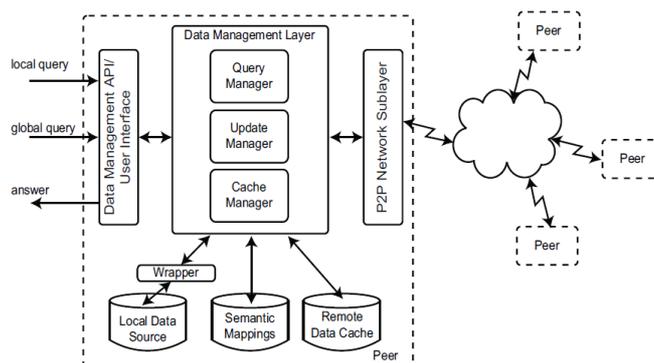
- Arquitetura
 - Uma interface usada para enviar as consultas
 - Uma camada de gerenciamento de dados que manipula processamento de consulta e informações de meta-dados
 - Uma infra-estrutura P2P, que é composto da subcamada de rede P2P e da própria rede P2P

PDMS

- Arquitetura
 - As consultas são processadas através da interface
 - Dados são manipulados pela camada de gerenciamento
 - Dados são armazenados localmente ou globalmente no sistema
 - O wrapper pode encapsular dados, consultas, ou incompatibilidade entre o BD local e os dados da camada de gerenciamento

PDMS

- Arquitetura



MAPEAMENTO DE ESQUEMA

- **Por que definir mapeamentos e correspondências?**
- **Consulta entre pontos com diferentes esquemas**
- **Mapeamento em Pares**
- **Mapeamento por Acordo Comum**
 - Common Schema Description (CSD)
- **Outras Abordagens:**
 - Mapeamento baseado em Técnicas de Aprendizado de Máquina
 - Mapeamento usando Técnicas de Recuperação de Informação

DÚVIDAS?

REFERÊNCIAS

- http://www.cin.ufpe.br/~speed/Proposta_CESP_Banca.pdf
- <http://pt.scribd.com/doc/8957836/TCC-Estudo-de-Aplicacoes-Distribuidas-P2P>
- <http://www.dct.ufms.br/gestor/titan.php?target=openFile&fileId=477>
- http://pt.wikipedia.org/wiki/P2P#Redes_P2P_estruturadas_e_n.C3.A3o-estruturadas
- [http://en.wikipedia.org/wiki/Chord_\(peer-to-peer\)](http://en.wikipedia.org/wiki/Chord_(peer-to-peer))
- ÖZSU, M. Tamer; VALDURIEZ, Patrick. *Principles of distributed database systems*. Ontario: Springer, 2011.