

Formação de DBAs SQL Server 2008

Parte 8: Banco de Dados Distribuído

Computação Distribuída



- Um grupo de elementos autônomos de processamento (não necessariamente homogêneos) que estão interconectados por uma rede de computadores e que cooperam na realização de tarefas a eles atribuídas

O que está distribuído?



- Lógica de Processamento
- Funções
- Dados
- Controle

O que é um SBDD?

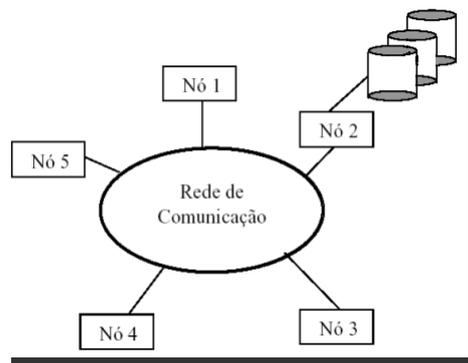


- Base de Dados Distribuída - BDD
 - É uma coleção de diversas bases de dados, interligadas logicamente através de uma rede de computadores
- Sistema de Gerência de Base de Dados Distribuídas - SGBDD
 - É o sistema de software que possibilita a gerência da base de dados distribuída e torna a distribuição transparente para o usuário
- Sistema de Banco de Dados Distribuídas - SBDD
 - É a combinação das bases com o sistema

O que não é um SBDD?



- Sistema de Banco de Dados que reside em um dos nós da rede de computadores

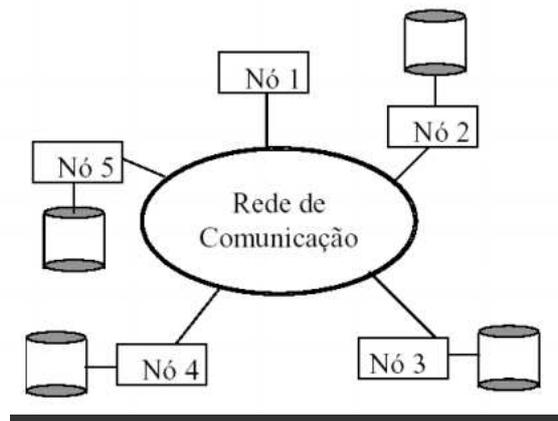


Principais Características



- Dados armazenados em diversos locais (ou nós)
- Processadores dos nós interconectados através de uma rede de computadores

Um Ambiente de SBDD



Aplicações



- Companhias Aéreas
- Redes de Lojas
- Cadeias de Hotéis
- Qualquer organização que possua uma estrutura descentralizada

Vantagens de SBDDs



- Transparência na Gerência dos Dados Distribuídos, Fragmentados e Replicados
- Confiabilidade através de Transações Distribuídas
- Aumento de Desempenho
- Facilidade de Expansão

Transparência



- Transparência é a separação entre a semântica de alto nível de um sistema e seus detalhes de implementação
- A questão fundamental é prover Independência de dados no ambiente distribuído
- Desta forma, os usuários do banco de dados enxergariam uma única imagem da base de dados logicamente integrada, embora ela estivesse fisicamente distribuída

Transparência - níveis



- Transparência da rede
- Transparência da replicação
- Transparência da fragmentação

Exemplo

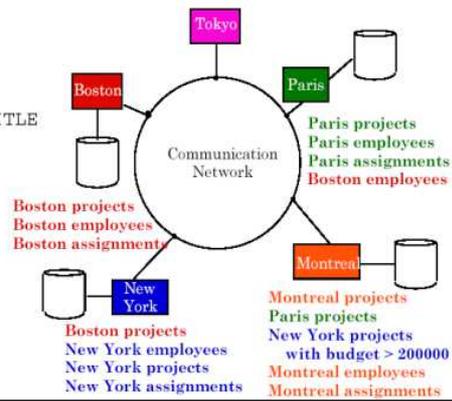


EMP			ASG			
ENO	ENAME	TITLE	ENO	PNO	RESP	DUR
E1	J. Doe	Elect. Eng.	E1	P1	Manager	12
E2	M. Smith	Syst. Anal.	E2	P1	Analyst	24
E3	A. Lee	Mech. Eng.	E2	P2	Analyst	6
E4	J. Miller	Programmer	E3	P3	Consultant	10
E5	B. Casey	Syst. Anal.	E3	P4	Engineer	48
E6	L. Chu	Elect. Eng.	E4	P2	Programmer	18
E7	R. Davis	Mech. Eng.	E5	P2	Manager	24
E8	J. Jones	Syst. Anal.	E6	P4	Manager	48
			E7	P3	Engineer	36
			E7	P5	Engineer	23
			E8	P3	Manager	40

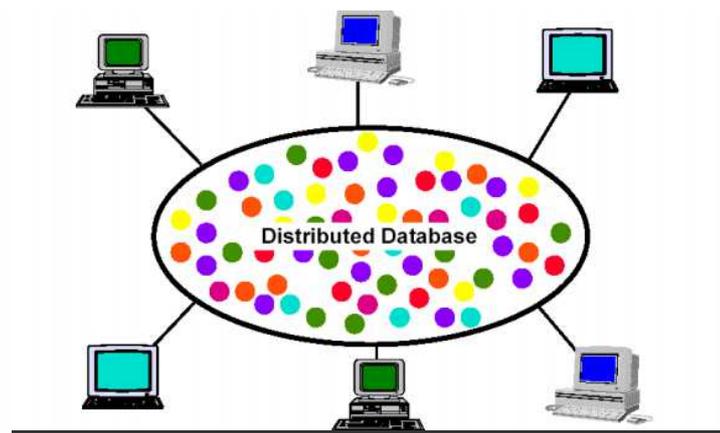
PROJ			PAY	
PNO	PNAME	BUDGET	TITLE	SAL
P1	Instrumentation	150000	Elect. Eng.	40000
P2	Database Develop	135000	Syst. Anal.	34000
P3	CAD/CAM	250000	Mech. Eng.	27000
P4	Maintenance	310000	Programmer	24000

Acesso Transparente

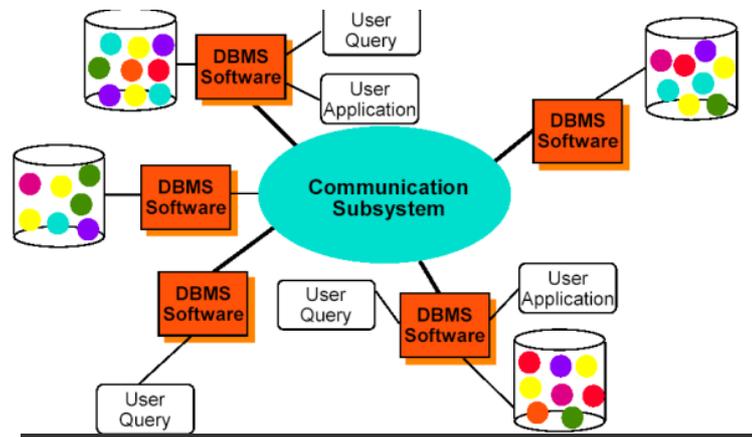
```
SELECT ENAME, SAL
FROM EMP, ASG, PAY
WHERE DUR > 12
AND EMP.ENO = ASG.ENO
AND PAY.TITLE = EMP.TITLE
```



Base Distribuída Visão do usuário



SBDD - realidade



Confiabilidade

- Espera-se que os SBDDs ofereçam confiabilidade por trabalharem com componentes replicados, eliminando assim pontos únicos de falha

Potencial aumento de desempenho



- Proximidade dos dados de seus pontos de uso
- Execução Paralela
 - Paralelismo entre consultas

Para tirar proveito do paralelismo



- Ter os dados distribuídos pelos processadores envolvidos no processamento paralelo
- Como tratar atualizações?
 - Atualizações sobre dados replicados implica na implementação de controle de concorrência distribuído e protocolos de finalização (commit)

Expansão do Sistema



- Facilidade: aumentando-se a capacidade de processamento e armazenamento da rede
- Custo: custa muito menos formar um sistema usando computadores menores com capacidade equivalente a uma única máquina de grande porte

Desvantagens



- Complexidade: problemas não resolvidos
- Custo: replicação da força de trabalho
- Distribuição de controle: sincronização e coordenação
- Segurança: problemas de segurança em uma rede de computadores

Principais áreas de problemas



- Projeto de banco de dados distribuídos
 - Separação do banco de dados em fragmentos e a distribuição ótima desses fragmentos
- Processamento distribuído de consultas
 - Decidir um estratégia para a execução das consultas através da rede da maneira mais econômica

Principais áreas de problemas



- Controle distribuído da concorrência
 - Sincronização de acessos para que a integridade do BD seja mantida
- Gerenciamento distribuído de impasses (ou Deadlocks)
 - Competição entre usuários pelo acesso aos dados

Funcionalidades exigidas



- Rastreamento de dados
- Processamento de consultas distribuídas
- Gerenciamento de transações distribuídas

Funcionalidades exigidas



- Gerenciamento de dados replicados
- Recuperação de banco de dados distribuído
- Segurança
- Gerenciamento de diretório (catálogo) distribuído

SqlServer e Banco de dados distribuidos

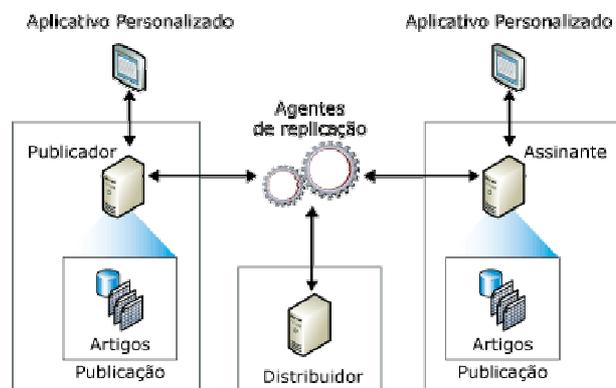


- Disponibilidade
- Segurança
- Gerenciamento
- Integridade
- Produtividade
- Desempenho
- Escalabilidade

Distribuição de dados utilizando o SQL Server



- Publicador
- Distribuidor
- Assinante



Replicação de Dados



- Cópia original dos dados
- Configuração.
- Sincronização
- Tipos
 - Snapshot Replication
 - Merge Replication e
 - Transactional Replication.

Snapshot Replication



- Distribui os dados exatamente da forma como eles aparecem em um momento específico
- **Não faz monitoramento** para atualizações dos dados
- Reproduzir os dados que sofrem poucas mudanças.
- Consome muito recurso.

Merge Replication



- Atualizações em qualquer computador serão replicadas para outro computador futuramente.
- Processo de distribuição de dados do Editor para Assinantes
- Unificar (merge) as atualizações.

Transactional Replication



- Transações, declarações INSERT, UPDATE ou DELETE, executados em um computador são replicados para outro computador
- A replicação transacional é útil quando: Mudanças incrementais precisam ser propagadas para assinantes assim que elas ocorrem.

Microsoft Distributed Transaction Coordinator



- O Microsoft Distributed Transaction Coordinator (MS DTS) é um serviço do SQL Server que coordena as transações entre os servidores através de uma rede de sistemas baseados no Microsoft Windows e pode atualizar dados em mais de um SQL Server ou origem de dados OLE DB.

Two-Phase Commit

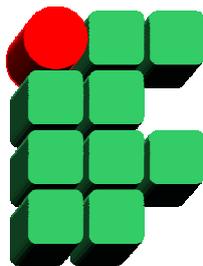


- A efetivação em duas fases é um protocolo para tornar as transações atômicas e duráveis, quando mais de uma origem de dados está envolvida na transação.

Efetuar uma Transação Distribuída



- BEGIN DISTRIBUTED TRANSACTION
- Chama automaticamente o MS DTC
- COMMIT ou ROLLBACK



Formação de DBAs SQL Server 2008
Parte 8: Banco de Dados Distribuído

V1.0