

Antônio Junior  
Vinicius Cezar

# Interoperabilidade de Banco de Dados

## Roteiro

- Alguns Conceitos
- Projeto de esquemas conceituais globais
  - Conversão de esquemas
  - Integração de esquemas
- Processamento de Consultas
- Gerenciamento de Transações
- Orientação a objetos na interoperabilidade de banco de dados

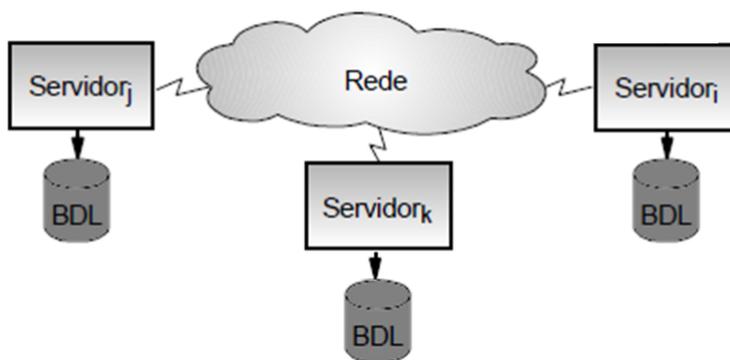
## Banco de Dados Distribuídos

- Banco de Dados Distribuídos (BDD) se refere a uma coleção de vários bancos de dados logicamente inter-relacionados e distribuídos por uma rede de computadores.
- Tipos:
  - Homogêneos.
  - Heterogêneos.

## SGBDs Distribuídos Homogêneos

- BD Distribuído Homogêneos :
  - Banco de dados armazenado em vários servidores
  - Conectados por uma rede de comunicação de dados
  - Abordagem de projeto: *Top-Down*
- SGBD Distribuído Homogêneos :
  - Estende as técnicas de armazenamento de dados, processamento de consultas e gerência de transações
  - Permitir a implementação de BDDs

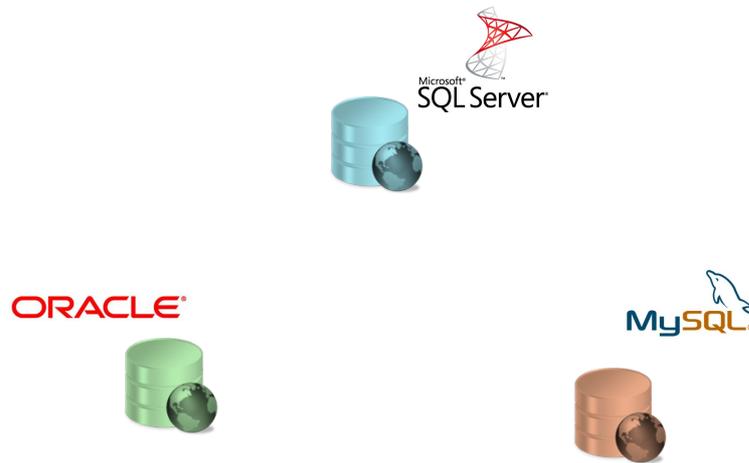
## SGBDs Distribuídos Homogêneos



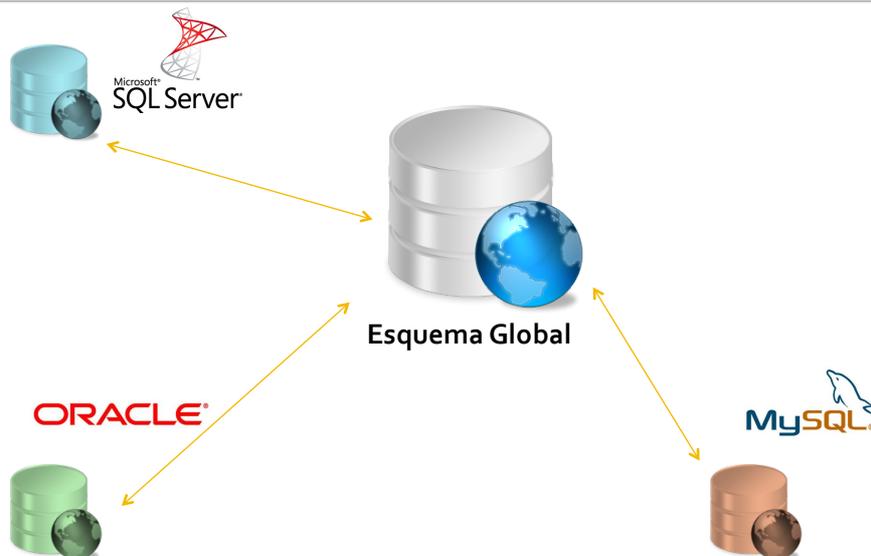
## O que são Bancos de dados Heterogêneos?

- Bancos que entre si:
  - Utilizam modelos de dados distintos
  - Representam os dados de uma maneira diferente
  - Rodam em diferentes plataformas de hardware
  - Normalmente são gerenciados por software diferentes.
- Abordagem de projeto: *Bottom-Up*

## O que são Bancos de dados Heterogêneos?



## Banco de Dados Múltiplos Distribuídos



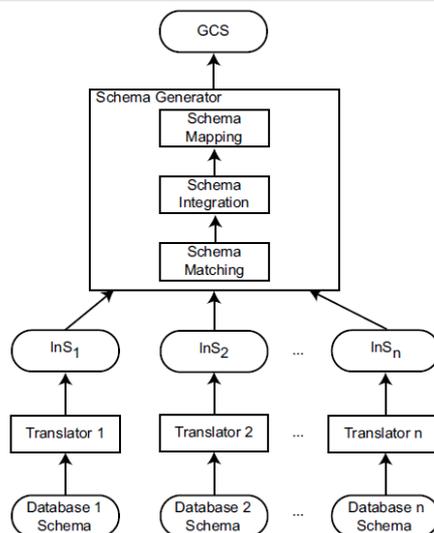
## O que é Interoperabilidade?

- É a capacidade de um sistema se comunicar de forma transparente (ou o mais próximo disso) com outros sistemas (semelhante ou não).

## Integração de Banco de Dados

- Processo pelo qual as informações de banco de dados participantes podem ser integradas conceitualmente para formar uma única definição coesa de um banco de dados múltiplo.
- Ocorre em duas etapas:
  - Conversão de Esquema
  - Integração de Esquema

## Processo de Integração de Banco de Dados



## Conversão de Esquemas

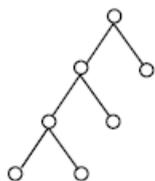
- Tarefa de mapeamento de um esquema para outro
  - Banco de dados componentes heterogêneos
- Exige a especificação de um modelo de dados de destino
- Cada esquema componente é convertido numa representação intermediária comum

## Integração de Esquemas

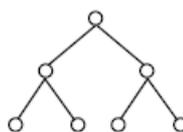
- Gerar o Esquema Conceitual Global integrando os esquemas intermediários
- Mecanismos de metodologias de Integração
  - Binários
  - N-ários

## Metodologias de Integração

- Métodos de integração binários
  - Envolve a manipulação de dois esquemas de cada vez
    - Passo a passo (escada)
    - Puramente binário



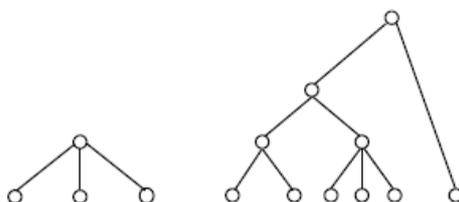
(a) Stepwise



(b) Pure binary

## Metodologias de Integração

- Métodos de integração n-ários
  - Integram mais de dois esquemas em cada iteração
    - Uma passagem
    - Iterativa



(a) One-pass

(b) Iterative

## Integração de Esquemas

- A integração envolve duas tarefas:
  - Homogeneização
    - Envolve a determinação de conflitos estruturais e semânticos de cada banco de dados componente.
  - Integração
    - Envolve a mesclagem dos esquemas de vários banco de dados para criar um esquema conceitual global.

## Homogeneização

- Resolução de conflitos de heterogeneidade semântica e estrutural.
- Heterogeneidade semântica
  - Conflitos de Nomenclatura
    - Homônimos
      - Uso de Prefixos
    - Sinônimos
      - Ontologias

## Homogeneização

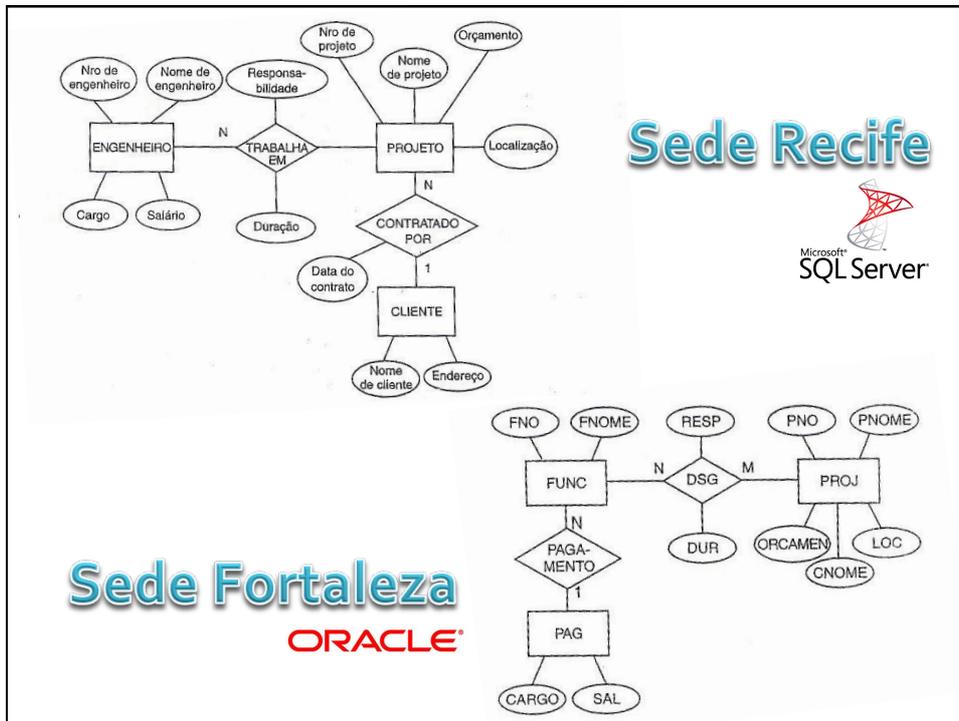
- Dois esquemas podem estar relacionados de 4 modos:
  - Idênticos
  - Subconjunto
  - Intersecção
  - Completamente distintos
- A identificação dos relacionamentos deve considerar a sua sintaxe e semântica

## Homogeneização

- Heterogeneidade estrutural
  - Tipo
    - Objeto representado como um atributo em um esquema e como uma entidade em outro.
  - Dependência
    - Diferentes modos de relacionamento são usados para representar o mesmo item (por exemplo, *m:n versus 1:n*).
  - Chave
    - Quando chaves primárias diferentes são selecionadas em esquemas distintos.
  - Comportamental
    - Implicados pelo mecanismo de modelagem

## Exemplo

- Considere o seguinte cenário:
  - Uma empresa de construção civil com sede em Recife, adquire uma outra localizada em Fortaleza.
  - Cada sede possui sua própria bases de dados em diferentes SGBDs.
  - Deseja-se ter uma visão global dos dados.

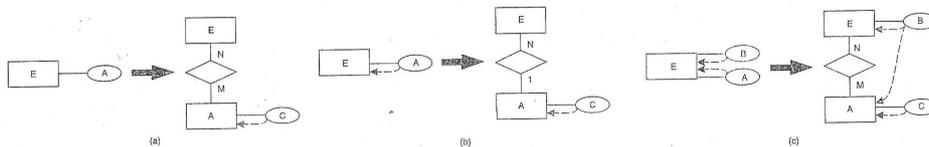


## Homogeneização

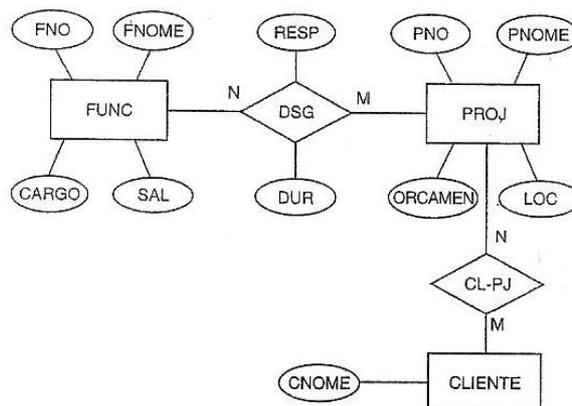
Esquema Recife	Esquema Fortaleza
<b>ENGENHEIRO</b>	<b>FUNC</b>
Nro de engenheiro	FNO
Nome de engenheiro	FNOOME
Salário	SAL
<b>TRABALHA EM</b>	<b>DSG</b>
Responsabilidade	RESP
Duração	DUR
<b>PROJETOS</b>	<b>PROJ</b>
Nro de projeto	PNO
Nome de projeto	PNOOME
Localização	LOC

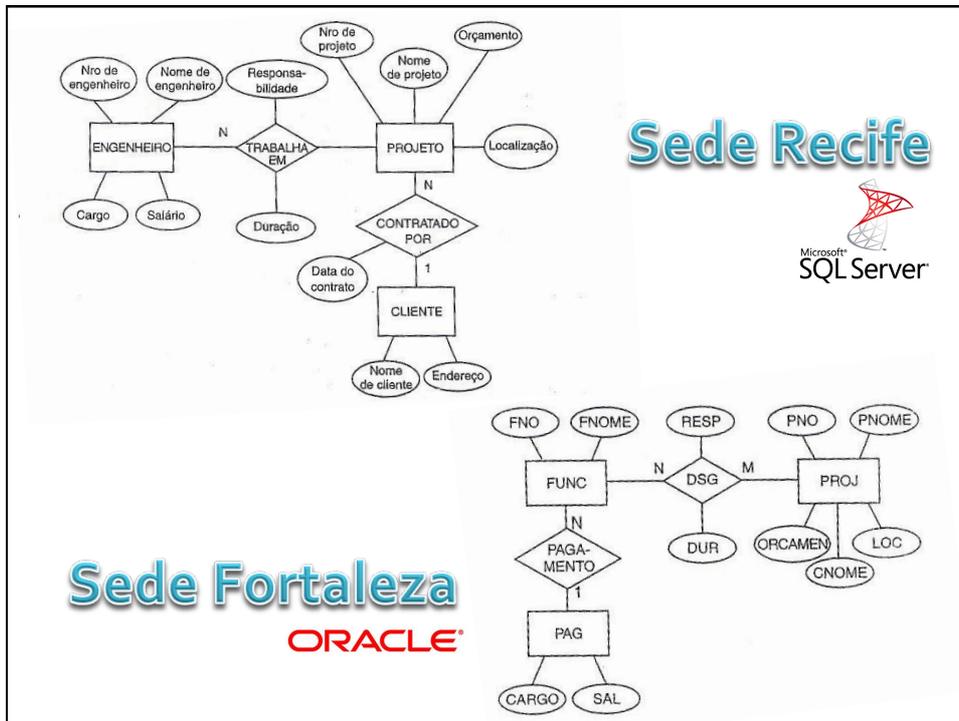
## Homogeneização

- A transformação de entidades / atributos / relacionamentos entre si é um modo de manipular conflitos estruturais.

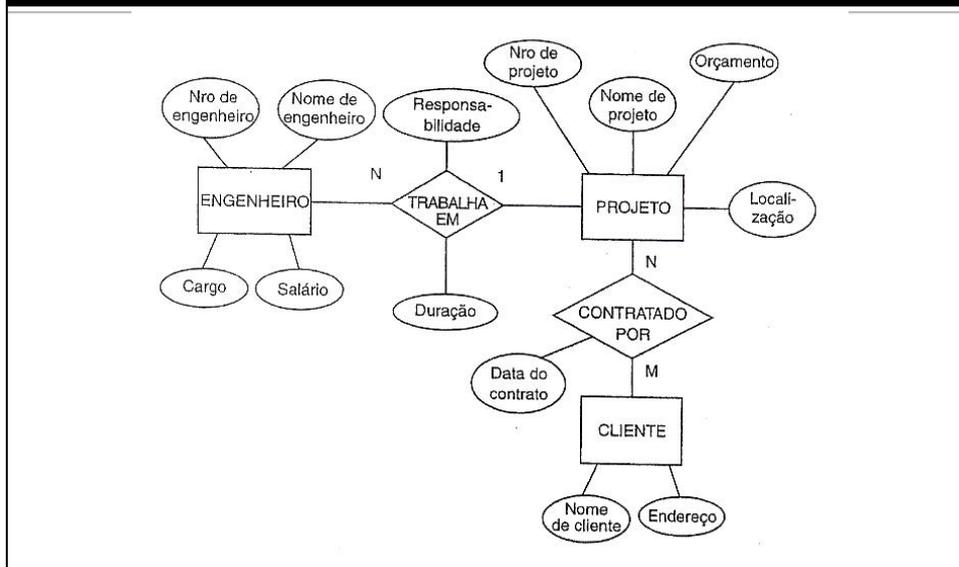


## Transformação atributos / entidades





## Esquema Global



## Integração

- Mesclagem dos sistemas intermediários em um banco de dados único e reestruturação para criar o melhor esquema integrado.
- Preserva as informações contidas nos esquemas participantes.

## Integração

- Podemos definir três dimensões de mesclagem e reestruturação:
  - Completeza
    - Todas as informações de todos os esquemas estão integradas ao esquema comum
  - Minimalidade
    - Diz respeito à presença de informações redundantes de relacionamentos em um esquema integrado.
  - Inteligibilidade
    - Facilidade de compreensão do esquema pós-integração.

## Processamento de Consultas

- Em SGBDs Distribuídos:
  - Decomposição de Consultas
    - Decompõe a consulta distribuída em uma consulta algébrica sobre relações globais*
  - Localização dos Dados
    - Localizar os dados com uso de informações de distribuição dos dados*
  - Otimização Global
    - Encontrar uma estratégia de execução global ótima
  - Otimização Local
    - Otimizada com o esquema local

## Processamento de Consultas

- SGBD múltiplo é uma camada de software que funciona sobre os SGBDs componentes.
- Cada SGBD componente tem seu próprio processador de consultas.

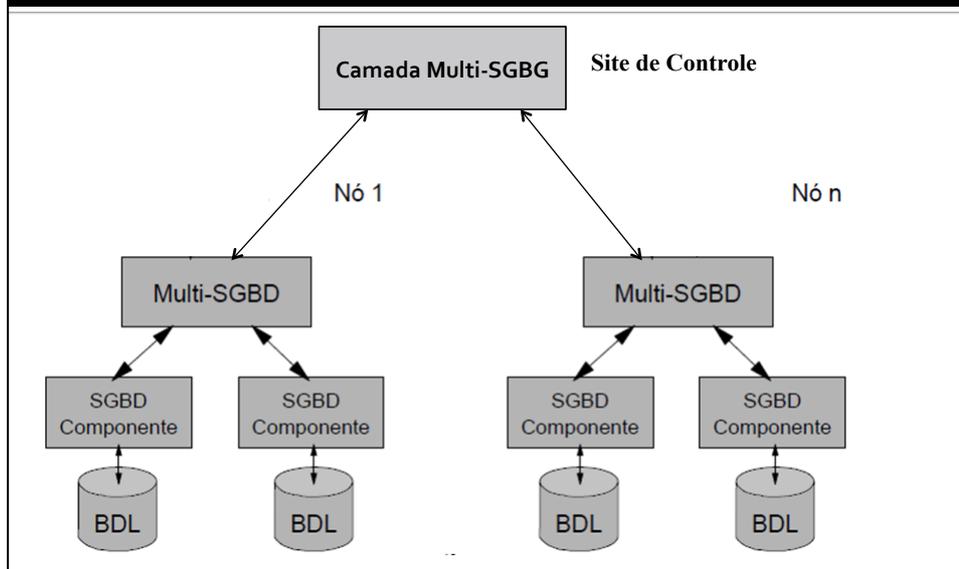
## Processamento de Consultas

- O processamento de consultas em SGBD Múltiplo é mais complexo do que em um SGBD distribuído.
  - Capacidade SGBDs componentes podem ser diferentes
  - Custo de processar consultas pode ser diferente
  - Movimentação de dados entre SGBDs
  - Capacidade de otimização local de cada SGBD diferentes

## Processamento de Consultas

- A autonomia dos SGBDs componentes
- Os dados são distribuídos não apenas pelos sites, mas também por vários bancos de dados cada qual em um SGBD autônomo.
  - Exige técnicas de processamento de consultas tolerantes à indisponibilidade do sistema.
  - Dificuldade em aplicar estratégias de otimização de consultas.

## Arquitetura de um SGBD múltiplo distribuído



## Exemplo Consultas

### Consulta Global:

```
select cli.[Nome de Cliente], cli.[Endereco]
from Cliente cli
inner join [Contratado Por] con on cli.[Nome de Cliente] = con.[Nome de Cliente]
inner join Projeto pro on con.[Nro de Projeto] = pro.[Nro Projeto]
where pro.[Orçamento] > 100.000,00
```



### Sede Recife:

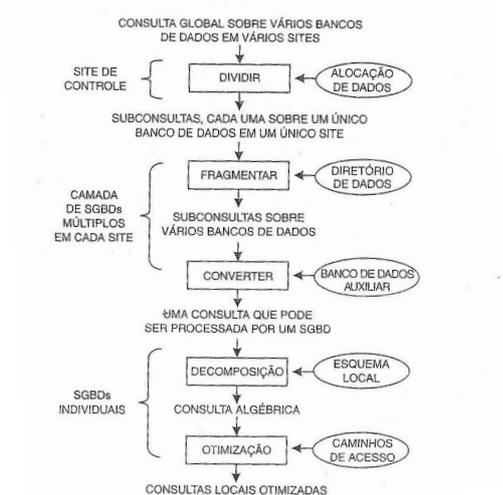
```
select cli.[Nome de Cliente], cli.[Endereco]
from Cliente cli
inner join Projeto pro on pro.[Nome de Cliente] = cli.[Nome de Cliente]
where pro.[Orçamento] > 100.000,00
```



### Sede Fortaleza:

```
select [CNOME], Null as Endereco
from PPROJ
where [ORCAMENT] > 100.000,00
```

## Camadas de Processamento de consulta



## Otimização de consultas

- Semelhante ao processamento em SGBDs homogêneos em alguns aspectos.
- Podem se basear em:
  - Heurística
    - Decompor a consulta global nas menores subconsultas possíveis.
      - Menos trabalho para o processador de consultas componente
    - Decompor a consulta global nas maiores subconsultas possíveis.
      - Menos trabalho para o processador de consultas global

## Otimização de consultas

- Custo
  - Calculados baseados nos SGBDs componentes.
  - Alternativas para determinar o custo de cada componente:
    - SGBDs Componentes como caixa preta
    - Conhecimento anterior sobre os SGBDs
    - Monitoração do comportamento em tempo de execução.

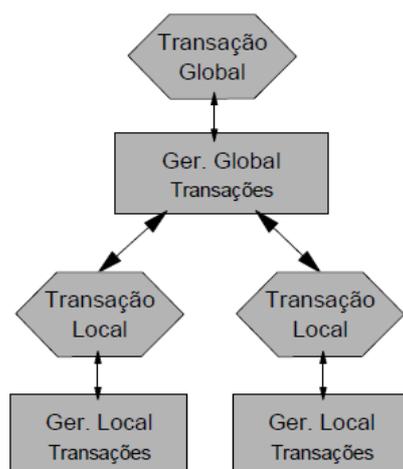
## Gerenciamento de Transações

- O desafio é permitir atualizações globais concorrentes aos bancos de dados componentes sem violar suas autonomias.
  - Gerenciador de transações
  - Controle de concorrência
  - Consolidação

## Gerenciamento de Transações

- Gerenciador de transações
  - Global (GTM)
    - Gerencia a execução de transações globais
    - Transações que atuam sobre vários bancos de dados
  - Local (LTM)
    - Gerencia a execução de transações de um único SGBD
    - Transações que atuam sobre um único banco de dados
- Uma transação global pode ser subdividida em subtransações globais.

## Gerenciamento de Transações



## Controle de Concorrência

- Não é fácil pro GTM identificar a existência de conflitos .
  - Existem outras transações sendo enviadas diretamente aos SGBDs componentes
  - Transações locais podem causar conflitos com subtransações globais.

## Exemplo Concorrência

- Considere uma transação GT para atualização de salário do engenheiro João com aumento de 15%.

- Sede Recife: GT1

```
select salario into temp1 from ENGENHEIRO where [ Nome de Engenheiro]= 'João'
if temp1 is empty then
  abort
else begin
  update ENGENHEIRO set Salario = Salario * 1.15 where Nome = 'João'
  commit
endif
```

- Sede Fortaleza: GT2

```
select Cargo into temp1 from FUNC where FNome = 'João'
if temp1 is empty then
  abort
else begin
  update pag set SAL = SAL * 1.15 where FNome = 'João'
  commit
endif
```

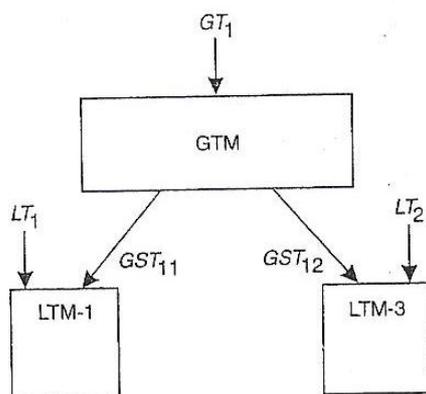
## Exemplo Concorrência

- Simultaneamente, considere que na sede de Recife alguém envia uma transação local para alteração do salário de João com um aumento 30% .

- LT1:

```
select salario into temp1 from ENGENHEIRO where [ Nome de Engenheiro]= 'João'
if temp1 is empty then
  abort
else begin
  update ENGENHEIRO set Salario = Salario * 1.3 where Nome = 'João'
  commit
endif
```

## Gerenciamento de Transações



## Gerenciamento de Transações

- Condições para transações globais executarem em segurança um SGVBD
  - Trazer a responsabilidade da execução correta das transações para o LTM.
    - Escalonamento correto de subtransações globais e transações locais.
  - Cada LTM deve manter a ordem de execução das subtransações determinadas pelo GTM.
    - O GTM é responsável por lidar com impasses globais

## Gerenciamento de Transações

- Em um SGBD distribuído o gerenciador de transações de um site pode se comunicar diretamente com os escalonadores de seu site e de outros sites.
  - Isso não ocorre com os SGBDs múltiplos
    - SGBDs componentes não sabem se comunicar diretamente entre eles

## Gerenciamento de Transações

- Modelo de Execução
  - *Abordagem de Solicitação de Serviços*
    - *A camada de SGBD múltiplo que se comunica com os SGBDs componentes por uma interface de alto nível.*
    - *Interface de Operações de Transações*
      - *SGBDs componentes permitam que o GTM envie transações*
  - O GTM terá controle mais preciso sobre a execução das transações.

## Gerenciamento de Transações

- Como lidar com o controle de concorrência em SGBDs Múltiplos distribuídos.
  - Os SGBDs pode reordenar subtransações globais
- Abordagem da bilhetagem:
  - Cada site possui um item de dado especial
  - Cada subtransação global lê o bilhete e grava o bilhete no site
  - Assim o valor do bilhete indica a ordem de cada subtransação global para o LTM

## Recuperação de Vários Bancos de dados

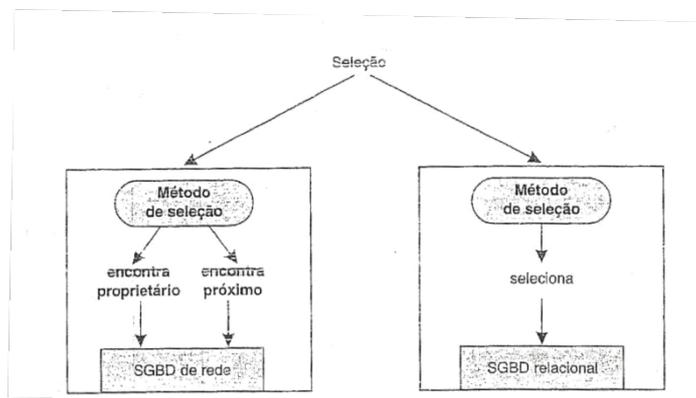
- Identificar se SGBDs componentes disponibilizam uma interface para o GTM consolidar e assegurar a atomicidade das transações.
  - Caso positivo: Utilizam o protocolo 2PC.
  - Caso negativo:
    - 1. "Pseudo 2PC": callback nas subtransacoes globais.
    - 2. Refazer a subtransação global que não se consolidou.
    - 3. Transações Compensatórios.

## Orientação a Objeto e Interoperabilidade

- Papel significativo na solução de algumas modelagens relacionadas à Interoperabilidade entre bds.
  - Modelagens que envolvam SGBDs relacionais e SGBDs de objetos o esquema conceitual deve ser orientado a objeto
- Características:
  - Encapsulamento: Permite ocultar as diferenças em interfaces e implementações de SGBDs componentes
  - Especialização: Permite a criação de tipos que abstraem as semelhanças entre as entidades de armazenadas em diferentes bancos de dados.

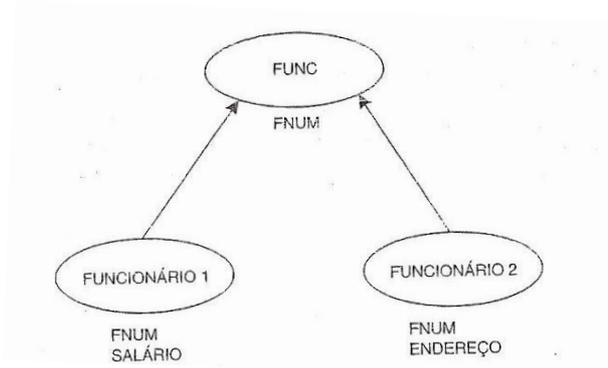
## Orientação a Objeto e Interoperabilidade

- Encapsulamento:



## Orientação a Objeto e Interoperabilidade

- Especialização:



## Orientação a Objeto e Interoperabilidade

- Além das vantagens de modelagens.
- Existem várias plataformas de computação com objetos distribuídos.
  - Permitindo que aplicativos se comuniquem facilmente uns aos outros.
  - OMA (Object Management Architecture)
    - CORBA (mecanismo de comunicação)
  - OMG (Object Manager Group)
  - DCOM (Distributed Componente Object Model)
  - OLE (Object Linking and Embedding)

■ Dúvidas?

■ Obrigado!