

FATEC SP – FACULDADE DE TECNOLOGIA DE SÃO PAULO
CURSO EM PROCESSAMENTO DE DADOS

CESAR APOLO NUMATA

BANCO DE DADOS DISTRIBUÍDOS

SÃO PAULO – SP
2012

CESAR APOLO NUMATA

BANCO DE DADOS DISTRIBUÍDOS

Monografia para Término de
Conclusão de Curso em
Processamento de Dados da
Faculdade de Tecnologia de São
Paulo.

ORIENTADOR:
PROFESSORA SANDRA
HARUMI TANAKA

SÃO PAULO – SP

2012

Folha de aprovação

CESAR APOLO NUMATA

BANCO DE DADOS DISTRIBUÍDOS

Monografia para Término de
Conclusão de Curso em
Processamento de Dados da
Faculdade de Tecnologia de São
Paulo.

Banca examinadora

ORIENTADOR: SANDRA HARUMI TANAKA

São Paulo, ____ de _____ de 2012

Dedicatória

À minha família, que me incentivou e me deu condições de adquirir conhecimento e educação.

Aos meus amigos, que me deram alegria nas horas difíceis e força de vontade para que eu sempre seguisse em direção aos meus sonhos.

Agradecimentos

A todos os que me apoiaram de alguma forma para que eu tivesse condições de realizar este trabalho.

Aos professores e aos colegas de trabalho que me deram o conhecimento para que isto fosse possível.

RESUMO

O escopo do trabalho é demonstrar uma estrutura de Banco de Dados Distribuídos, levando em conta a integridade dos dados, a segurança, e comparar a tecnologia distribuída e a centralizada. As redes de computadores surgiram com a necessidade de se utilizar computadores em locais diferentes. Ao se conectar vários computadores em rede, tem-se uma melhoria na capacidade de processamento. Às vezes em um banco de dados acontece que certos dados interessem a somente um lugar da rede e seria razoável que esses dados ficassem guardados nesse lugar. É da descentralização do armazenamento desses dados que resulta em um Banco de Dados Distribuídos (BDD). Para dados utilizados com frequência por dois ou mais locais, cria-se cópias para cada lugar. Existem casos em que os dados armazenados estão de formas diferentes em cada lugar. Dessa maneira, obriga-se a fazer uma conversão nas informações antes de serem enviadas para os outros locais da rede. A base de dados pode ser mantida em um único local ou então em cada um dos lugares ligados pela rede. A principal vantagem de se utilizar um sistema de dados distribuídos é compartilhar e acessar dados de uma maneira segura e eficiente. A desvantagem principal é o acréscimo de complexidade exigida para assegurar coordenação própria entre os locais. Um dos principais cuidados é com a sincronização dos bancos nos locais da rede, que deve garantir que um determinado local tenha todas as informações dos demais locais.

Abstract

The scope of work is to demonstrate a framework of Distributed Database, considering data integrity, security, and to compare centralized and distributed technology. The networks of computers brought the need of using computers at different locations. When you connect multiple computers on a network you have an improvement in processing capability. Sometimes in a database, it happens that certain data interest to only one place in the network and it is reasonable that the data stored in that place stay there. It is the decentralization of storage of data that results in a Distributed Database (DDB). For data used frequently by two or more locations, it creates copies for each place. There are cases in which data are stored in different ways at different places. Thus, it undertakes to do a conversion on the information before being sent to other network locations. The database can be kept in one location or locations in each of the connected network. The main advantage of using a system of distributed data is to share and access data in a safe and efficient manner. The main disadvantage is the additional complexity required to ensure coordination between the local itself. One of the primary precautions is the synchronization of local banks in the network, which must ensures that a particular location has all the information from other sites.

SUMÁRIO

1. INTRODUÇÃO.....	9
2. SISTEMAS DE BANCO DE DADOS	10
2.1 Tipos de SBDs Distribuídos	13
2.2 Vantagens dos SBDDs	15
2.3 Desvantagens dos SBDDs	15
2.4 Áreas de Problemas nos Bancos de Dados Distribuídos	17
2.5 Conclusão.....	19
3. ESTUDO DE CASO.....	20
3.1 Justificativas e necessidades.....	20
3.2 Especificação do problema.....	21
3.3 Modelo de dados.....	21
3.4 Projeto.....	23
3.4.2 Arquitetura.....	24
3.4.1 Projeto do banco de dados.....	25
3.5 Implementação.....	27
3.6 Resultados.....	32
3.6.1 Vantagens.....	33
3.6.2 Desvantagens.....	37
4. CONCLUSÃO	39
REFERÊNCIAS BIBLIOGRÁFICAS	40

1. INTRODUÇÃO

Um sistema de bancos de dados distribuídos é uma junção de redes de computadores com os sistemas de bancos de dados. Os sistemas de bancos de dados mostraram, com o passar do tempo, ser algo além de um simples aplicativo para gerenciar dados de forma centralizada, como também, um aplicativo para gerenciar dados de forma distribuída por meio de uma de uma rede de computadores.

O sistema de bancos de dados distribuídos nada mais é do que um conjunto de vários bancos de dados logicamente inter-relacionados, fisicamente separados em diferentes lugares, dispersos geograficamente, distribuídos por uma rede de computadores.

O objetivo deste trabalho é mostrar o funcionamento de uma estrutura de Bancos de Dados Distribuídos, considerando os aspectos de segurança, concorrência e integridade dos dados.

O capítulo 2 apresenta as principais características de um sistema de bancos de dados, os tipos de sistema de gerenciamento de banco de dados distribuídos, vantagens e desvantagens da distribuição dos dados e os principais problemas que podem existir com bancos de dados distribuídos.

No capítulo 3 é apresentado um estudo de caso de mostrando a implantação de um banco de dados distribuído como solução para gerenciamento de dados de uma empresa de grande porte, os resultados, vantagens e desvantagens na utilização de um sistema de gerenciamento de banco de dados distribuídos para esta empresa.

No capítulo 4 é feita a conclusão do trabalho desenvolvido.

2. SISTEMAS DE BANCO DE DADOS

Um sistema de banco de dados (SBD) é basicamente um sistema de manutenção de registros por computador. O objetivo dos SBDs é manter os registros e disponibilizar quando solicitados. Estes registros ficam gravados em arquivos com uma estrutura de um banco de dados. (DATE, 2000).

Um banco de dados é uma grande quantidade de informações armazenadas de forma estruturada em um computador. Para ter acesso as informações gravadas dentro do banco de dados, é necessário um sistema de gerenciamento de banco de dados.

Um sistema de gerenciamento de bancos de dados (SGBD) consiste em uma coleção de dados inter-relacionados e em um conjunto de programas para acessá-los. O objetivo de um SGBD é criar um ambiente adaptado para a recuperação e armazenamento das informações do banco de dados. (KORTH, 1995)

Os sistemas de banco de dados existem de duas formas: os sistemas de bancos de dados centralizados e os sistemas de bancos de dados descentralizados ou distribuídos. Em um banco de dados centralizado, todos os integrantes do sistema ficam em um mesmo computador ou *site* (um local onde algo se baseia).

Os integrantes dos sistemas são os dispositivos para armazenamento, o *software* (aplicativos) de gerenciamento de banco de dados e todos os dados. Este tipo de banco de dados centralizado pode ser acessado por vários terminais diferentes conectados ao *site*, porém o SGBD e todos os dados estão situados em um único *site*.

Com o progresso, melhoria, nas áreas de bancos de dados e tecnologia de processamento em rede, a distribuição de sistemas de computadores em vários *sites* conectados em rede começaram a surgir. Começaram então a estudar uma nova forma de se trabalhar com dados de forma distribuída.

Pode-se definir como bancos de dados distribuídos (BDD), como um conjunto de vários bancos de dados logicamente inter-relacionados, fisicamente separados em diferentes *sites*, dispersos geograficamente, distribuídos por uma rede de computadores. Um sistema de gerenciamento de bancos de dados distribuídos (SGBD Distribuídos) é conceituado como um sistema que possibilita o gerenciamento dos bancos de dados distribuídos e que transforma a distribuição dos dados de forma clara para os usuários, como se fosse um sistema centralizado. (CASANOVA, 2012).

O termo “sistema de bancos de dados distribuídos” (SBDD) atribui ao conjunto entre os bancos de dados distribuídos e o SGBD distribuído. As figuras 2.1 e 2.2 ilustram a diferença de uma arquitetura de um SBDD de uma arquitetura centralizada.

A figura 2.1 mostra que para cada trabalho realizado por um *site*, as informações têm de passar pela rede e chegar ao *site* onde está o banco de dados, podendo assim fazer com que o serviço seja lento, pois a rede pode ficar congestionada.

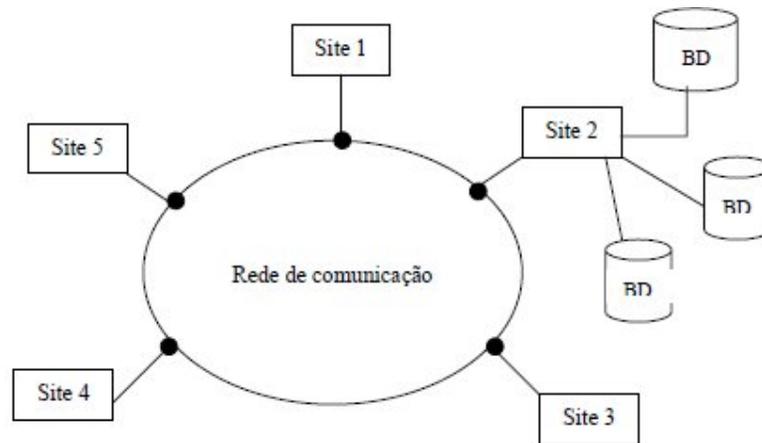


Figura 2.1 - Banco de dados central em uma rede
Adaptado de: MATTOSO, 2012

A figura 2.2 apresenta vários bancos de dados entre diferentes *sites*, e cada um dos *sites* possui um SGBD, onde trabalham de forma local. E por intermédio da rede é possível a comunicação entre eles. Os SGBDs podem ou diferentes em todos os *sites*.

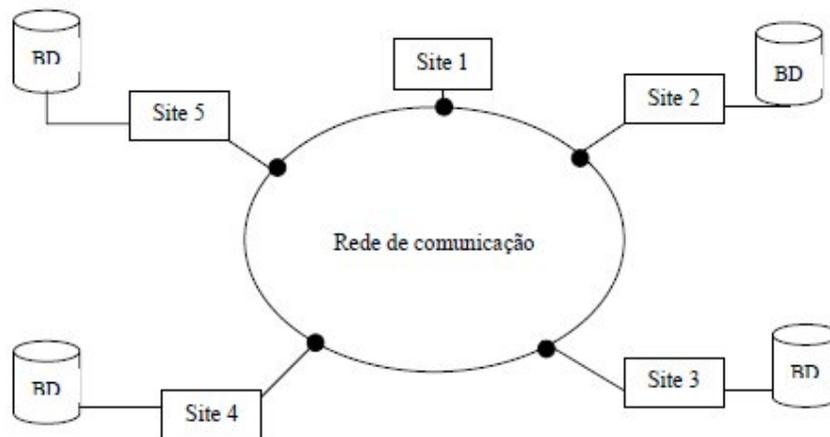


Figura 2.2 -Ambiente de SBDD
Adaptado de: MATTOSO, 2012

2.1. TIPOS DE SGBDs DISTRIBUÍDOS

Existem dois tipos de SBDDs, o SBDD homogêneo e o SBDD heterogêneo. No SBDD homogêneo, ilustrado pela figura 2.3, os SGBDs locais são semelhantes em todos os *sites*, ou a estrutura do banco de dados é igual em todos os *sites*. Já no SBDD heterogêneo, ilustrado pela figura 2.4, pode existir dois ou mais SGBDs diferentes espalhados entre os *sites* ou a estrutura do banco de dados também é diferente. Podem existir diversos nós e cada nó pode conter um ou mais SGBDs locais. (KORTH, 1995)

Os SGBDs homogêneos consistem em SGBDs locais que ofereçam interfaces idênticas ou pelo menos da mesma família (mesma empresa ou versões diferentes) e que forneçam os mesmos serviços aos usuários em diferentes sites. (CASANOVA, 2012)

Os SGBDs distribuídos heterogêneos existem quando há necessidade de integrar sistemas já existentes. (CASANOVA, 2012)

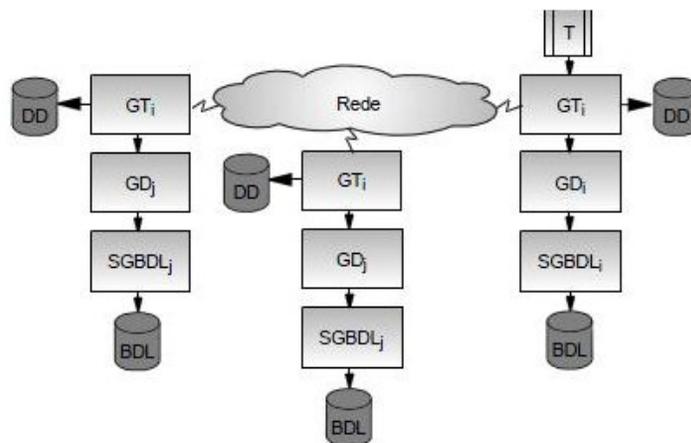


Figura 2.3 SGBDs Homogêneos
Fonte: CASANOVA, 2012

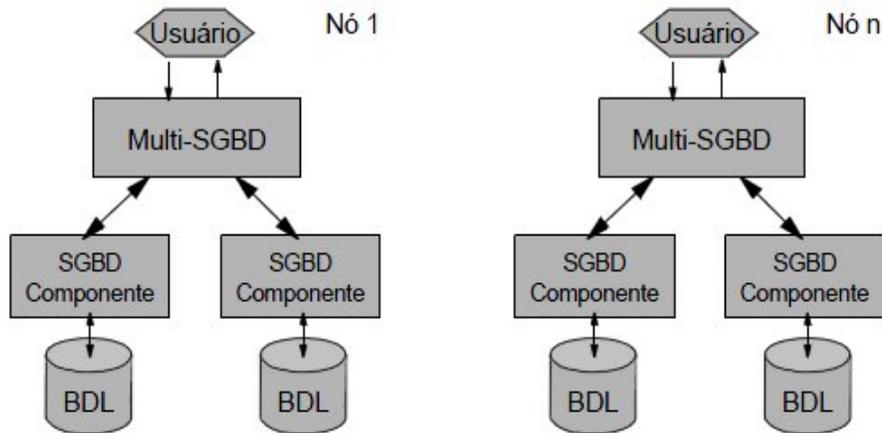


Figura 2.4 SGBDs Heterogêneos
Adaptado de: CASANOVA, 2012

Quando os bancos de dados são heterogêneos (tanto no SGBD quanto na estrutura), é necessário fazer uma conversão entre os SBDDs. Em geral, esse mecanismo de conversão envolve uma forma para facilitar a conversão de dados, assim como programas para converter instruções de manipulação de dados. Normalmente, essa heterogeneidade é introduzida quando se está construindo um SGBD distribuído a partir de vários SGBDs autônomos e centralizados. De fato, tais devem ser considerados complementares em relação aos SGBDs distribuídos.

Existem dois tipos de distribuição dos dados no SBDD: particionar e replicar. Particionando, o banco de dados se separa em diversas partições disjuntas, e cada partição é colocada em um *site*. Replicando, pode ser totalmente ou parcialmente replicado. Totalmente replicado, cópias do banco de dados inteiro são armazenados em cada *site*. E parcialmente replicado, cada partição do banco de dados é armazenado em mais de um *site*. (FACHIN, 2012)

2.2. VANTAGENS DOS SBDDs

São consideradas algumas características como suas principais vantagens, sendo elas:

- Transparência na gerência dos dados distribuídos: separação de um sistema de alto nível e os detalhes de implementação. O objetivo é que se tenha uma independência de dados. A transparência pode ser dividida em três níveis: Transparência da rede, Transparência da replicação e Transparência da fragmentação (que é dividida em híbrida, vertical e horizontal);
- Maior confiabilidade na transação de dados distribuídos: o SBDD funciona de acordo com o projeto;
- Maior desempenho em consultas: podem ser executadas em paralelo as subconsultas;
- Expansão do sistema: facilidade em adaptar ao crescimento da base de dados;
- É possível o compartilhamento de recursos e dados;
- Utilização por maior quantidade de usuários ao mesmo tempo. (O'BRIEN, 2008)

2.3. DESVANTAGENS DOS SBDDs

Existem algumas desvantagens também para se implantar um sistema de gerenciamento de bancos de dados distribuídos das quais pode-se citar:

- Processamento de consultas distribuídas e otimização necessita de algoritmos adequados;
- Seu controle e gerência devem trabalhar de forma integrada;
- Dificuldade no seu desenvolvimento, manutenção e gerenciamento;

- Dificuldade em evitar que erros ocorridos nas máquinas ou na rede atrapalhem o sistema;
- Garantir uma segurança dos dados compartilhados entre as máquinas através do sistema. (O`BRIEN, 2008)

2.4. ÁREAS PROBLEMÁTICAS NOS BANCOS DE DADOS DISTRIBUÍDOS

É necessário resolver diversos problemas técnicos para alcançar o todo o potencial dos SGBDs distribuídos, uma vez que são muito mais complexos. Esta complexidade pode influenciar a estabilidade de um SBDD. Destacam-se alguns problemas técnicos para distribuição de dados: (ÖZSU, 2006)

- *Projeto de bancos de dados distribuídos*: é o principal argumento a ser discutido ao se distribuir os dados por diversos *sites*. Existem duas alternativas: particionar e replicar. A utilização de particionamento pode implicar estudos matemáticos para minimizar o custo de armazenamento do banco de dados pelos *sites*, o tráfego de informações pela rede e a sincronização entre os *sites*.
- *Processamento distribuído de consultas*: a distribuição das informações acelera a consulta, mas diminui a eficiência nas atualizações. O objetivo é otimizar os fragmentos, utilizando paralelismo para melhorar o desempenho nas consultas e nas atualizações.
- *Gerenciamento de diretório distribuídos*: diretório ou catálogo do banco de dados, contém informações como estrutura e localização sobre os itens das de informações no banco de dados. Um destes diretórios pode permanecer centralizado em um único local ou distribuído por vários *sites*, onde pode haver uma cópia ou várias cópias do banco de dados e deverão estar sempre em constante atualização. (ÖZSU, 2006)

- Controle distribuído da concorrência: envolve a sincronização de acessos entre os bancos de dados distribuídos, de forma que a integridade dos dados sejam mantidas. O problema do controle da concorrência em um contexto distribuído é um pouco diferente do que ocorre com a estrutura centralizada. Além de se preocupar com a integridade das informações de um único banco de dados, a consistência de várias cópias do banco de dados também é muito importante.
- Gerenciamento distribuído de impasses: a competição entre usuários pelo acesso a um conjunto de informações, pode resultar em impasse, se o mecanismo de sincronização se basear em bloqueios. As alternativas bem conhecidas de prevenção, anulação e detecção também se aplicam a SBDDs.
- Confiabilidade de SGBDs distribuídos: foi mencionado anteriormente que uma das vantagens dos SBDDs era a maior confiabilidade e disponibilidade. É importante mencionar também os mecanismos para assegurar uma boa consistência dos bancos de dados e, além de detectar defeitos e se recompor deles. Nos casos dos SBDDs, ocorrendo um defeito em que vários *sites* fiquem inoperantes e sem acessos, os bancos de dados existentes nos *sites* que permanecem operacionais devem continuar estáveis e atualizados. Além disso, quando o sistema do computador ou da rede se recuperar da falha, o SBDD deve ser capaz de recuperar e manter atualizados os bancos de dados nos *sites* em que ocorreram a falha. (ÖZSU, 2006)

- Suporte do sistema operacional: o suporte oferecido pelos sistemas operacionais para operações de bancos de dados não corresponde propriamente aos requisitos do software de gerenciamento dos bancos de dados. Os principais problemas são os sistemas de arquivos, o gerenciamento de memória, os métodos de acesso, a recuperação de falhas e o gerenciamento de processos.

2.5. CONCLUSÃO

A principal diferença entre bancos de dados (BD) centralizados e BD distribuídos é o posicionamento das bases de dados, onde no BD centralizado a base de dados está em um único local e no BD distribuído as bases de dados estão espalhadas pelos *sites* da rede.

Existem alguns fatores importantes para se desenvolver um sistema de bancos de dados distribuídos, como por exemplo, o gerenciamento transparente dos dados, a autonomia, a integridade, a economia com *hardware*, dentre outros.

Existem também algumas desvantagens para desenvolver um sistema de bancos de dados distribuídos, pois a complexidade é bem maior do que o sistema centralizado. O custo com treinamento para pessoas sem experiência e com aplicativos é alto. E a segurança é um fator importantíssimo, pois a rede do SBDD tem de ser muito bem protegida, já que a segurança de redes é muito complexa.

3. ESTUDO DE CASO

Para exemplificar o funcionamento de um banco de dados distribuído, vamos citar o caso da rede de locadoras de automóveis denominada Localiza Rent-Car. Localizada em Belo Horizonte, possui 320 agências espalhadas pelo Brasil e pelo mundo, contando com uma frota de cerca de 30.000 automóveis de diversos tipos e marcas.

3.1 Justificativas e necessidades

A Localiza Rent-Car foi escolhida como estudo de caso pelas razões enumeradas abaixo:

- Cada filial deve armazenar os dados de seus veículos localmente, mas deve poder acessar os dados das demais filiais para atender pedidos específicos de clientes.
- Os clientes devem poder fazer reservas *on-line*, acessando somente um *site* (da empresa), sem precisar verificar o *site* de cada filial.
- Não é necessário que o cliente saiba em qual filial o carro está sendo reservado, importando apenas que o cliente vai retirar e entregar o automóvel no local por ele indicado.
- Redução de custos/facilidade de controle, isto é, com as 320 agências espalhadas pelo Brasil e pelo mundo, o cliente consegue com muita comodidade locar um carro sem sair de casa.
- Reduz o risco da utilização constante de dinheiro, ou seja, o cliente efetua o pagamento no momento da reserva. (LEMES, 2012)

3.2 Especificação do problema

Quando se fala em locadora de veículos, logo, imaginamos uma empresa situada em uma cidade onde atende somente a população local. Com isso, certamente um único sistema de controle de frotas e reservas juntamente com um banco de dados centralizado já seria suficiente para tal gerenciamento.

Mas a empresa Localiza não possui apenas uma única agência, conforme visto a Localiza possui várias agências espalhadas pelo Brasil e pelo mundo. Sendo assim estaremos demonstrando como seria extremamente vantajosa e importante a aplicação de um SGBD distribuído. (LEMES, 2012)

Para que possamos estudar este caso tomamos como prioridade algumas regras de negócios, são elas:

- A entrega e retirada do carro são combinadas previamente no momento da reserva no *site*.
- Cada locadora possui igualdades com relação à sua estrutura física, por exemplo, comercial, manutenção e gerência. Com isso, cada uma possui uma meta a ser atingida, seja ela em satisfação ou financeira.

3.3 Modelo de dados

Os dados de todas as locadoras estarão definidos de uma única forma, ou seja, as tabelas serão iguais com relação a sua estrutura, alterando somente o seu conteúdo. Por exemplo, em uma locadora pode não haver um determinado veículo enquanto na outra se encontra disponível, com isso, não perdendo a venda/aluguel. Por fim, como dito, as tabelas estarão organizadas igualmente com relação aos seus campos, tamanho, características e validações. A estrutura do banco e o relacionamento das tabelas estarão organizados da seguinte forma:

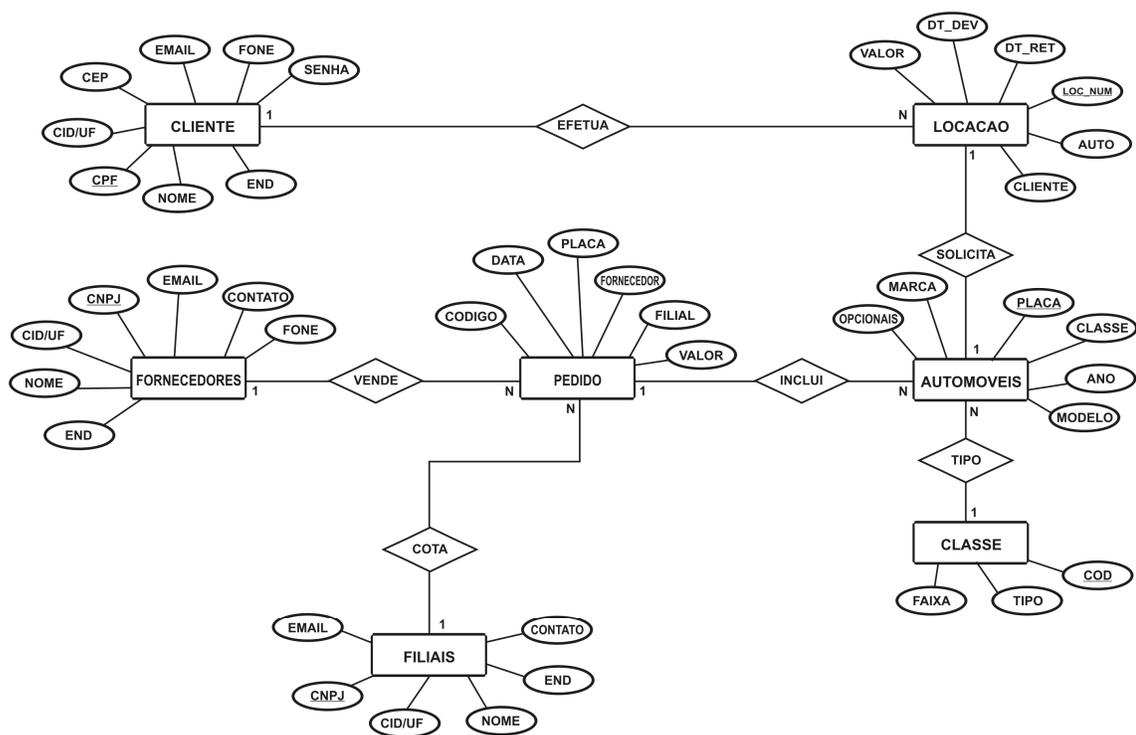


Figura 3.1 - Estrutura física do sistema de banco de dados distribuídos
 Fonte: LEMES, 2012

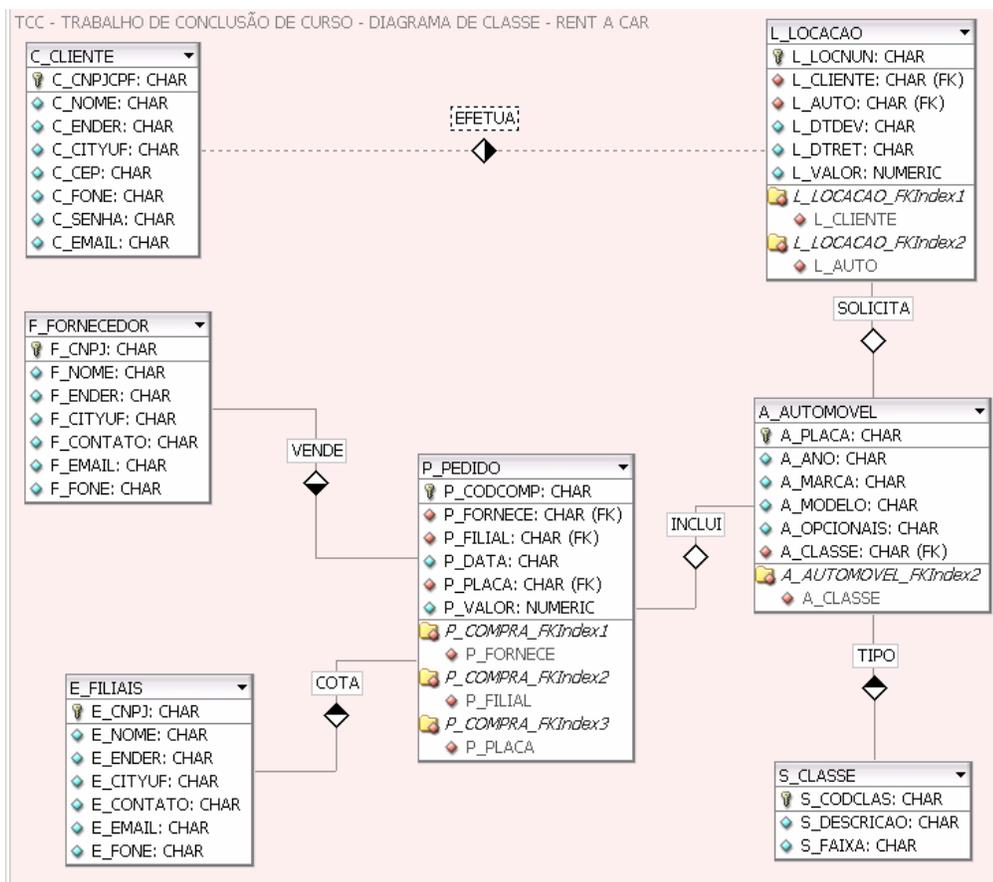


Figura 3.2 - Modelo Relacional do estudo de caso
Fonte: LEMES, 2012

3.4 Projeto

O projeto do estudo de caso será baseado na área em que o cliente tem a possibilidade de efetuar as reservas *on-line*. A figura abaixo mostra o momento da reserva de um determinado veículo:

Figura 3.3 - Tela principal onde o usuário define a retirada e a devolução
 Fonte: LEMES, 2012

3.4.1 Arquitetura

A arquitetura do sistema, mostrada na figura 3.1, está distribuída entre as locadoras da empresa da seguinte forma:

- Localiza - A (sede do grupo - escritório central). Irá conter um banco de dados com as devidas tabelas, conforme citado na figura 3.2. Conseqüentemente este banco de dados estará configurado para se tornar um Sistema de Banco de Dados Distribuídos.

- Localiza - B, Localiza - C e Localiza - D (possui uma unidade de negócio). Irá conter um banco de dados idêntico, em termos de estrutura, ao descrito no item anterior, contendo diferenças somente no conteúdo das tabelas.

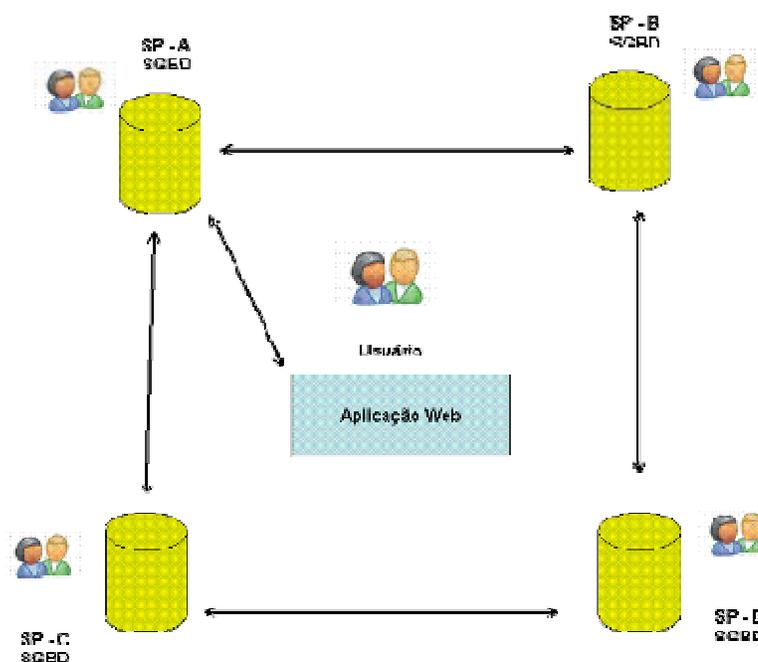


Figura 3.4 - Representação da arquitetura do protótipo do sistema
 Fonte: LEMES, 2012

3.4.2 Projeto do banco de dados

Esta seção mostra a estrutura do banco de dados distribuído definido para a solução, incluindo a criação das tabelas e seus devidos campos e também a configuração necessária no SGBDD.

Para que um sistema de banco de dados se torne distribuído é necessário verificar se o SGBD tem suporte a distribuição. Uma vez verificado o próximo passo é configurá-lo, para isso destacam-se os principais passos de configuração de um SGBDD:

- Primeiramente determina quem será o DISTRIBUIDOR, ou seja, o servidor no qual será responsável em distribuir os dados.

- Em seguida é necessário definir o PUBLICADOR, ou seja, as bases de dados no qual cria-se um tipo de “compartilhamento” entre eles.

- Por último cria-se os ASSINANTES, ou seja, esses serão adicionados nos publicadores de acordo com cada banco de dados. Um assinante do tipo *PUSH* é definido pela publicadora, ou seja, quem publicou os dados define quem vai recebê-los. Já o assinante do tipo *PULL* escolhe as publicações que quer assinar. A figura abaixo demonstra um SGBDD: (LEMES, 2012)

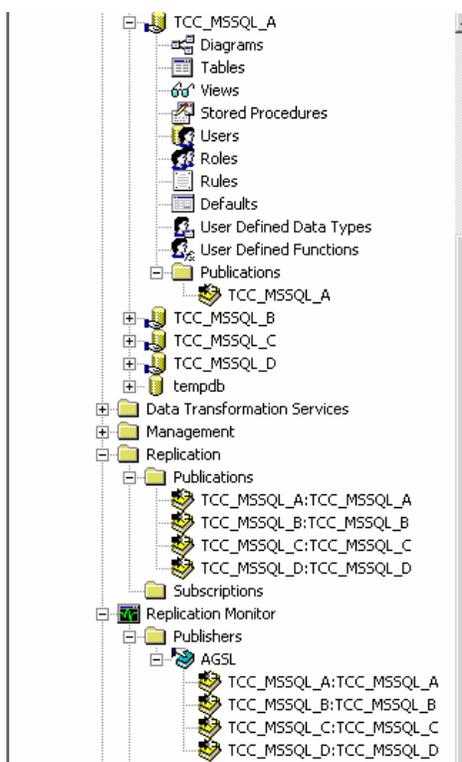


Figura 3.5 - SGBDD configurado
Fonte: LEMES, 2012

3.5 Implementação

O banco de dados escolhido foi o SQL Server 2000, da Microsoft. Para um banco de dados se tornar distribuído o mesmo sofre modificações como replicação e fragmentação, com isso poderemos mostrar a idéia de um SGBDD. (GUNDERLOY, 2000)

Principais características do SQL Server 2000:

- Possui ferramentas gráficas para administração e desenvolvimento (*Query Analyzer* e *Enterprise Manager*) e replicação de dados.
- Possui ferramenta para aferição de desempenho (*Profiler*) , que permite ao administrador capturar, armazenar e analisar eventos que responsáveis por gargalos no servidor.
- *Views* distribuídas (*Distributed Partitioned Views*) , que podem ser utilizadas para balanceamento e distribuição de carga entre servidores.
- *Failover Clustering*, que fornece mecanismos para criar cópias *on-line* de sua base de dados à partir de um servidor virtual, responsável pelo manejo das requisições.

As técnicas abaixo aplicadas são necessárias pois só assim o SGBD SQL Server 2000 se tornará distribuído. Antes de iniciar sobre as principais considerações, devemos levar em conta que as bases e as tabelas já estão devidamente criadas no SGBD (GUNDERLOY, 2000). Veja as etapas:

- a) Definir qual servidor será o Distribuidor

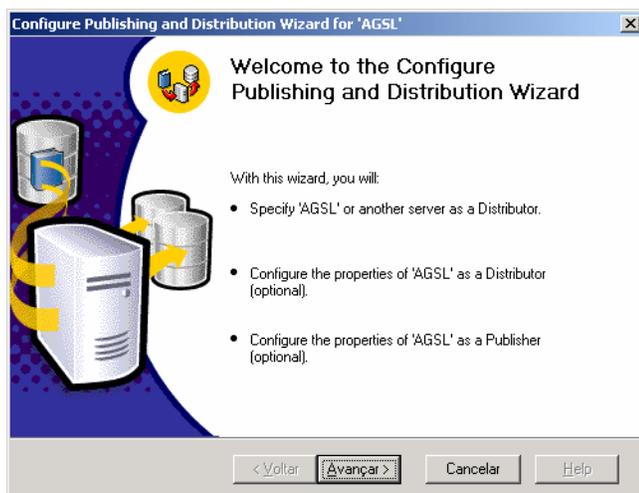


Figura 3.5 - Tela inicial da definição do distribuidor
Fonte: LEMES, 2012

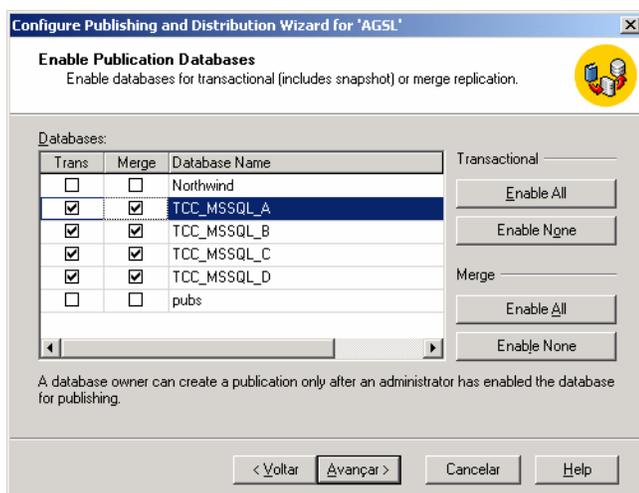


Figura 3.6 – Escolha das bases a serem publicadas
Fonte: LEMES, 2012

b) Uma vez definido o distribuidor, nota-se que as bases de dados ficam com a seguinte aparência:

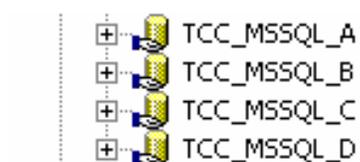


Figura 3.7 – Bases selecionadas para publicação
Fonte: LEMES, 2012

c) Para que os bancos se tornem visíveis entre si, a próxima etapa é criar as publicações.

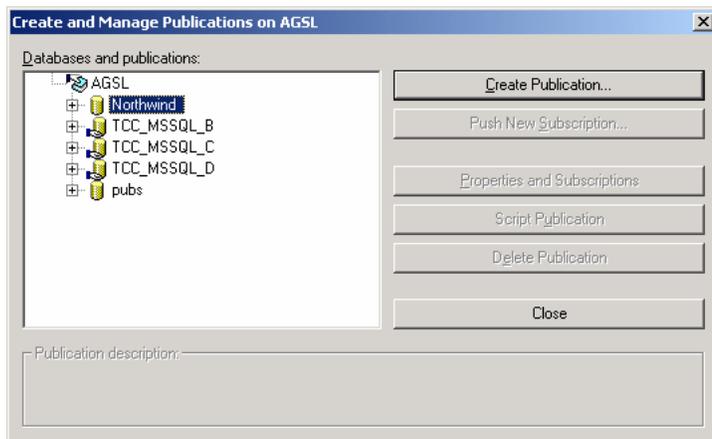


Figura 3.8 - Esta é a tela inicial na qual define-se qual base será publicado.
Fonte: LEMES, 2012

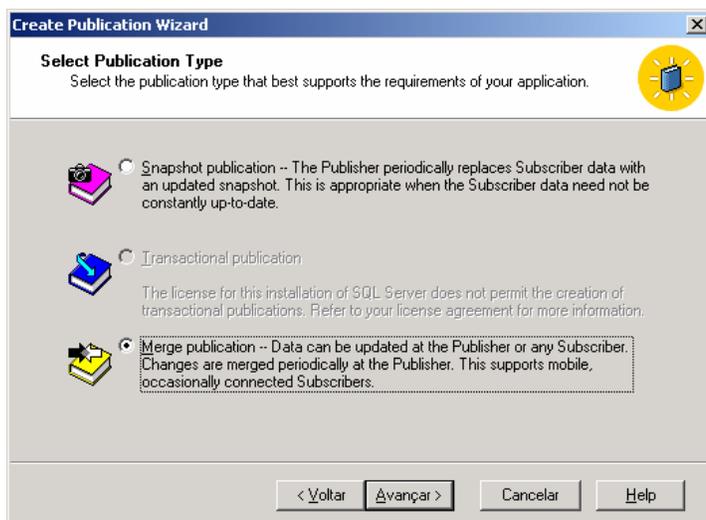


Figura 3.9 – Definição do tipo de publicação dos dados
Fonte: LEMES, 2012

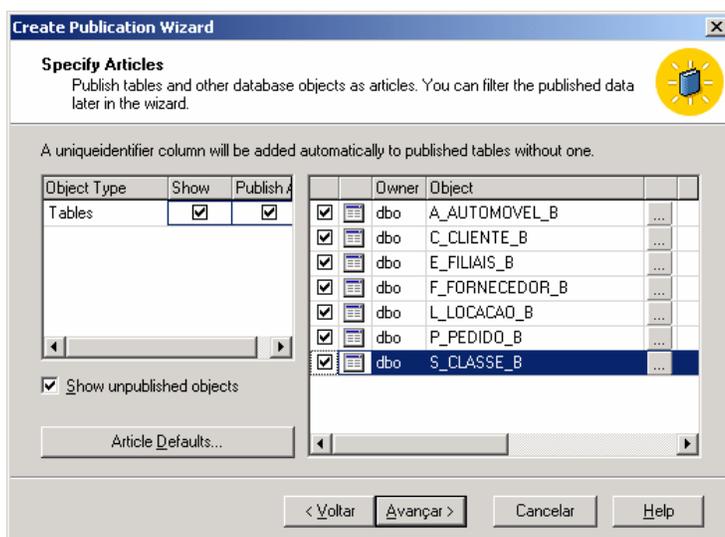


Figura 3.10 – Definição de quais tabelas/artigos farão parte da publicação
Fonte: LEMES, 2012

d) A próxima e última etapa trata dos assinantes, ou seja, quais tabelas de quais bases poderão assinar as publicações das outras bases de dados. Para que tudo seja feito com êxito é extremamente importante seguir todas as etapas anteriores.

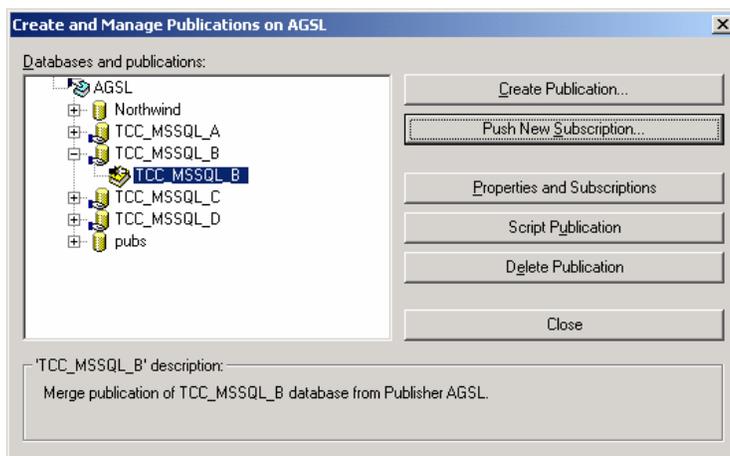


Figura 3.11 – Definição de qual banco será assinado por uma determinada publicação
Fonte: LEMES, 2012



Figura 3.12 - Define-se em qual distribuidor serão assinados quais artigos/tabelas.
Fonte: LEMES, 2012

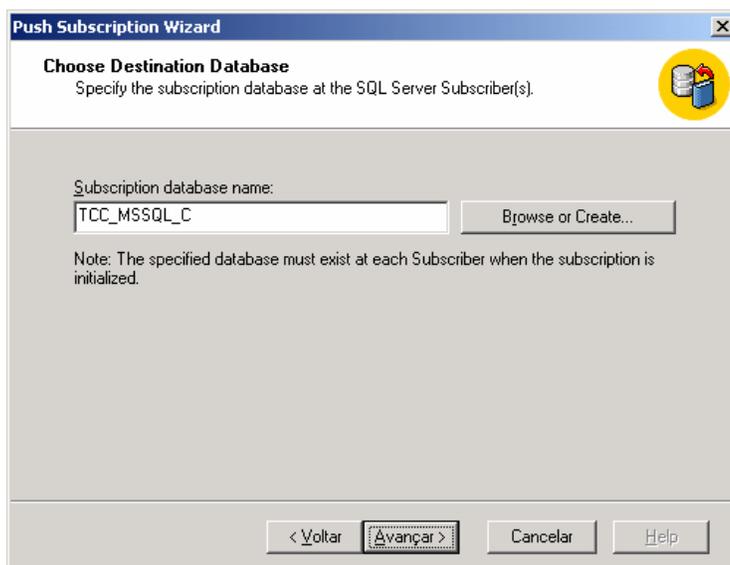


Figura 3.13 – Escolha de qual publicação irá receber um determinado assinante/tabela.
Fonte: LEMES, 2012



Figura 3.14 - Tela de conclusão da definição do assinante na publicação.
Fonte: LEMES, 2012

Para concluir devemos considerar alguns dados importantes para o desenvolvimento deste estudo de caso:

- Existe 1 distribuidor local.
- 4 bases de dados, na qual todas serão assinantes e publicadora.
- Apenas uma, na qual conterà os dados da matriz, será a publicadora de todas as outras assinantes, ou seja, ela será a responsável por publicar os dados das outras filiais.

3.6 Resultados

O estudo de caso apresentado possibilitou a coleta de resultados por meio de testes efetuados num mesmo sistema, no qual um utiliza um BD centralizado e o outro um BDD. (LEMES, 2012)

Os resultados foram concluídos com base em:

- Duas aplicações *web* foram desenvolvidas conectadas em bases de dados configuradas de forma distinta, ou seja, a primeira aplicação foi desenvolvida com o objetivo de demonstrar a funcionalidade de um banco de dados centralizado; já a segunda utiliza um banco de dados distribuído devidamente preparado e configurado.

- As aplicações realizam as mesmas operações de consulta, só que a primeira conecta várias vezes em bancos diferentes até que encontre o resultado da busca; já a aplicação com BDD conecta-se uma única vez ao banco principal retornando assim o resultado da SELECT efetuada.

- O grande resultado observado é quanto à velocidade de busca das consultas efetuadas pelos clientes. Enquanto uma aplicação com BD Centralizado demora, dependendo do servidor, em torno de 6 segundos para retornar o resultado da SELECT, a aplicação com BD Distribuído demora 2,5 segundos. Isso se dá pelo fato do mesmo efetuar apenas uma conexão, portanto podemos concluir que podemos ter um ganho de aproximadamente 60%, lembrando que isso dependerá da conexão e do servidor. (LEMES, 2012)

3.6.1 Vantagens

A principal vantagem da utilização de um banco de dados distribuído em uma aplicação *web* é a característica de partilhar e acessar dados de uma maneira confiável e eficiente, adicionalmente, aumento de velocidade no processamento de consultas e do crescimento incremental, de possuir disponibilidade, confiabilidade e um controle distribuído, além da sua transparência e autonomia. (LEMES, 2012)

Com relação ao estudo de caso apresentado podemos destacar uma necessidade muito interessante que só foi possível graças a utilização de um

gerenciador de banco de dados distribuído, essa necessidade diz respeito ao compartilhamento dos dados entre as filiais, sendo que cada compartilhamento possui um controle de distribuição, caracterizando assim confiabilidade e disponibilidade.

Além disso, outras relações importantes são: rapidez no processamento das consultas e a transparência para o usuário, ou seja, em momento algum o usuário sabe em qual local tal consulta esta sendo executada.

Considerando as vantagens no compartilhamento de dados, confiabilidade e disponibilidade, temos: (LEMES, 2012)

- Compartilhamento de Dados e Controle de Distribuição: a principal vantagem de partilhar informações pela distribuição de dados é que cada nó é capaz de reter um grau de controle sobre os dados armazenados localmente. Em sistemas distribuídos, existe um administrador global do banco de dados que é responsável pelo sistema como um todo. Uma parte dessas responsabilidades é delegada ao administrador do banco de dados local para cada nó. Dependendo do projeto do banco de dados distribuído, cada administrador pode ter um grau diferente de autonomia local.

- Confiabilidade e Disponibilidade: confiabilidade significa que um sistema funciona conforme foi projetado. Disponibilidade quer dizer que o sistema realiza suas funções sempre que é requerido.

Ambas são necessárias: um sistema confiável que não esteja sempre disponível é impraticável; um sistema disponível que não seja confiável também. Se um nó falhar em um sistema distribuído, os nós remanescentes podem ser capazes de continuar operando. A falha de um nó pode ser detectada pelo sistema e uma ação apropriada pode ser necessária para recuperar a falha. Quando o nó que falha se recupera ou é reparado, deve

haver mecanismos disponíveis para integrar habilmente o nó de volta ao sistema. (LEMES, 2012)

Embora a recuperação de falhas seja mais complexa em sistemas distribuídos, a habilidade da maioria dos sistemas para continuar a operar apesar da falha de um nó resulta no aumento da disponibilidade, que, por exemplo, é crucial em sistemas de banco de dados utilizados em aplicações de tempo real.

Aceleração no processo de consultas:

Se uma consulta envolve dados em diversos nós, é possível dividi-la em sub-consultas que podem ser executadas em paralelo. Entretanto, em um sistema distribuído, não há o compartilhamento da memória principal, assim sendo, nem todas as estratégias para processadores paralelos podem ser aplicados diretamente a sistemas distribuídos. Nesses casos, em que os dados são duplicados, as consultas podem ser direcionadas pelo sistema para os nós com menos carga.

Transparência e Autonomia:

A transparência de rede é o grau em que os usuários do sistema podem estar despreocupados com os detalhes do projeto do sistema distribuído, ou seja, 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. (LEMES, 2012)

Os usuários do banco de dados visualizariam uma única imagem da base de dados logicamente integrada, embora ela estivesse fisicamente distribuída. A autonomia local é o grau em que um projetista ou administrador de um nó pode ser independente do restante do sistema distribuído.

Considerando as questões de transparência e autonomia, pode-se ter:

- Autonomia de Nomeação e de Local: em banco de dados distribuídos, cuidados precisam ser tomados para garantir que dois nós não usem o mesmo nome para itens de dados distintos. Uma solução para este problema é requerer que todos os nomes sejam registrados em um servidor de nome central. Para o aumento de autonomia local é requerer que cada nó prefixe seu próprio identificador para qualquer nome que ele gerar. Isto assegura que dois nós não gerem o mesmo nome (uma vez que cada nó tem um único identificador).

- Transparência de Reprodução e Fragmentação: quando um item de dado é requisitado, a réplica específica não precisa ser nomeada. Em vez disso, uma tabela de catálogo é usada pelo sistema para determinar todas as réplicas para o item de dado. Da mesma forma, um usuário não deve ser solicitado a saber como um item é fragmentado.

- Transparência de Localização: a transparência de localização é obtida criando-se um conjunto de nomes alternativos ou *aliases* para cada usuário. Um usuário pode então se referir aos itens de dados por meio de nomes simples, os quais são traduzidos pelo sistema para nomes completos. Com *aliases*, o usuário pode despreocupar-se em relação à localização física de itens de dado. Ele não é afetado se o administrador do banco de dados decidir remover um item de um nó para outro. (LEMES, 2012)

- **Transparência e Atualização:** o problema principal é assegurar que todas as réplicas de um item de dado e todos os fragmentos afetados sejam atualizados. O problema de atualização para dados reproduzidos e fragmentados está relacionado ao problema de atualização de visões.

3.6.2 Desvantagens

Todo sistema, apesar de eficiente, também têm suas desvantagens e algumas delas estão relacionadas aos custos e à complexidade, podendo gerar uma degradação do desempenho e um aumento do número de “*bugs*”. (LEMES, 2012)

Custos:

O custo em nível de *software* é bem mais alto em BDD do que em um banco de dados centralizado. Os fatores que determinam o aumento do custo são:

- Desenvolvimento de algoritmos eficientes e rápidos para controlar a fragmentação, replicação de dados e processamento de consultas.
- Prevenção de falhas em nós da rede. Quando houver falha, o sistema tem que continuar funcionando. Para isto, deve-se desenvolver algoritmos contendo um número de replicação dos dados satisfatória para suprir esta necessidade.
- Os *softwares* desenvolvidos devem ser portáteis para serem executados em qualquer *hardware* e sistema operacional, ou o sistema de um ambiente para outro deve sofrer pequenas alterações.
- Os algoritmos de consulta devem ser rápidos e eficientes para satisfazer a expectativa de tempo x resposta que o usuário tem em relação ao sistema.

- Os algoritmos de *backup* devem ser capazes de guardar as últimas atualizações do sistema, além de especificar se os mesmos devem ser feitos à noite, ao meio-dia ou se devem parar todo sistema.

O mais importante é que o SGBDD e os programas devem garantir, acima de tudo, dados consistentes em todas as atualizações e consultas feitas.

Complexidade:

A complexidade adicional requerida para assegurar a própria coordenação entre os nós é a principal desvantagem de um SGBDD. Por meio dela são gerados um maior potencial de erros; uma vez que os nós que formam o sistema distribuído operam em paralelo, sendo mais difícil assegurar a correção dos algoritmos, há uma degradação do desempenho de processamento, uma vez que a mudança de mensagens e a computação adicional requerida para realizar a coordenação interna não surgem em sistemas centralizados, gerando assim esse baixo desempenho. (LEMES, 2012)

Dependendo de como a consulta foi desenvolvida (algoritmo) pode tornar-se muito complexa e levar muito tempo para ser executada.

4. CONCLUSÃO

Este trabalho mostrou a funcionalidade do Banco de Dados Distribuídos, onde foram mostradas suas vantagens, desvantagens, principais características e funcionalidades, mostrando que possui funcionalidades essenciais para atender a uma empresa.

Atualmente, com a intensa necessidade de armazenar dados de uma empresa, são utilizados diferentes tipos de banco de dados para manter esse controle. Para se utilizar um banco de dados é preciso primeiro analisar os objetivos da empresa, qual será o melhor tipo para suprir suas necessidades.

Em empresas consideradas de grande porte, um sistema de banco de dados distribuídos (SGBD) torna-se necessário. No entanto, em empresas de pequeno e médio porte um SGDB pode ser desejável, mas não necessário caso o custo e da complexidade sejam grandes demais.

Existem vários métodos e sistemas de gerenciamento de bancos de dados (SGBD) para controlar um SBDD. Logicamente existem métodos melhores do que outros. O método utilizado no caso da empresa Localiza Rent-Car produziu ganhos, principalmente em disponibilidade, confiabilidade e velocidade nas consultas e alterações.

Ao se utilizar um bom método, pode-se tornar uma empresa mais informatizada, com recursos mais baratos. Sendo assim, as empresas poderão expandir seus negócios em diferentes localidades mantendo suas bases de dados sempre em sincronia.

REFERÊNCIAS BIBLIOGRÁFICAS

ÖZSU, M. Tamer; VALDURIEZ, Patrick. **Principles of Distributed Database Systems**. 2ª Ed, Pearson, 2006.

KORTH, Henry F., **Sistema de Banco de Dados**, tradução [da 2ª ed. ver] Maurício Heihachiro Galvan Abe. São Paulo: Makron Books, 1995.

CASANOVA, Marco Antônio. **Banco de Dados**. Disponível em: <<http://www.inf.puc-rio.br/~casanova/INF1731-BD/modulo13.pdf>>. Acesso em 25 de fevereiro de 2012.

O`BRIEN, James A. **Management Information Systems**, 7ª Ed, Mcgraw-Hill/Irwin, 2008.

LEMES, Adriano G. Souza. **Banco de Dados Distribuídos em Aplicações Web**. Disponível em: <<http://bibdig.poliseducacional.com.br/document/?down=37> >. Acesso em: 16 de julho de 2012.

FACHIN, André. **Banco de Dados Distribuídos**. Disponível em: <<http://www.slideshare.net/andrefachin/banco-de-dados-distribuidos-66037> >. Acesso em 2 de fevereiro de 2012.

MATTOSO, Marta. **Introdução a banco de dados distribuídos**. Disponível em: <<http://www.cos.ufrj.br/~marta/IntroductionP.pdf>>. Acesso em: 25 de fevereiro de 2012.

DATE, C. J.. **Introdução a Sistemas de Banco de Dados**. 8ª Ed, Campus, 2000.

GUNDERLOY, Mike. **Mastering SQL Server 2000**. 1ª Ed, John Wiley & Sons, 2000.