

# BANCO DE DADOS DISTRIBUÍDOS

Douglas Vinício Ribeiro Schneid <douglas89\_ribeiro@hotmail.com>

Lelio Ronei Knop <leliork@gmail.com>

Mauro da Silva Unger <msunger89@gmail.com>

Universidade Luterana do Brasil (Ulbra) – Curso de Ciência da Computação – Campus Canoas

Av. Farroupilha 8001 – Bairro São Luis – CEP 92420-280 – Canoas – RS

28 de maio de 2013

## RESUMO

Este artigo descreve as características de um banco de dados distribuídos, apresentando suas características, vantagens e desvantagens, os problemas que podem ocorrer como fragmentação, concorrência de acesso e inconsistências, assim como as medidas que visam contornar estas ocorrências para um melhor funcionamento dispondo de segurança e garantindo a integridade dos dados.

**Palavras-chave:** Sistemas Distribuídos, Banco de Dados.

## ABSTRACT

*Title: “Distributed Database”*

*This article describes the characteristics of a distributed database, presenting their characteristics, advantages and disadvantages, problems that may occur as fragmentation, competition for access and inconsistencies, as well as measures to overcome these occurrences vision for a better operation providing security and ensuring data integrity.*

*Key-words: Distributed Systems, Database.*

## 1 INTRODUÇÃO

Este artigo irá abordar as características de um banco de dados distribuído, que é um banco que possui mais de uma base que podem se localizar fisicamente em locais distantes entre si. Sendo útil quando há necessidade de haver uma base que esteja sempre disponível. Porém o fato de ser possuir dados distribuídos pode gerar inconsistências e fragmentações, que podem ser evitadas utilizando réplicas e protocolos de acesso com o Timestamp.

Um banco de dados distribuído apresenta diferentes formas de fragmentação, que podem ser horizontal, vertical e mista, diferentes formas de armazenamento, e diferentes formas de replicação, que são podem se síncrona ou assíncrona. Também havendo dois tipos de banco de dados distribuídos, homogêneos e heterogêneos.

## 2 CONCEITO

BDD é um conjunto de varias Bases de Dados, logicamente conectadas, distribuídas por uma rede de computadores, tanto cabeada, quando sem fio. Possui dois tipos de BDD, falaremos de ambos a seguir:

- **HETEROGÊNEOS:** Como já diz no nome “Heterogêneo”, “contrário“, é aquele que dispõem de diferentes tipos de banco de dados.
- **HOMOGÊNIOS:** Já os homogêneos, é aquele banco distribuído com mesmo tipo de banco de dados. Existem dois métodos de Armazenamento de banco de dados, são eles Replicados ou Fragmentados. Os tipos são achados no decorrer dos servidores de um sistema distribuído.
  - **REPLICADOS:** Nas transições replicadas, existe uma cópia de cada base de dados em cada servidor. Podem ocorrer as transações de duas maneiras síncrona ou assíncrona. No caso das síncronas, cada transação é dada como concluída quando todos os servidores do BDD tiverem confirmado sua execução local com sucesso. Já a réplica assíncrona há um banco principal, o qual executa sua transação e após sua confirmação, este servidor encaminha a transação para os demais servidores.
  - **FRAGMENTADOS:** As transações fragmentadas as bases de dados são diferentes em cada servidor, porem ambas estão inter-relacionadas. Assim cada servidor de base de dados sabe para qual

servidor deve distribuir os dados, pois cada servidor possui as informações dos bancos adjacentes.

### 3 VANTAGENS E DESVANTAGENS

- Vantagens:
  - Em um banco de dados não pode haver falhas, tendo em vista que, as informações ficam armazenadas no mesmo e o rompimento destas pode acarretar uma perda imensurável dependendo da aplicação, por isso, o BDD tem a vantagem de realizar suas transações sem que haja uma falha generalizada no banco de dados inteiro. Havendo uma falha na aplicação, esta será local, ou seja, ficara somente no servidor com falha.
  - Mais um ponto positivo do banco de dados distribuídos é a melhora no desempenho, pois estão localizados próximo da aplicação com a chamada da função.
  - A modulação é um fator vantajoso, tendo em vista que, pode ocorrer diversas atualizações ou modificações sem que afete os outros módulos (servidores de banco de dados distribuídos)do sistema.
  - Pode haver um aproveitamento financeiro ao montar um sistema distribuído, pois, pode ser utilizado uma rede de computadores ao invés de um único computador servidor.
- Desvantagens:
  - A complexidade de infraestrutura de um banco de dados distribuídos é alta, tem de ser considerada a fragmentação dos dados, alocação dos fragmentos em lugares específicos e a replicação de dados.
  - A segurança é ameaçada quando as informações não ficam em um SGBD centralizado, tendo em vista que, devem haver firewall e proxy em cada servidor de banco distribuído.
  - A falta de padronização das metodologias e ferramentas para auxiliar os usuários decreta outra desvantagem de um sistema distribuído.

### 4 ARMAZENAMENTO

Em um banco de dados distribuídos podem existir vários tipos de enfoques para o armazenamento das relações ou tabelas:

#### 4.1 Replicação

A replicação pode ser definida como um processo onde é feito a geração e a reprodução de várias cópias (réplicas) de dados em um ou mais nós, dessa forma é feito várias cópias idênticas da relação, armazenadas em nós diferentes.

A vantagem dessa técnica é a disponibilidade das informações, mas por outro lado, essa técnica gera um overhead, em razão do aumento de armazenamento e processamento.

A replicação pode ser dividida em dois grupos:

- **Síncrona:** cada transação é concluída quando os nós confirmam que a transação local foi finalizada com sucesso.
- **Assíncrona:** o nó principal faz a execução da transação enviando a confirmação ao solicitante e então redireciona a transação aos demais nós.

A replicação pode ser considerada um pré-requisito para obter o melhor potencial de um sistema distribuído.

#### 4.2 Fragmentação

A relação é dividida (particionada) em vários fragmentos, sendo esses armazenados em lugares diferentes. A fragmentação pode ser dividida de três formas:

- **Horizontal:** é quando algumas relações são divididas de forma horizontal, agrupando linhas para criar subconjuntos de tuplas (registros) de uma relação. Um exemplo é a figura 1, que mostra a relação completa “PROJ” e alguns registros de uma relação fragmentada de forma horizontal, representado pela “PROJ<sub>1</sub>”.

PROJ				PROJ <sub>1</sub>			
PNO	PNAME	BUDGET	LOC	PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal	P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York	P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York				
P4	Maintenance	310000	Paris				
P5	CAD/CAM	500000	Boston				

Figura 1: Relação inteira (PROJ) e uma fragmentada horizontalmente (PROJ<sub>1</sub>). (SHIBAYAMA, 2004)

- **Vertical:** nesse tipo de fragmentação, é feito a divisão de uma relação de forma vertical, caso o nó não precise de todos os atributos de uma relação. Sendo que um fragmento de uma relação contém apenas alguns atributos da relação. Esse tipo de técnica não é muito recomendado, pois se dois fragmentos estiverem armazenados de forma separados, esses não poderão ser recolocados juntos as tuplas originais, pois não existe nenhum elemento em comum com os dois fragmentos, uma solução para isso seria colocar chave primária em todos os fragmentos, para que se possa então reconstruir a relação novamente.

PROJ				PROJ <sub>2</sub>		
PNO	PNAME	BUDGET	LOC	PNO	PNAME	LOC
P1	Instrumentation	150000	Montreal	P1	Instrumentation	Montreal
P2	Database Develop.	135000	New York	P2	Database Develop.	New York
P3	CAD/CAM	250000	New York	P3	CAD/CAM	New York
P4	Maintenance	310000	Paris	P4	Maintenance	Paris
P5	CAD/CAM	500000	Boston	P5	CAD/CAM	Boston

Figura 2: Relação inteira (PROJ) e uma fragmentada verticalmente (PROJ<sub>2</sub>). (SHIBAYAMA, 2004)

- **Mista:** podem ocorrer casos em que se pode ter tanto a fragmentação vertical, seguida por uma horizontal ou vice-versa, gerando assim uma fragmentação mista. Na figura 3, pode-se ver que a parte fragmentada (PROJ<sub>3</sub>) é formada através da aplicação dos métodos de fragmentação horizontal e vertical em uma relação inteira (PROJ).

PROJ				PROJ <sub>3</sub>		
PNO	PNAME	BUDGET	LOC	PNO	PNAME	BUDGET
P1	Instrumentation	150000	Montreal	P1	Instrumentation	150000
P2	Database Develop.	135000	New York	P2	Database Develop.	135000
P3	CAD/CAM	250000	New York			
P4	Maintenance	310000	Paris			
P5	CAD/CAM	500000	Boston			

Figura 3: Relação inteira (PROJ) e uma fragmentada (PROJ<sub>3</sub>). (SHIBAYAMA, 2004)

### 4.3 Replicação e Fragmentação

As duas técnicas anteriormente citadas, podem ser usadas várias vezes em uma mesma relação. Conforme o Exemplo de Almeida (2013):

“Um fragmento pode ser replicado, e as réplicas podem ser fragmentadas novamente e assim por diante.”.

## 5 TIPOS DE BANCOS DE DADOS DISTRIBUÍDOS

Um Sistema de Gerenciamento de Banco de Dados Distribuídos (SGBDD) pode ser classificado da seguinte forma:

- **Homogêneo:** onde o banco de dados é igual em todos os nós, geralmente usado em projetos que tem seu desenvolvimento iniciado do zero.
- **Heterogêneo:** o banco de dados é diferente nos nós, esse tipo é usado quando existe alteração ou até a migração de um projeto que já está desenvolvido.

## 6 TRANSAÇÕES E CONTROLE DE CONCORRÊNCIA

Em um banco de dados distribuídos onde temos uma base descentralizada permitindo acessos simultâneos é imprescindível um controle de concorrência para evitar um problema chamado anomalia de sincronização que apresentam três problemas, sendo eles:

- Gravação de dados Inconsistentes;
- Acessar dados de forma inconsistente;
- Perda de Dados.

Para evitar estas anomalias devem ser tomadas algumas medidas. Inicialmente é necessário haver um protocolo de acesso. Uma forma muito utilizada para este protocolo é a divisão por “visões”, onde cada usuário possui uma visão do banco. Também devem ser definidas permissões e restrições de acesso para cada nível de usuário, onde em cada nível haverá uma visão mais restrita ou ampla da base.

Porém apesar de auxiliarem os métodos citados a cima não evita por completo a ocorrência de anomalias decorrentes de concorrência, ainda se faz necessário um protocolo que gerencie conflitos de acesso, por exemplo, dois usuários de mesmo perfil terão a mesma visão do banco, e caso um esteja realizando uma consulta ao mesmo tempo em que outro realiza uma inserção, poderá causar acesso ou gravação de dados inconsistentes.

Para este gerenciamento pode-se citar entre outros o Timestamp, este processo atribui uma identificação para cada transação, e esta somente é aplicada de forma efetiva no banco quando estiver concluída, sendo cada transação efetuada de uma vez, e havendo transações conflitantes a última atualização a solicitada prevalece sobre a(s) anterior(es). No exemplo citado no parágrafo anterior este processo evitaria transtornos, pois caso a primeira transação a ser iniciada seja uma consulta (T1) e logo após a atualização (T2), ao pedir acesso ao banco cada transação recebe seu time, T1 recebe 0 e T2 recebe 100, T2 só terá acesso a base quando T1 tiver sido concluída.

## 7 CONCLUSÃO

Pelos embasamentos referenciados a cima, tiramos como conclusão que um banco de dados distribuído, é um banco mais ágil e confiável em relação aos bancos de dados centralizados, no momento de realizar um *deploy* da aplicação. Tiramos como lição o grande número de protocolos utilizados nesse processo para controle de concorrência e transação, onde damos ênfase no *timestamp*, o qual estabelece um tempo de acesso para cada transação. Atentamos para o baixo custo de montagem de servidores para estabelecer um banco de dados, tendo em vista, que pode ser utilizados computadores de baixo desempenho.

## REFERÊNCIAS

ALMEIDA, Roniere. **O que é um Banco de Dados distribuído?**. Disponível em: <<http://www.devmedia.com.br/o-que-e-um-banco-de-dados-distribuido/24762>>. Acesso em: 05 maio 2013.

CASANOVA, Marco Antonio, MOURA, Arnaldo Vieira. **Principio de Sistemas de Gerência de Banco de Dados Distribuídos**. Disponível em <[www.inf.puc-rio.br/~casanova/Publications/Books/1985-BDD.pdf](http://www.inf.puc-rio.br/~casanova/Publications/Books/1985-BDD.pdf)> acessado em: 21 mai. 2013.

COSTA, Alex da Silvaeira. **Um Protocolo Distribuído para Controle de Consistência em Bancos de Dados Replicados para Ambiente de Computação Móvel**. Disponível em: <[www2.dc.uel.br/nourau/document/?view=887](http://www2.dc.uel.br/nourau/document/?view=887)> acessado em: 23 mai. 2013

ELMASRI AND NAVATHE, *Fundamentals of database systems* (3rd edition), Addison-Wesley Longman.

OZSU, M. Tamer. VALDURIEZ, Patrick. *Principles of Distributed Databases (2nd edition)*. Ed. Springer, Londres – ING 2011.

SHIBAYAMA, Eric Teruo. **Estudo Comparativo entre Bancos de Dados Distribuídos**. Londrina. 2004. Disponível em: <[www2.dc.uel.br/nourau/document/?down=155](http://www2.dc.uel.br/nourau/document/?down=155)>. Acesso em: 21 maio 2013.

VASCONCELOS, Leandro Guarino De. **Banco de Dados Distribuídos**. 2012. Disponível em: <[http://www.4learn.pro.br/leandro/bdd/introducao\\_bdd.pdf](http://www.4learn.pro.br/leandro/bdd/introducao_bdd.pdf)>. Acesso em: 21 maio 2013.