

## HISTÓRIA DOS BANCOS DE DADOS

**E**m uma primeira abordagem sobre o assunto é muito importante conhecer como começou toda essa história Bancos de Dados e discutirmos como será o futuro dessa ferramenta tão importante nas empresas e em nossas vidas. Aqui conheceremos os primórdios dos bancos de dados, sua evolução e as perspectivas de futuro.

### 1.1 DÉCADA DE 60 – PRIMÓDIOS

Como muitas tecnologias na computação industrial, os fundamentos de bancos de dados surgiram na empresa IBM na década de 1960 através de pesquisas de funções de automação de escritório. Foi um período da história na qual as empresas descobriram que estava muito custoso empregar um número grande de pessoas para fazer trabalhos como armazenar e indexar (organizar) arquivos. Por este motivo, valia a pena os esforços e investimentos em pesquisar um meio mais barato e ter uma solução mecânica eficiente.

Nestes primórdios os dados eram armazenados diretamente em arquivos, os quais implementavam alguns inconvenientes na armazenagem de dados, como:

- redundâncias e inconsistências
- dificuldade de acesso
- falta de integridade lógica
- falta de atomicidade nas transações
- insegurança

Um dos primeiros sistemas de banco de dados foi o IMS (*Information Management System*) da IBM lançado no final da década de 60.

### 1.2 DÉCADA DE 70 – GÊNESIS

Na década de 1970 o pesquisador da IBM - Ted Codd - publicou o primeiro artigo sobre bancos de dados relacionais. Este artigo tratava sobre o uso de cálculo e álgebra relacional para permitir que usuários não técnicos armazenassem e recuperassem uma grande quantidade informações. Codd visionava um sistema onde o usuário seria capaz de acessar as informações através de comandos, nas qual as informações estariam armazenadas em tabelas.

Devido à natureza técnica deste artigo e a relativa complicação matemática, o significado e proposição do artigo não foram prontamente realizados. Entretanto ele levou a IBM a montar um grupo de pesquisa conhecido como *System R* (Sistema R), que deu origem a um sistema de banco de dados relacional com o mesmo nome.

O Sistema R evoluiu para SQL/DS, que posteriormente passou a chamar o DB2, o mais conhecido sistema gerenciador de banco dos dados da IBM. O Sistema R utilizava uma linguagem chamada *Structured Query Language* (SQL) - Linguagem de Consulta Estruturada. Linguagem esta que se tornou um padrão na indústria para bancos de dados relacionais e hoje em dia é um padrão ISO (*International Organization for Standardization* - Organização Internacional de Padronização).

Outro estudioso que propôs o modelo de dados relacional foi Edgar Frank Codd, entre 1970 e 1972, que se tornou um marco em como pensar em banco de dados. Ele desconectou a estrutura lógica do banco de dados do método de armazenamento físico. Este sistema tornou-se padrão desde então.

O Dr. Peter Chen também foi ofereceu grande contribuição ao propor o modelo Entidade-Relacionamento (ER) para projetos de banco de dados dando uma nova e importante percepção dos conceitos de modelos de dados.

Mesmo a IBM sendo a companhia que inventou o conceito original e o padrão SQL, eles não produziram o primeiro sistema comercial de banco de dados. O feito foi realizado pela *Honeywell Information Systems Inc.*, cujo sistema foi lançado em junho de 1976. O sistema era baseado em muitos princípios do sistema que a IBM concebeu, mas foi modelado e implementado fora da IBM.

### 1.3 DÉCADA DE 80 – DESENVOLVIMENTO

Nesta década surgiu o primeiro sistema de banco de dados construído baseado nos padrões SQL com a empresa Oracle através do Oracle 2 e depois com a IBM através do SQL/DS, servindo como sistema e repositório de informações de outras empresas.

Foi neste período que os softwares de banco de dados relacionais foram sendo refinados graças ao *feedback* (retorno) que os usuários destes sistemas faziam, devido ao desenvolvimento de sistemas para novas indústrias e ao aumento do uso de computadores pessoais e sistemas distribuídos.

Além dos bancos de dados relacionais surgiram os primeiros bancos orientados a objeto, distribuídos e destinados a funções específicas. Os pioneiros a implementarem orientação a objeto foram:

- O2 [1988]
- Exodus [1986]
- ORION [1986]

Na Metade da década, os bancos de dados passaram a combinar características de orientação a objeto com o modelo relacional, expandindo a arquitetura com novas possibilidades, como otimização de consultas configuráveis. Por exemplo:

- POSTGRES [1986]
- STARBURST

No final da década há a maturidade da tecnologia, onde os sistemas relacionais passaram a apresentar um desempenho aceitável. São exemplos dessa maturidade:

- DB2
- Ingres
- Oracle
- Sybase
- Informix

Além disso, há a padronização de uma linguagem para manipulação de bancos de dados, o SQL ANSI - *American National Standards Institute* [1986, 1989].

## 1.4 DÉCADA DE 90 – MATURIDADE

No início da década nota-se a maturidade que os bancos de dados atingiram com o surgimento dos primeiros Sistemas Gerenciadores de Banco de Dados – SGDB Orientados a Objeto comerciais, SGDBs paralelos, tempo real, etc, além de avanços na padronização de interfaces e interoperabilidade.

Na metade da década surgem novas classes de aplicação:

- *Data Mining*;
- Bibliotecas Digitais (Vídeo-sob-demanda, Animação);
- Hipermídia e Multimídia em geral;
- GIS (Sistemas de Informações Geográficas), Meteorologia, Física de Alta Energia (HEP).

Nos idos da década surgem novas abordagens:

- WIIS - *Web Information Integration System*: sistema para tratar dados oriundos de vários *websites*. Lidam com um grande número de *websites*, possui maior autonomia dos componentes e dados semi-estruturados.

- Enfoque de *Data Warehouse* (armazém de dados): dados são extraídos das fontes e armazenados em uma *warehouse*(*Armazém*).

- Enfoque de Multi-SGBD: os dados são mantidos nos Web sites, na qual as consultas são decompostas e enviadas aos vários *websites*.

Desde sua chegada, os bancos de dados tem tido aumento nos dados de armazenamento, desde os 8 MB (Megabytes) até centenas de petabytes de dados em listas de e-mail, informações sobre consumidores, sobre produtos, vídeos, informações geográficas, etc. Com este aumento de volume de dados, os sistemas de bancos de dados em operação também sofreram aumento em seu tamanho.

Um dos projetos mais ambiciosos de banco de dados já visto está no CERN (Conselho Europeu para Pesquisa Nuclear) que consiste em um banco de dados distribuído com a capacidade de armazenamento na casa dos Hexabytes (1 Hexabyte = 1,000 Petabytes =  $1 * 10^{18}$  Bytes) de dados.

O padrão SQL passou da IBM para a ANSI (*American National Standards Institute* - Instituto Nacional Americano para Padrões) e para a ISO, os quais formaram um grupo de trabalho para continuar o desenvolvimento. Este desenvolvimento ainda acontece com outras novas versões dos padrões definidos.

## 1.5 ONDE ESTAMOS?

Atualmente, é possível afirmar que a informação gira em torno dos bancos de dados.

O site *Business Intelligence Lowdown* publicou uma matéria em 2007 onde lista os 10 maiores bancos de dados do mundo e nos mostra o porquê desta afirmação:

### a) GOOGLE

Ainda não há muito conhecimento sobre a verdadeira dimensão do banco de dados do Google, mas basta considerar que a empresa mantém em *cache* quase todas as páginas da Internet, tem mais de 51 milhões de usuários no Gmail, armazena imagens do mundo todo para o *Google Earth* e para o *Google Maps*, tem milhões de cadastrados no Orkut, entre outros. Isso tudo sem contar que o YouTube faz parte de seu leque de serviços. Não se sabe o quanto de informações o Google armazena, mas estima-se que a empresa ultrapassa tranquilamente a casa dos petabytes (1 petabyte = 1024 terabytes);

Números: 91 milhões de pesquisas por dia representa 50% de todas as buscas Internet Virtual perfis de inúmeros número de usuários.

### b) AT & T

Semelhante a *Sprint*, os Estados Unidos “mais antiga empresa de telecomunicações AT & T mantém um dos maiores bancos de dados. Arquitetonicamente falando, a AT & T a maior base de dados é a “nata” da cultura, uma vez que ostenta títulos, incluindo o maior volume de dados em uma única base (312 terabytes), o segundo maior número de linhas em uma única base (1,9 trilhão), que inclui a AT & T’s extensa chamada de registros.

Números: 323 terabytes de informação 1,9 trilhões telefonemas em registros.

### c) *World Data Centre for Climate*

Acondicionado em um supercomputador de 35 milhões de euros usado para investigação extensa sobre o clima, capaz de dar a resposta para o aquecimento global. O *World Data Centre for Climate* (WDCC) como o nome indica, o WDCC é uma entidade que faz pesquisas climáticas no mundo todo. Essa é uma das áreas que mais exigem processamento e capacidade de armazenamento de dados. O WDCC possui 220 terabytes de dados facilmente acessíveis na Internet, bem como seus 110 terabytes (ou 24.500 DVD’s) usados para simulação de dados, e seis PETABYTES de dados internos da empresa.

Números: 6 PETABYTES é igual à aproximadamente 1.343.296 DVDs.

O que podemos concluir com isso:

- Trivialização do uso da tecnologia de banco de dados?
- Proliferação de produtores e consumidores de dados?
- Aplicações armazenando da ordem de petabytes ou mais?

## 1.6 E O FUTURO?

Fica a indagação para a quantidade de dados existente no mundo. É sabido que é necessário um novo modelo de armazenamento dados sob pena da escassez de meios para guarda de dados.