



Escola Superior de Tecnologia e Gestão
Instituto Politécnico da Guarda



Escola Superior de Tecnologia e Gestão
Instituto Politécnico da Guarda

INTRODUÇÃO ÀS BASES DE DADOS

José Carlos Fonseca





BASES DE DADOS

- ❑ Dados – Representação de factos do mundo. Ex. a altura de uma pessoa
- ❑ Informação – Dados contextualizados e mais estruturados. Ex. as várias características de uma pessoa
- ❑ Conhecimento – O nosso modelo do mundo. A forma como a realidade é reflectida no nosso pensamento. Ex. como representamos internamente essa pessoa
- ❑ Inteligência – Capacidade de sentir o ambiente, tomar decisões e controlar a ação. Ex. como reagimos com essa pessoa



BASES DE DADOS

Uma BD é essencialmente um repositório de informação de um modo organizado, metódico e seguro, permitindo formas eficazes de manutenção e consulta

O PROBLEMA

- ❑ Desde o início da era dos computadores que o armazenamento e manipulação de dados foi um dos objectivos, já que é também um aspecto fundamental das organizações
- ❑ A informação pode encontrar-se nos mais variados formatos e na sua forma mais simples pode ser armazenada em ficheiros de texto usando o sistema de ficheiros do computador
- ❑ Por exemplo, cada um destes ficheiros de texto poderá conter os diversos registos de uma determinada entidade, por exemplo colocando cada um numa linha diferente. Isto é o que se chama “flat file”.

O PROBLEMA

- ❑ Considerando uma empresa genérica:
 1. Possui uma **grande quantidade de dados** (2 TB) sobre empregados, produtos, vendas, etc.
 2. Os dados são **accedidos simultaneamente** por vários empregados
 3. As questões sobre os dados têm de ser **respondidas rapidamente**
 4. As alterações aos dados feitas pelos diferentes utilizadores devem ser **aplicadas com consistência**
 5. O acesso a alguns dados deve ser **restrito**
 6. Se houver problemas como se conseguem **recuperar os dados**

O PROBLEMA

1. Não temos 2 TB de memória, pelo que é necessário guardar os dados em disco. Como garantir a **escalabilidade**?
2. O acesso simultâneo de vários utilizadores fica lento por falta de optimização dos recursos. Pode-se ler algo que está a ser alterado? Que alterações são visíveis?
3. É necessário escrever código para responder a cada uma das questões sobre os dados. A **pesquisa sequencial** de dados em ficheiros é **lenta**. Como torná-la mais rápida?



O PROBLEMA

4. Temos de **proteger os dados** de alterações que o tornem inconsistentes. Como inserir simultaneamente dados de dois utilizadores? Como garantir a unicidade dos dados?
5. Tem de haver **mecanismos de segurança** que permitam que diferentes utilizadores consigam aceder a diferentes dados. O sistema operativo não permite isso. Desenvolver código? Será esse mesmo código seguro? Será eficiente?
6. Tem de haver mecanismos que permitam a **recuperação de dados** em caso de falha do sistema. Se um utilizador estava no meio de uma operação, será que essa operação foi concluída? Será que foi só parcialmente? Que dados foram modificados?



O PROBLEMA

A dificuldade de garantir todas estas características, sem um sistema informático especificamente desenvolvido para o efeito é notória!



A SOLUÇÃO

- ❑ Para se resolverem os problemas relacionados com a gestão da informação foram criadas as **Bases de Dados (BD)**
- ❑ Para se gerirem as BD podem ser usados sistemas de ficheiros e o software por medida usado para os gerir, que não conseguem resolver convenientemente os problemas. Para tal foram criados os **Sistemas Gestores de Bases de Dados (SGBD)**
- ❑ Um SGBD é um software especificamente desenhado para assistir na manutenção e utilização de grandes quantidades de dados

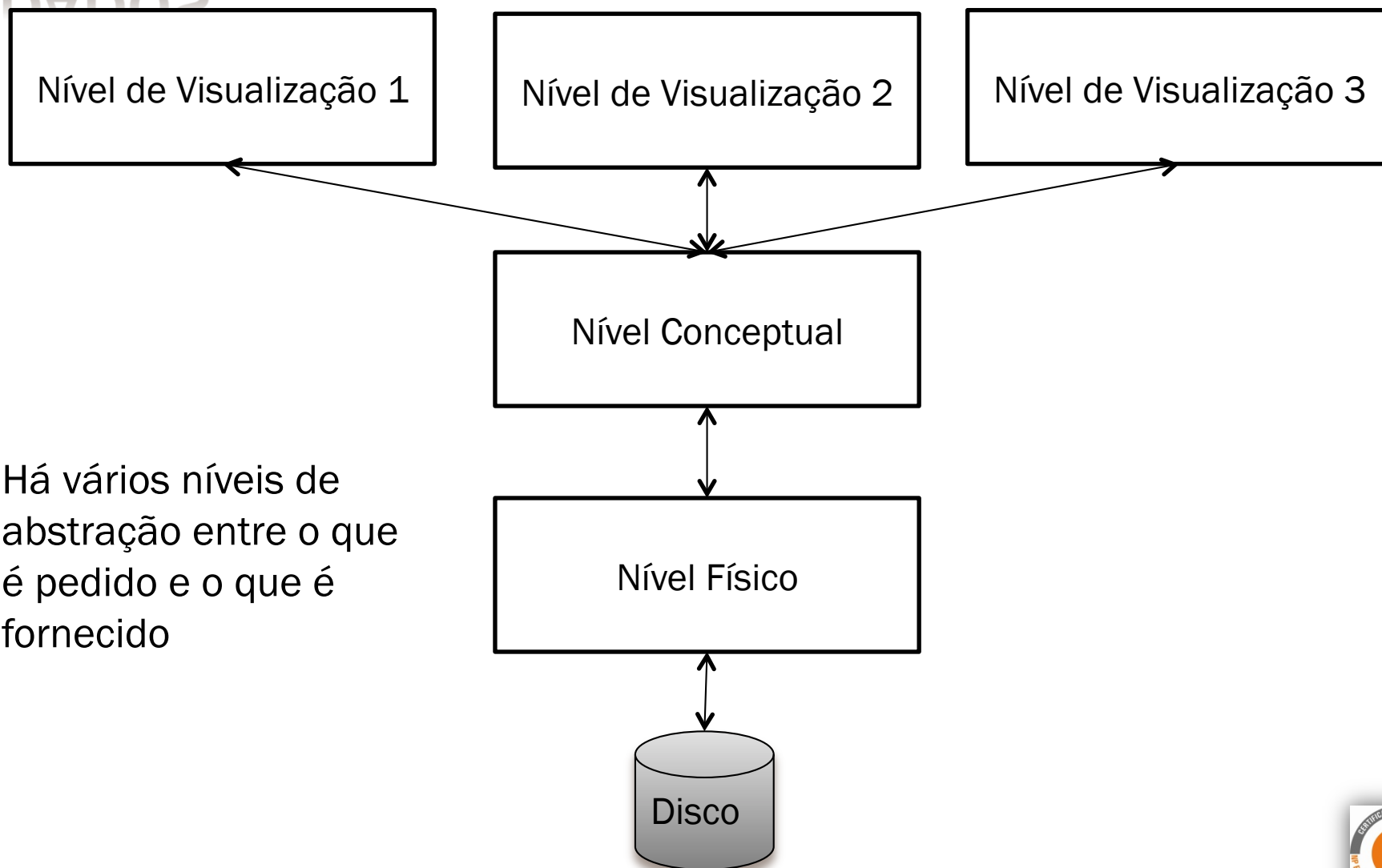


A SOLUÇÃO

□ Vantagens de um SGBD:

1. **Independência dos dados:** 3 níveis de abstração dos dados fornecem independência lógica e física
2. **Acesso eficiente aos dados**
3. **Integridade dos dados e segurança**
4. **Administração de dados**
5. **Acesso concorrenciais e recuperação em caso de falha**
6. **Diminuição do tempo de desenvolvimento das aplicações**

3 NÍVEIS DE ABSTRAÇÃO DOS DADOS



Há vários níveis de abstração entre o que é pedido e o que é fornecido



3 NÍVEIS DE ABSTRAÇÃO DOS DADOS

- ❑ **Nível Conceptual.** Num modelo relacional, descreve as relações: as entidades e os respectivos relacionamentos.
- ❑ **Nível Físico.** Descreve como as relações descritas no nível conceptual são de facto armazenados pelo SGBD no suporte (disco, SSD, etc.), os índices usados, os parâmetros de configuração.
- ❑ **Nível de Visualização.** Descreve parcialmente os dados, de acordo com as necessidades de cada utilizador ou grupo de utilizadores. Consiste nas Views e seus relacionamentos.

3 NÍVEIS DE ABSTRAÇÃO DOS DADOS

- ❑ **Independência lógica.** Se houver alterações no nível conceptual, o nível de visualização pode ser modificado de forma a que as relações sejam calculadas como anteriormente
- ❑ **Independência física.** O nível conceptual isola os utilizadores das alterações feitas no armazenamento dos dados, nomeadamente na forma como são armazenados e na escolha dos índices



OS MODELOS DE DADOS

- ❑ **Modelo de dados:** conjunto de ferramentas conceptuais, para descrever os dados, a semântica dos dados e as restrições dos dados



O MODELO HIERÁRQUICO

- ❑ As BD começaram por utilizar o Modelo Hierárquico, cuja estrutura em árvore pressupõe uma relação hierarquizada dos registos dos dados.
- ❑ Tinham limitações na capacidade de pesquisas e uma estrutura muito rígida e dificilmente alterável.



O MODELO HIERÁRQUICO

- ❑ Por exemplo, desejando-se armazenar os empregados e os respectivos departamentos, poderíamos ter, seguindo o modelo hierárquico:

Empregados e Departamentos usando o Modelo Hierárquico

10, Contabilidade, 7782, Clark

20, Pesquisa, 7369, Smith, 7566, Jones, 7788, Scott

30, Vendas, 7499, Allen, 7521, Ward, 7654, Martin, 7698, Blake

40, Operações



O MODELO HIERÁRQUICO

Empregados e Departamentos usando o Modelo Hierárquico

10, Contabilidade, 7782, Clark

20, Pesquisa, 7369, Smith, 7566, Jones, 7788, Scott

30, Vendas, 7499, Allen, 7521, Ward, 7654, Martin, 7698, Blake

40, Operações

- ❑ Pesquisas de empregados muito rápidas se soubermos o departamento, mas lentas se não soubermos
- ❑ Alterações de dados problemáticas se for necessário movimento entre registos
- ❑ Tem algumas restrições: neste exemplo é possível ter um departamento sem empregados, mas não é possível ter empregados sem departamento



O MODELO EM REDE

- ❑ Posteriormente surgiu o Modelo em Rede, onde qualquer registo estava ligado praticamente a todos os outros.
- ❑ As consultas eram realizadas mais rapidamente, mas a complexidade da rede dificultava muito a sua gestão.



O MODELO RELACIONAL

- ❑ Em 1970, o Dr. E.F. Codd propôs o modelo relacional para bases de dados através de um documento intitulado "A Relational Data Model of Data for Large Shared Data Banks", publicado em Communications of the ACM(Vol. 13, No. 6, June 1970, pp. 377-87).
- ❑ Baseado na teoria dos conjuntos, o modelo relacional era a primeira forma de organizar os dados de tal forma que estes constituíssem informação que pudesse ser facilmente utilizável.



O MODELO RELACIONAL

- ❑ Entre outros artigos sobre o mesmo tema que Codd publicou encontra-se o “Twelve Rules for Relational Databases”, publicado em dois artigos da Computerworld "Is Your DBMS Really Relational?" e "Does Your DBMS Run By the Rules?" em 14 de Outubro de 1985, e 21 de Outubro de 1985, respectivamente.



VANTAGENS DO MODELO RELACIONAL

- ❑ Os sistemas relacionais têm benefícios tais como:
 1. Facilidade de acesso a os dados
 2. Facilidade na modelização de dados
 3. Redução no armazenamento de informação e redundância
 4. Independência do armazenamento físico em relação ao desenho lógico
 5. Linguagem de manipulação de dados de alto nível: SQL



VANTAGENS DO MODELO RELACIONAL

- ❑ É muito eficiente para muitos tipos de aplicações e tipos de dados, mas não para todos, nomeadamente quando são necessárias estruturas não relacionais
- ❑ É muito eficiente para aplicações de Online Transaction Processing (OLTP) e Decision Support Systems (DSS)



MODELOS ORIENTADOS A OBJECTOS

- ❑ A próxima evolução, já em curso, é o modelo orientado a objectos.
- ❑ Para além da informação com significado este modelo acrescenta o comportamento da informação.
- ❑ Juntam-se no mesmo repositório os dados e as regras de manutenção e validação.



A STRUCTURED QUERY LANGUAGE (SQL)

- ❑ Decorrendo das 12 regras de Codd aparece uma linguagem declarativa que permite definir, manipular e pesquisar os dados da BD.
- ❑ A linguagem SQL foi originalmente desenvolvida pela IBM da pela maioria dos grandes fabricantes de SGBD.



A STRUCTURED QUERY LANGUAGE (SQL)

- ❑ SQL quer dizer Structured Query Language e foi adoptada como um standard ANSI/ISO.
- ❑ Embora tenha sido revista em 1999 (referenciada como SQL99 ou SQL3) muitos vendedores ainda nem sequer adoptaram na totalidade a versão de 1992.
- ❑ O standard de 1992 é mais reduzido e simples e só algumas partes do standard de 1999 é que se encontram correntemente implementados.



A STRUCTURED QUERY LANGUAGE (SQL)

- ❑ A maioria das linguagens de computador são imperativas (muitas vezes usando programação sequencial ou então usando programação procedimental).
 - ❑ O programador indica ao computador o que deve fazer passo a passo, especificando um procedimento a realizar.



A STRUCTURED QUERY LANGUAGE (SQL)

- ❑ No SQL o programador diz o que quer obter e o computador é que determina como o irá fazer.
- ❑ Este facto tem duas vantagens:
 1. As pesquisas não dependem da representação dos dados, pelo que o SGBD pode representar os dados da forma que melhor lhe serve
 2. É mais difícil existir um pequeno bug numa query SQL do que num programa procedimental. Geralmente os dados são descritos da forma que desejamos ou então dá erro de uma forma perfeitamente visível



BASES DE DADOS COMERCIAIS - ORACLE

- ❑ Larry Ellison, Bob Miner e Ed Oates fundaram a Oracle Corp. há cerca de 30 anos atrás, em 1977.
- ❑ A Oracle Corp. é a maior vendedora mundial de sistemas de informação e a segunda maior companhia informática.
- ❑ O seu volume de negócios (tem cerca de 50% do mercado) é superior ao conjunto dos seus mais directos adversários.
- ❑ Os seus produtos encontram-se nos data centers de 98 das Fortune 100 Companies.
- ❑ O nome Oracle vem do código de um projecto da CIA no qual Ellison trabalhou quando era empregado da Ampex

[www.oracle.com]



BASES DE DADOS COMERCIAIS - SYBASE

- ❑ Sybase tornou-se o segundo maior SGBD, logo a seguir ao Oracle, depois de um acordo de partilha do código fonte com o SQL Server da Microsoft.
 - ❑ Até à versão 4.9 o Sybase e o SQL Server eram virtualmente idênticos.
 - ❑ Acabaram por decidir separar-se, embora algumas características comuns ainda se mantenham, como o Transact-SQL e arquitectura.
 - ❑ A Sybase disponibiliza vários produtos, incluindo o PowerDesigner que é um software de modelização de dados muito potente.
 - ❑ Em 2010 a Sybase foi adquirida pela SAP
 - ❑ Neste momento o Sybase terá cerca de 4% do mercado dos SGBDs.
- [www.sybase.com]



BASES DE DADOS COMERCIAIS - INFORMIX

- ❑ A Informix foi criada por Roger Sippl e Laura King em 1981.
- ❑ Chegou a ser o 3º maior SGBD, em 1995, logo a seguir à Oracle e Sybase, tendo depois decaído.
- ❑ Foi adquirido pela IBM em 2001 à Informix Software.

[www-306.ibm.com/software/data/informix/]

José Fonseca



BASES DE DADOS COMERCIAIS – DB2

- ❑ Há quem considere a DB2 como sendo o primeiro SGBD.
- ❑ É um produto da IBM e foi lançado em 1983 para o mainframe MVS da IBM.
- ❑ O DB2 tem as suas origens na década de 1970 quando o Dr. Edgar Codd, que trabalhava para a IBM, descreveu a sua teoria de bases de dados relacionais.
- ❑ Durante muitos anos o DB2 só estava disponível para mainframes da IBM, mas actualmente já corre em OS/2, Windows, Linux e PDAs.
- ❑ A IBM e o DB2 encontram-se frequentemente nos topos dos benchmarks da indústria TPC-C e TPC-H.

[www-306.ibm.com/software/data/db2/]

José Fonseca



BASES DE DADOS COMERCIAIS – SQL SERVER

- ❑ É o SGBD da Microsoft.
- ❑ Usa o Transact-SQL comum ao Sybase.
- ❑ A versão 1.0 apareceu em 1989 para o OS/2 e tinha o código comum ao Sybase 3.0.
- ❑ Tem tido um grande crescimento devido à popularidade do Sistema Operativo Windows da Microsoft

[www.microsoft.com/sql/default.mspix]

José Fonseca



COMPARATIVO

Lucros de SGBS em 2006 (Milhões de Dollars)

Companhia	2006	2006 Market Share (%)	2005-2006 Growth (%)
Oracle	7,168.0	47.1	14.9
IBM	3,204.1	21.1	8.8
Microsoft	2,654.4	17.4	28.0
Teradata	494.2	3.2	5.7
Sybase	486.7	3.2	8.2
Outros	1,206.3	7.9	5.0
Total	15,213.7	100.0	14.2

[www.gartner.com/it/page.jsp?id=507466]

José Fonseca



COMPARATIVO

Lucros de SGBS em 2010-2011 (Milhões de Dollars)

Empresa	2010	2011	2010 Share (%)	2011 Share (%)	Crescimento (%)
Oracle	9,990.5	11,787.0	48.2	48.8	18.0
IBM	4,300.4	4,870.4	20.7	20.2	13.3
Microsoft	3,641.2	4,098.9	17.6	17.0	12.6
SAP/Sybase	744.4	1,101.1	3.6	4.6	47.9
Teradata	754.7	882.3	3.6	3.7	16.9
Outros	1,315.3	1,389.7	6.3	5.8	5.7
Total	20,746.6	24,129.5	100.0	100.0	16.3

[itknowledgeexchange.techtarget.com/eye-on-oracle/oracle-the-clear-leader-in-24-billion-rdbms-market/]



BASES DE DADOS LIVRES- POSTGRESQL

- ❑ O PostgreSQL é por muitos considerado a mais avançada BD de código aberto.
- ❑ Corre na generalidade dos Sistemas Operativos (SO).
- ❑ É completamente compatível com ACID e tem chaves estrangeiras, chaves primárias, joins, triggers e procedimentos.
- ❑ O PostgreSQL começou por ser o projecto de investigação Ingres da Universidade de Berkeley, na Califórnia.
- ❑ Posteriormente evoluiu chamando-se Post-Ingres tornando-se no PostgreSQL. A primeira versão foi lançada em 1989. É distribuído sob a licença BSD.

[www.postgresql.org/]



BASES DE DADOS LIVRES- MYSQL

- ❑ O MySQL tem, actualmente, mais de 11 Milhões de instalações (
<http://www.informationweek.com/news/showArticle.jhtml?articleID=206900327>).
- ❑ O MySQL era propriedade de uma companhia Sueca, MySQL AB, que foi adquirida pela Sun Microsystems em 2008 que por sua vez foi adquirida pela Oracle em 2010.
- ❑ O MySQL é distribuído sob a licença GNU General Public Licence.

[www.mysql.com/]



BASES DE DADOS LIVRES- MYSQL

- ❑ É uma BD muito popular em aplicações web sendo um dos componentes das plataformas LAMP, MAMP e WAMP.
- ❑ A sua popularidade está associada à popularidade do PHP e Ruby on Rails.
- ❑ Muitas aplicações Content Management Systems (CMS) populares (Drupal, PHP Nuke, WordPress, Joomla!, etc.) correm em PHP e MySQL.
- ❑ É criticada por algumas divergências em relação ao standard SQL no que diz respeito ao tratamento de NULL e valores por omissão. Nas últimas versões já possui triggers, procedimentos e views.

[www.mysql.com/]



BASES DE DADOS LIVRES- SQLITE

- ❑ O SQLite é uma biblioteca que implementa um motor de BD transaccional auto-contido, sem necessidade de servidor e sem necessidade de configuração.
- ❑ É considerado como sendo a BD mais usada no mundo inteiro.
- ❑ É usada em aplicações informáticas, bem como em electrónica de consumo incluindo telemóveis, PDAs e leitores MP3.
- ❑ O código fonte é livre e do domínio público.

[www.sqlite.org/]



BASES DE DADOS LIVRES- SQLITE

- ❑ O SQLite implementa a grande maioria do standard SQL-92, incluindo as propriedades das transacções (atomicidade, isolamento e persistência).
- ❑ O código fonte consiste em 74 ficheiros em C, mas também existe uma versão chamada "Amalgamation", na qual todo o código se encontra num único ficheiro em C.
- ❑ Este ficheiro contém tudo o que é necessário para desenvolver aplicações usando o SQLite e ainda o torna mais rápido.
- ❑ Também existe um wrapper para Java.
- ❑ O SQLite ocupa apenas 256KBytes podendo ficar a ocupar somente 170KBytes desabilitando algumas características durante a compilação.
- ❑ Para grandes BD o SQLite necessita de 256Bytes de memória por cada 1MByte da BD.

[www.sqlite.org/]



BASES DE DADOS LIVRES- JAVADB

- ❑ JavaDB é a distribuição da Sun Microsystems da BD transaccional Apache Derby, da Apache Software Foundation.
- ❑ É uma BD leve, já que ocupa em memória somente 2MBytes, podendo ser usada em dispositivos móveis.
- ❑ O Apache Derby foi lançado pela Cloudscape em 1997. Em 1999 a Informix Software Inc. adquiriu a Cloudscape. Em 2001 a IBM adquiriu a Informix Software Inc. Em 2004 a IBM forneceu o código à Apache Software Foundation como Derby.
- ❑ A Sun Microsystems juntou-se ao projecto com o intuito de o desenvolver nos seus produtos, sendo agora distribuída por ela como JavaDB.
- ❑ Agora é da Oracle, desde que a Sun Microsystems foi adquirida pela Oracle em 2010.

[www.oracle.com/technetwork/java/javadb/overview/index.html]



BASES DE DADOS LIVRES- ORACLE DATABASE 11G EXPRESS EDITION

- ❑ Oracle Database XE é fácil de instalar e gerir. Pode o usar-se o seu interface baseado em browser web, o Database Home Page, para administrar a base de dados; criar tables, views, e outros schema objects; importar, exportar, e ver os dados das tabelas; executar queries e scripts SQL; e gerar reports.
- ❑ O Oracle Database XE inclui o Oracle HTML DB 2.1, que é um ambiente de desenvolvimento para criar aplicações web baseadas em bases de dados.
- ❑ Também se podem usar outras linguagens populares, tais como o Visual Studio.NET ou o PHP para desenvolver as aplicações.

[www.oracle.com/technetwork/products/express-edition/overview/index.html]



BASES DE DADOS LIVRES – ORACLE DATABASE 11G EXPRESS EDITION

- ❑ O Oracle Database XE pode ser instalado numa máquina de qualquer tamanho e com qualquer nº de CPUs (uma base de dados por máquina), mas o XE só permite armazenar até 4GB de dados, só usa até 1GB de memória e só usa um CPU do servidor.

[www.oracle.com/technetwork/products/express-edition/overview/index.html]



BASES DE DADOS LIVRES- ORACLE DATABASE 11G EXPRESS EDITION

- ❑ Oracle Database XE é aconselhado para:
 - ❑ Programadores que trabalham em PHP, Java, .NET, XML, e aplicações Open Source
 - ❑ DBAs que precisam de uma base de dados inicial gratuita para treino e colocação em funcionamento
 - ❑ Independent Software Vendors (ISVs) e hardware vendors que necessitam de uma base de dados que possam distribuir livre de encargos
 - ❑ Instituições de ensino e estudantes que necessitam de uma base de dados livre de encargos no seu curriculum

[www.oracle.com/technetwork/products/express-edition/overview/index.html]



BIBLIOGRAFIA

- ❑ Pereira, José L., Tecnologia de Bases de Dados, FCA, 1998
- ❑ Ramakrishnan, R., Gehrke, J., Database Management Systems, Third Edition, McGraw-Hill, 2007
- ❑ Mullins, Craig S., Database Administration: The Complete Guide to Practices and Procedures Addison Wesley, 2002
- ❑ [en.wikipedia.org/wiki, wikipedia](http://en.wikipedia.org/wiki/wikipedia), 2008