

Desenvolvimento de uma Ferramenta de Monitoramento de Espaço em Disco de Servidores de Banco de Dados

Fernando H. Tonon¹, Jean D. H. M. Andreazza¹

¹Faculdade de Tecnologia de Bauru (Fatec)
Rua Manoel Bento Cruz, nº 3-30, Centro, Bauru/SP – Brazil

tonon@fc.unesp.br, jeandaniel@bol.com.br

Abstract. *This paper describes the development of the DBCare tool, which purpose is the monitoring of database systems focusing on the disk space use, foreseeing the approximated date in which the disk will be full, making possible to acquire more disks or to free space in the existing drives in time. By the end of the paper, it's verified that the application achieves its goal by determining an approximate date of the full utilization of the disc, alongside with a graph and a table showing the evolution of the size of the monitored database in a specific moment.*

Resumo. *Este artigo tem como objetivo descrever o desenvolvimento da ferramenta DBCare, cuja meta é monitorar sistemas de banco de dados com foco na utilização de espaço em disco, prevendo a data aproximada em que o disco estará cheio para que seja possível providenciar mais discos ou limpeza destes em tempo hábil. Ao final do artigo, verifica-se que a aplicação consegue determinar uma data aproximada de utilização total do disco, além de um gráfico e de uma tabela demonstrando a evolução do tamanho do banco de dados monitorado em um determinado momento.*

1. Introdução

A administração de bancos de dados é uma tarefa complexa, que dispensa tempo, e a utilização de ferramentas para agilizá-la tem se tornado fundamental para administradores que lidam com diversas bases de dados, que, na maioria das vezes, são heterogêneas e fisicamente distantes.

Na gama de informações a que o administrador de bancos de dados (database administrator) deve manter-se atento, está a disponibilidade de espaço de armazenamento, que, além de variar bastante, é difícil de mensurar constantemente, visto que em grandes bases de dados, informações são consultadas, acrescentadas e removidas (embora em escala menor à anterior) e são gerados logs de transações que ininterruptamente.

Assim sendo, uma ferramenta para monitorar o estado dos discos e prever possíveis necessidades de expansão torna-se crucial para o bom funcionamento do banco de dados, pois, além de atividades de manutenção, pode ser necessário providenciar a compra de unidades de armazenamento e nem sempre o orçamento permite ou o fator tempo não é favorável.

Tal ferramenta deve estar disponível on-line para que, de onde estiver, o Database Administrator (DBA) possa monitorar o estado do armazenamento do banco e solicitar providências administrativas ou técnicas, mesmo à distância.

2. Referencial Teórico

Qualquer organização que necessite da automatização de seu banco de dados usará alguma tecnologia de banco de dados. As vantagens dos bancos de dados são oferecer acessibilidade, disponibilidade, segurança e gerenciabilidade [Zeis, 2009].

O termo bancos de dados remete a quaisquer coleções de itens de dados, como por exemplo, uma lista telefônica, uma folha de pagamento ou o cardápio de um restaurante. Entretanto, na tecnologia da informação, existem diversas definições para o termo, mas para Taylor (2003), bancos de dados podem ser definidos como sendo coleções auto descritivas de registros integrados.

Apesar de existirem vários tipos de bancos de dados, cada um deles baseado num modelo conceitual diferente, o modelo mais utilizado de armazenamento de dados, atualmente, é o modelo relacional, que envolve um conceito em que os dados são organizados de forma numa estrutura lógica, em forma de tabelas. Estas tabelas são formadas por colunas, cada qual armazenando determinados tipos de dados, que são chamados de registros. Um registro é a representação de um objeto físico ou conceitual, na tabela, ele é representado por uma linha. Cada registro possui atributos (representados pelas colunas) e um atributo de um registro é o dado [Taylor, 2003].

Outras informações importantes que os bancos de dados armazenam, são os chamados metadados. Estes descrevem a estrutura dos dados dentro um banco de dados. Desta forma, os bancos de dados são auto descritivos, pois eles possuem uma estrutura que descreve a forma como a informação está organizada. As informações de metadados são armazenadas no dicionário de dados.

O conceito relacional também necessita que as tabelas se relacionem umas com as outras, formando uma estrutura integrada, sendo utilizadas, para isto, as chaves primárias e estrangeiras [Zeis, 2009].

2.1. O Padrão SQL

A linguagem Structured Query Language (SQL) foi desenvolvida por pesquisadores da IBM para acesso aos Sistemas Gerenciadores de Bancos de Dados (SGBDs) de forma padronizada. Sua primeira versão foi lançada em 1986, o SQL-86. Foi devido a essa definição de padrão para a linguagem que os SGBDs se tornaram um sucesso [Costa, 2006].

Esta linguagem foi desenvolvida com o propósito de permitir que a criação e manutenção de bancos dados, a inserção, manipulação e destes dados e a seleção de partes destes dados. Recentemente, foi incorporado no padrão SQL o modelo de estruturação a objetos, tornando-se uma estrutura híbrida e, sendo classificado, então, como banco de dados objeto-relacionais [Taylor, 2003].

Os comandos da SQL podem ser categorizados em sub-linguagens, sendo as principais: Linguagem de Manipulação de Dados (LMD ou Data Manipulation Language - DML) e a Linguagem de Definição de Dados (LDD ou Data Definition Language - DDL), entre outras.

2.2. DDL

A Linguagem de Definição de Dados é utilizada para criar, alterar ou excluir os elementos básicos de um banco de dados relacional, ou seja, tabelas, visões, esquemas e outras estruturas dos bancos de dados. Os principais comandos da DDL incluem CREATE, ALTER E DROP [Taylor, 2003].

2.3. DML

Esta é a parte do SQL que realiza as operações com os dados, inclui os comandos responsáveis por inserir, alterar e excluir dados e os comandos de seleção de dados que, tanto por sua função quanto por sua complexidade, podem ser considerados os comandos mais importantes da DML. Podemos listar, entre os principais comandos, INSERT, UPDATE, DELETE e SELECT. Apesar de o comando SELECT parecer simples, pode gerar grandes linhas de código incluindo múltiplas expressões que, por sua vez, podem incorporar uma variedade conectores lógicos, funções agregadas e sub-consultas [Taylor, 2003].

2.4. A Linguagem de Programação Java

De acordo com Lemay (2005), o Java surgiu em meados dos anos 90 com a finalidade de substituir o C++ como linguagem para um projeto de TV interativa da Sun. Apesar de não ter tido sucesso no projeto original, o Java despontou como linguagem popular para o desenvolvimento de páginas web interativas. Até o momento, estima-se que existam mais profissionais Java do que programadores C++ nos dias de hoje.

O Java é uma linguagem de programação orientada a objetos segura, que pode ser executada em qualquer plataforma desde que haja a máquina virtual Java instalada. Isto porque as aplicações Java são compiladas em bytecode e podem ser executadas diretamente pelo interpretador e, este sim, deve ser compatível com o sistema operacional local [Lemay, 2005].

2.5. Sistemas Gerenciadores de Bancos de Dados

Os Sistemas Gerenciadores de Bancos de Dados (SGBDs) são conjuntos de programa usados para definir, administrar e processar bancos de dados e suas aplicações associadas. Existem vários bancos de dados no mercado atualmente, inclusive soluções gratuitas e é importante que o DBA saiba escolher a que mais se adapta à resolução do seu problema.

No entanto, para o usuário, a comunicação com o banco de dados normalmente é transparente. Isso se dá ao fato do SGBD intermediar essa comunicação, ficando a cargo dele cuidar dos detalhes físicos do armazenamento de banco de dados deixando para a aplicação a necessidade de entender apenas a parte lógica dos dados e não como eles são armazenados [Taylor, 2003].

2.6. O SGBD MySQL

MySQL é um sistema de bancos de dados relacional de baixo custo, pois pode ser baixado gratuitamente da Internet. Além disso, é mais rápido e confiável do que alguns sistemas de banco de dados comerciais como Oracle e DB2. É notável que o uso do MySQL está crescendo continuamente e seus usuários estão satisfeitos. Apesar disso tudo, ele ainda não possui algumas características que outros sistemas possuem há tempos [Kofler, 2001].

Entre essas características, Kofler (2001) menciona: arquitetura cliente-servidor; suporte a linguagem SQL; suporte a variadas linguagens de programação no lado cliente; interface ODBC; independência de plataforma; e a velocidade.

2.7. Máquinas Virtuais e Virtualização

A virtualização oferece a flexibilidade e a facilidade de manutenção, tanto de hardware como de software, necessárias para a maior parte das organizações mundiais. Além disso, oferece a rapidez e robustez necessária para aperfeiçoar e adequar os novos processos de TI que podem surgir ao longo do tempo [Veras, 2011].

O conceito de virtualização envolve desvincular o sistema operacional e as aplicações do ambiente físico. Assim sendo, servidores virtuais são mais facilmente instalados, gerenciados e migrados do que servidores reais, facilitando, inclusive a preparação de ambientes de teste e desenvolvimento, tão necessários para manutenção de aplicações empresariais.

De acordo com o International Data Corporation (IDC), em média 85% da capacidade dos servidores empresariais se mantém ociosa, contata-se, então, um grande desperdício de capacidade de processamento e memória destas máquinas. Com a virtualização, é possível aproveitar estes recursos através da consolidação de servidores gerando, inclusive, economia de energia, espaço físico e facilitando o gerenciamento do ambiente de TI em geral [Veras, 2011].

2.8. Linux

Originário do UNIX, sistema desenvolvido em 1969 por um grupo de projetistas da Bell Labs, o Linux foi criado nos anos 1990 por Linus Torvalds e rapidamente garantiu seu espaço como o sistema operacional de código aberto mais utilizado no mundo. Seu sucesso se deve ao fato de, além de ser registrado na General Public Licence (GPL), sempre foi apoiado pelos usuários, que auxiliavam no seu desenvolvimento [Farias, 2004].

A sua proposta principal é ser uma alternativa barata e funcional em relação ao UNIX, da qual se originou, principalmente por permitir a utilização em processadores x86/x64 comuns. Há várias distribuições Linux disponíveis no mercado, algumas gratuitas e outras proprietárias, cujo foco variam muito. As principais são Debian, Redhat, CentOS, SUSE, Slackware e Ubuntu [Farias, 2004].

O Linux possui várias características que o tornam atrativo como sistema operacional de servidor, entre elas, podemos destacar: ser multitarefa e multiusuário; gratuito; não exige uma arquitetura robusta para rodar; é um sistema estável; possui código fonte aberto [Farias, 2004].

3. Desenvolvimento de Aplicação

3.1. Determinação das tabelas e respectivos relacionamentos para a base de dados

A principal necessidade de armazenamento desta aplicação é dos dados de espaço em disco livre num determinado momento. Porém, outros dados são igualmente importantes para fins estatísticos, de detecção de anomalias e tuning de banco de dados. Todos estes dados podem se obtidos pelo administrador do sistema via sistema operacional ou consultas ao SGBD, porém como a intenção deste trabalho é centralizar estas informações numa única aplicação tornou-se necessário a idealização de uma estrutura de dados que permitisse tal armazenamento de forma objetiva.

Foram criadas três tabelas principais: a primeira para cadastro dos servidores a serem monitorados, a segunda para cadastro dos bancos de dados a serem monitorados nestes servidores e, enfim, a tabela de cadastro do tamanho do banco de dados e espaço livre no servidor em um momento específico.

Além destas tabelas, também foram criadas tabelas de apoio, com informações sobre os sistemas de arquivos, tipos de disco e esquemas RAID para cadastro. Estas tabelas são meramente para organização dos dados a serem exibidos pelos formulários.

3.2. Máquinas do ambiente de teste

O sistema foi instalado numa máquina real, Windows 7, com MySQL, phpmyadmin e Tomcat configurados. Nesta máquina real é que está sendo executado o sistema. Com a ferramenta VirtualBox, foi criada uma máquina virtual com MySQL instalado num sistema Ubuntu versão 12.04 LTS.

A fim de gerar dados o suficiente para a análise do consumo de espaço em um tempo hábil à realização do projeto, foi utilizada uma aplicação de desenvolvimento próprio, para inserir grande quantidade de informações no banco de dados em um curto espaço de tempo.

3.3. Desenvolvimento da aplicação

Tendo a proposta de prover um ambiente web e a necessidade da plataforma Java, optou-se por desenvolver a aplicação em jsp, visto que é uma plataforma homologada pela Sun e suportada pelo NetBeans, IDE (Integrated Development Environment, ou Ambiente de Desenvolvimento Integrado).

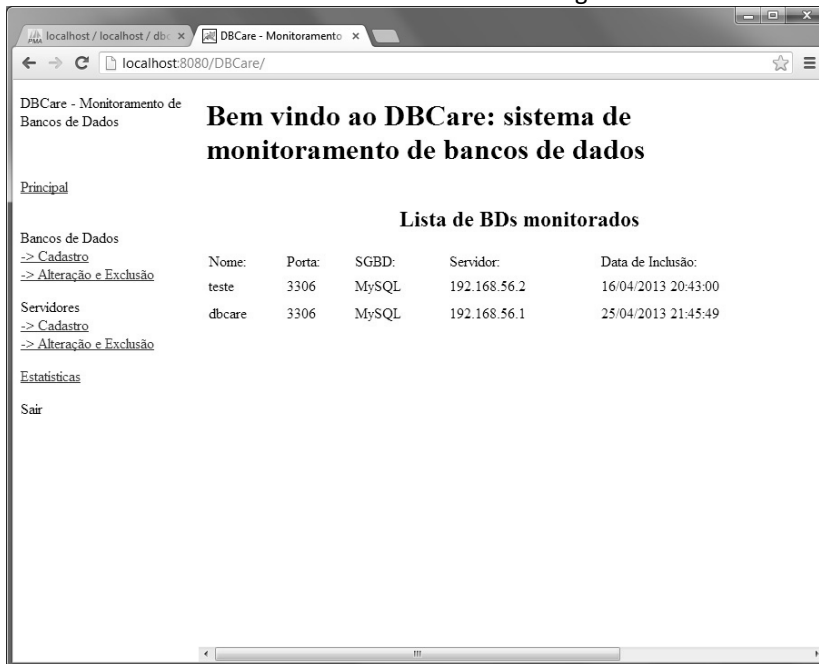


Figura 1. Tela do aplicativo DBCare

Nota-se que o NetBeans e seu ambiente de desenvolvimento foram utilizados apenas para o desenvolvimento do código Java e da conexão do banco de dados, o ambiente gráfico da aplicação e determinação de layout foram desenvolvidos utilizando uma ferramenta de desenvolvimento web gratuita, com o intuito de agilizar a organização dos elementos gráficos e textuais na tela.

A aplicação foi organizada em duas funções específicas, o cadastro, alteração e exclusão de bancos de dados e o monitoramento destes.

O cadastro, alteração e exclusão, foram organizados em formulários específicos para bancos de dados e servidores. Um banco de dados só pode ser incluído para monitoramento se houver um servidor cadastrado. Da mesma forma, qualquer servidor só poderá ser excluído, via aplicação, se não houver banco de dados relacionado à ele.



Figura 2. Tela do aplicativo DBCare mostrando tentativa de exclusão de servidor atrelado a banco de dados

A segunda função da aplicação, que é o monitoramento dos bancos de dados cadastrados é acessada a partir da página principal ou através do link “estatísticas” na barra lateral esquerda da página web da aplicação. Ao acessar esta página, são exibidos dados dos bancos de dados ativos, incluindo os servidores em que estão cadastrados.

Ao selecionar um destes bancos de dados, o usuário acessa uma página em que são exibidas as informações de monitoramento de disco do sistema de banco de dados em geral, onde o tamanho de disco total se refere ao tamanho total da partição onde está armazenado o banco de dados e o espaço utilizado se refere ao espaço total utilizado pela aplicação de banco de dados na máquina. Segue-se este padrão pois nem sempre é possível saber quantos bancos de dados e/ou aplicações estão instaladas na máquina remota.

As informações exibidas nesta página de estatísticas incluem o nome do banco de dados, tamanho total da partição onde ele está instalado, o tamanho total ocupado pelos dados armazenados no banco de dados, espaço livre remanescente no disco, momento da medição, incluindo data e hora, espaço de tempo desde a medição inicial, consumo total de disco e a diferença de consumo em relação à medição anterior. Estas informações ficam armazenadas no banco de dados da própria aplicação, que fica armazenado na mesma máquina em que está rodando a aplicação. Além destas informações, é exibida uma previsão do momento quando o disco estará cheio. Esta previsão é calculada baseando-se na diferença de tamanho entre a primeira medição e a última medição realizada. Mais especificamente, multiplica-se o valor do tamanho de espaço em disco livre pela divisão do número resultante da subtração do timestamp da última data de medição pelo timestamp da primeira data de medição pelo valor do tamanho de disco usado. Do valor resultante, adiciona-se o valor do timestamp do momento atual e obtém-se um valor de timestamp que, após convertido para formato de data e hora, retorna a data em que o disco possivelmente estará cheio.

A fórmula utilizada está melhor explicada abaixo:

$$\begin{aligned} & \textit{timestamp do tempo adicional} \\ & \textit{para encher a partição} \\ & = \textit{espaço livre atual} \times \frac{\textit{timestamp da última medição}}{\textit{espaço consumido na última medição}} \end{aligned}$$

Figura 3. Equação utilizada para calcular o tempo adicional para uso completo da partição

$$\textit{timestamp do tempo adicional para encher a partição} + \textit{timestamp do momento em que a partição estará cheia} = \textit{timestamp do momento em que a partição estará cheia}$$

Figura 4. Equação utilizada para determinar o o momento de uso completo da partição

Estas informações são disponibilizadas por um aplicativo agente, a ser executado na mesma máquina onde está hospedado o sistema DBCare. Este aplicativo é responsável pelo acesso aos bancos de dados, baseando-se nas informações fornecidas pelo usuário, e captura e armazenamento das informações de tamanho do banco de dados.

3.3.1. Ferramenta de desenvolvimento de gráficos

A fim de tornar a visualização da evolução do tamanho do banco de dados mais intuitiva, foi utilizada a biblioteca Java JFreeChart, que permite a criação de gráficos e sua exibição na aplicação Java.

A ferramenta possui vários tipos pré-definidos de gráficos, como gráficos em barra, pizza, linha, entre outros. Na aplicação DBCare, para melhor visualização do conteúdo, foi utilizado a opção de desenvolver gráficos de linha.

Para que o gráfico fosse construído e exibido corretamente no ambiente web, foi necessário gerar um arquivo de imagem, do tipo Portable Network Graphics (extensão PNG), para então exibi-lo na página de estatísticas.

3.4. Funcionamento do aplicativo agente

O aplicativo agente se mantém em execução contínua e possui um timer que, a cada cinco minutos, executa uma rotina que verifica quais bancos de dados estão cadastrados no sistema e acessa os seus respectivos servidores, consultando informações de espaço ocupado pelos diversos bancos de dados. Entre as informações resgatadas, estão o espaço ocupado pelo banco de dados selecionado, o espaço livre na partição de dados e o momento da medição.

É importante mencionar que o espaço livre da partição de dados é obtido através de console, com conexão ssh no ambiente MySQL e através do próprio SGBD no ambiente Oracle. As informações de espaço ocupado pelo bancos de dados são obtidas através de consultas específicas para verificar o tamanho de todos os bancos de dados existentes no servidor e, em nível de aplicação, calcula-se o total destes.

O comando ssh utilizado para obter a informação de uso do disco é o "`df /var | grep / | awk '{print $3}'`". Nota-se que é suposto que a instalação do MySQL tenha sido a padrão, alocando os arquivos físicos do banco de dados na partição /var. Além disso, alterando-se o parâmetro de impressão de 3 para 2, obtém-se o tamanho total da partição e com o parâmetro alterado para 4, é possível obter o espaço disponível em disco.

Utiliza-se o parâmetro em 2 para obter-se o tamanho total da partição no momento de cadastro do banco de dados. Evita-se, assim, que o usuário cadastre um tamanho incorreto e a informação fornecida pela aplicação seja inconsistente.

O aplicativo, tendo estas informações disponíveis, armazena-as no servidor do programa agente no banco de dados do DBCare, onde estão, inclusive, as informações dos outros servidores cadastrados, nada é gravado nos sistemas monitorados.



Figura 5. Tela do aplicativo DBCare Agent

4. Resultados

No ambiente de teste foi instalado um banco de dados com informações fictícias e aleatórias que são geradas por um outro aplicativo, desenvolvido especificamente para este fim.

Apesar de se tratar de uma pequena quantidade de dados, foi possível obter uma previsão de disco cheio em ambas as máquinas.

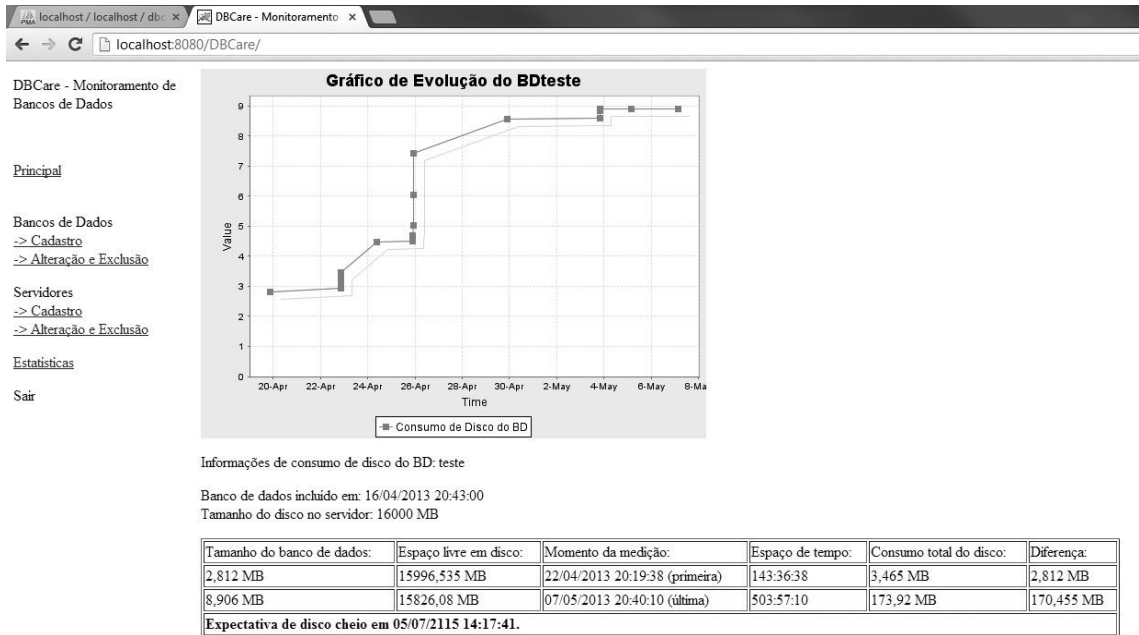


Figura 6. Tela do aplicativo DBCare mostrando estatísticas de um banco de dados monitorado

Nota-se, através da figura 4, que houve um crescimento de 6,094 MB desde a primeira medição exibida e a última exibida, comparando os dados iniciais, onde havia sido constatado o tamanho de 2,812 MB no dia 22 de abril e a informação da última medição que apresenta o tamanho de 8,906 MB, podemos concluir que o banco de dados quase triplicou de tamanho no espaço de aproximadamente quinze dias. Apesar disso, o consumo de disco saltou de 3,465 MB para consideráveis 173,92 MB. Isso pode ser justificado pelo fato de a partição /var da máquina analisada, que é o ponto de montagem onde foi instalado o banco de dados MySQL monitorado pode ter recebido outros arquivos, como por exemplo arquivos de log, tão comumente armazenados nestas partições.

Por efeito de facilidade de visualização, foi optado por exibir apenas o primeiro e o último registros, ao invés de todos eles.

A previsão apresentada é distante, visto a pequena quantidade de dados envolvidos, num banco de dados de produção, espera-se que seja obtido uma data mais próxima à realidade.

Referências

Costa, R. (2006). **SQL: Guia prático**. Rio de Janeiro, Brasport, 2ª edição.

Farias, P. (2004). **Curso Essencial de Linux**. Digerati.

Kofler, M. (2001). **MySQL**. Apress.

Lemay, L. Cadenhead, R. (2005). **Aprenda em 21 dias Java 2**. Elsevier Brasil.

Taylor, A. (2003). **SQL For Dummies**. John Wiley & Sons.

Veras, M. (2011). **Virtualização**. Rio de Janeiro, Brasport.

Zeis, C.; Ruel, C.; Wessler, M. (2009). **Oracle 11g For Dummies**. John Wiley & Sons.