

República Federativa do Brasil

Luiz Inácio Lula da Silva

Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues

Ministro

Empresa Brasileira de Pesquisa Agropecuária - Embrapa

Conselho de Administração

José Amauri Dimárzio

Presidente

Clayton Campanhola

Vice-Presidente

Alexandre Kalil Pires

Dietrich Gerhard Quast

Sérgio Fausto

Urbano Campos Ribeiral

Membros

Diretoria Executiva da Embrapa

Clayton Campanhola

Diretor-Presidente

Gustavo Kauark Chianca

Herbert Cavalcante de Lima

Mariza Marilena T. Luz Barbosa

Diretores-Executivos

Embrapa Informática Agropecuária

José Gilberto Jardine

Chefe-Geral

Tércia Zavaglia Torres

Chefe-Adjunto de Administração

Sônia Ternes Frassetto

Chefe-Adjunto de Pesquisa e Desenvolvimento

Álvaro Seixas Neto

Supervisor da Área de Comunicação e Negócios



Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento

ISSN 1677-9274
Novembro, 2003

Documentos 33

Organização e Armazenamento de Arquivos de *Back up* de Banco de Dados

Maria Angelica de Andrade Leite

Campinas, SP
2003

Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Av. André Tosello, 209
Cidade Universitária "Zeferino Vaz" Barão Geraldo
Caixa Postal 6041
13083-970 Campinas, SP
Telefone (19) 3789-5743 Fax (19) 3289-9594
URL: <http://www.cnptia.embrapa.br>
e-mail: sac@cnptia.embrapa.br

Comitê de Publicações

Carla Geovana Nascimento Macário
Ivanilde Dispatto
Luciana Alvim Santos Romani (Presidente)
Marcia Izabel Fugisawa Souza
Marcos Lordello Chaim
Suzilei Almeida Carneiro

Suplentes

Carlos Alberto Alves Meira
Eduardo Delgado Assad
José Ruy Porto de Carvalho
Maria Angélica de Andrade Leite
Maria Fernanda Moura
Maria Goretti Gurgel Praxedis

Supervisor editorial: *Ivanilde Dispatto*
Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*
Editoração eletrônica: *Área de Comunicação e Negócios (ACN)*

1ª. edição on-line - 2003

Todos os direitos reservados.

Leite, Maria Angelica de Andrade.

Organização e armazenamento de arquivos de *back up* de banco de dados / Maria Angelica de Andrade Leite. — Campinas : Embrapa Informática Agropecuária, 2003. 17 p. : il. — (Documentos / Embrapa Informática Agropecuária ; 33).

ISSN 1677-9274

1. *Back up*. 2. Ferramenta Agência. 3. Agência de Informação Embrapa. I. Título. II. Série.

CDD - 005.86 21st ed.

Autor

Maria Angelica de Andrade Leite
M.Sc. em Ciência da Computação,
Pesquisadora da Embrapa Informática
Agropecuária, Caixa Postal 6041, Barão
Geraldo 13083-970 - Campinas, SP.
Telefone (19) 3789-5796
e-mail: angelica@cnptia.embrapa.br

Apresentação

A Empresa Brasileira de Pesquisa Agropecuária - Embrapa, criou o serviço da Agência de Informação da Embrapa visando disponibilizar informações referentes aos produtos pesquisados por ela através da Internet. A Agência de Informação da Embrapa vai disponibilizar a informação através de uma árvore de conhecimento, descrevendo a cadeia produtiva relativa a cada produto, e através de mecanismos de busca de documentos referenciados pela árvore.

Para a implementação e implantação da Agência de Informação da Embrapa foi criado o projeto Agência envolvendo as unidades da Embrapa responsáveis por seu desenvolvimento. Estas unidades se constituem do Serviço de Comunicação para Transferência de Tecnologia, a Embrapa Informática Agropecuária e a Embrapa Gado de Corte.

A base de dados, onde são armazenados os dados inseridos na Agência, é centralizada na Embrapa Informática Agropecuária e, por este motivo, a unidade tem a responsabilidade de manter um esquema de *back up* dos dados inseridos. Existem soluções de software de *back up* de domínio público mas, dentro do projeto Agência, optou-se pelo desenvolvimento de *scripts* simples e de fácil manutenção e evolução caso seja necessária a incorporação de novas funcionalidades.

Este documento vai descrever os requisitos para a realização de procedimento de *back up* no projeto Agência, o esquema de atribuição de nomes, organização e armazenamento de arquivos de *back up* em função destes requisitos bem como os algoritmos criados para implementar este procedimento.

José Gilberto Jardine
Chefe-Geral

Sumário

Introdução.....	9
Esquema de <i>Back up</i> Proposto.....	10
Apresentação dos Algoritmos de Procedimento de <i>Back up</i>	12
Algoritmo de <i>Back up</i> Diário, Semanal e Mensal no Micro.....	13
Algoritmo de <i>Back up</i> na Estação de Trabalho.....	14
Tecnologias Utilizadas na Implementação.....	15
Referências Bibliográficas.....	17

Organização e Armazenamento de Arquivos de *Back up* de Banco de Dados

Maria Angelica de Andrade Leite

Introdução

Este documento foi elaborado com o intuito de difundir o esquema de organização e armazenamento de arquivos de *back up* de banco de dados utilizado no projeto de desenvolvimento da Agência de Informação da Embrapa (Embrapa Comunicação para Transferência de Tecnologia, 2001; Sarmiento, 2000). A motivação para elaboração deste documento é que o esquema de organização de armazenamento de arquivos de *back up*, aqui apresentado, possa ser utilizado em outros projetos.

O problema de realização de *back up* é uma constante preocupação das empresas. Um dos objetivos de se realizar o *back up* consiste em minimizar a possibilidade de perda de dados devido a falhas de hardware, software ou mesmo causas adversas como incêndios ou acidentes na empresa e, também, reduzir, ao mínimo, a indisponibilidade de um sistema devido a algum problema com os seus dados. O processo de *back up* consiste em manter uma cópia dos dados em um outro dispositivo que permita a sua recuperação em caso de necessidade.

O projeto Agência de Informação da Embrapa centraliza o armazenamento dos dados de todas as unidades da Embrapa em tabelas do banco de dados Oracle8i Enterprise Edition Release 8.1.7.0.0 (Oracle Corporation, 2000). Desta forma o projeto Agência possui a responsabilidade de prover uma forma de *back up* para estes dados.

Existem vários softwares de domínio público disponíveis para realização de *back up* sendo que um deles é o Amanda (University of Maryland at College Park, 2003). O software Amanda é um sistema de *back up* bastante completo, com várias funcionalidades e que permite ao administrador de rede montar um servidor de *back up* para realizar o *back up* de várias máquinas em uma fita de grande capacidade. Inicialmente foi desenvolvido para trabalhar em estações de trabalho com o sistema operacional UNIX mas as versões mais recentes podem usar o serviço Samba (Samba..., 2003) para realizar *back up* em máquinas Windows.

Como o esquema de *back up* do projeto Agência é simples, optou-se pelo desenvolvimento de *scripts* fáceis de manter e de adaptar caso surja a necessidade de novas funcionalidades. A política de *back up* foi estabelecida logo que as unidades da Embrapa começaram a fazer a inserção dos dados. O *back up* armazena dados históricos, isto é, a intenção é manter cópia dos dados inseridos desde o início da implantação da Agência para permitir que, caso os usuários alterem os dados atuais, e depois, decidam recuperar dados previamente inseridos, o sistema tenha estes dados disponíveis de acordo com uma data específica ou a época aproximada da inserção. Assim, o projeto Agência se propõe a manter uma cópia dos dados tanto para os casos em que haja necessidade de recuperação dos mesmos, por motivos de perda devido a falhas de hardware e/ou software, quanto para permitir que um usuário possa recuperar dados inseridos numa fase anterior, dos quais ele mesmo não possui uma cópia, mas que necessite ter acesso. Para este fim foi elaborado um esquema de organização, atribuição de nomes e armazenamento de arquivos agrupando os mesmos por dia, semana e mês facilitando assim a sua recuperação em função da época em que os dados foram inseridos.

Este documento apresenta a política de *back up* desenvolvida para os dados inseridos na base de dados centralizada do projeto Agência. É apresentada a forma de organização e atribuição de nomes aos arquivos de *back up* bem como os algoritmos desenvolvidos para implementar esta política. Ao final do documento são apresentadas as tecnologias utilizadas na implementação dos algoritmos.

Esquema de *Back up* Proposto

Conforme discutido na introdução, dentro do projeto da Agência, os requisitos para a organização do *back up* foram bastante específicos. De acordo com estes requisitos estabeleceu-se os itens, listados a seguir, para serem obedecidos. Para que os arquivos de *back up* possam ser recuperados em função de uma data ou época de inserção foi estabelecido um esquema de atribuição de nomes dos mesmos de modo que possa ser possível recuperar os arquivos por dia ou por uma semana relativa no mês ou por um determinado mês do ano. Os requisitos são apresentados a seguir:

1. armazenar os arquivos de *back up* indicando o dia da realização do mesmo;
2. gerar um arquivo de *back up* semanal contendo os arquivos diários da respectiva semana;
3. gerar um arquivo de *back up* mensal contendo os arquivos diários do mês;
4. os arquivos de *back up* devem ser armazenados em fita de forma que um mesmo arquivo de *back up* possa estar em mais de uma fita evitando perda de dados, no caso de ocorrer problema de recuperação com alguma fita;

5. o banco de dados Oracle está instalado em um micro com sistema Windows NT e o dispositivo de leitura/gravação de fita encontra-se conectado numa estação de trabalho Sun com sistema operacional Unix. Desta forma o *back up* do banco de dados é realizado no micro e os arquivos devem ser copiados para a estação de trabalho para que se possa realizar o *back up* em fita.

A atribuição de nomes aos arquivos de *back up* e o esquema de *back up* proposto foi estabelecido para gerar arquivos de acordo com os itens 1, 2 e 3. Assim, foi estabelecida a realização de um *back up* diário e a necessidade de uma fita para armazenar arquivos de *back up* para cada dia da semana. No final de uma semana os arquivos diários da semana compõem o *back up* semanal correspondente e deve ser alocada uma fita para cada semana do mês. No final de cada mês é elaborado o *back up* mensal considerando os arquivos semanais relativos ao mês em questão. Os arquivos de *back up* mensal são armazenados em uma fita própria para cada mês. Desta forma, no final de 1 mês, cada arquivo diário estará armazenado em pelo menos 2 fitas (semanal e mensal). Os arquivos de *back up* semanal e mensal são armazenados na fita de forma incremental. Os arquivos diários armazenados referem-se apenas à semana de trabalho corrente. Em cada um dos casos os arquivos de *back up* são compactados e possuem uma referência do dia, semana ou mês associada ao seu nome facilitando, assim, a recuperação dos dados em função de uma data ou época.

Atualmente são utilizadas 7 fitas para armazenar o *back up* diário (uma por dia), 5 fitas para armazenar os *back ups* semanais (uma para cada semana, considerando que um mês de 30 dias possui 4 semanas de 7 dias mais 1 semana de 2 dias) e uma fita para cada mês. No total são utilizadas 24 fitas anualmente:

7 fitas diárias + 5 fitas semanais + 12 fitas mensais = 24 fitas

O número acima não considera o tempo de desgaste de uso das fitas que é de 6 meses. Desta forma, quando uma fita completar 6 meses de uso ela deve ser reposta por uma nova fita.

Os arquivos de *back up* são nomeados da seguinte forma:

- arquivos de *back up* diário: agdia < dia da semana > .zip, sendo que o dia da semana é dado pelas 3 primeiras iniciais (Dom, Seg, Ter, Qua, Qui, Sex, Sab). Exemplo: agdiaSeg.zip
- arquivos de *back up* semanais: agsem < número da semana no mês > < mês e ano > .zip, sendo que o número da semana varia de 1 a 5 e o mês e ano é dado no formato mmaaaa. Exemplo: agsem3072003.zip
- arquivos de *back up* mensais: agmen < mês e ano > .zip sendo que o mês e ano são no formato mmaaaa. Exemplo: agmen062003.zip

Apresentação dos Algoritmos de Procedimento de *Back up*

Dadas as condições de infra-estrutura computacional do projeto Agência, estabelecidas no requisito 5, o desenvolvimento do procedimento de *back up* resultou na elaboração de dois algoritmos. O primeiro algoritmo é executado diariamente no micro onde encontra-se instalado o banco de dados Oracle e o segundo algoritmo é executado na estação de trabalho conectada a um dispositivo de leitura/gravação de fita.

No micro onde encontra-se instalado o banco de dados Oracle, inicialmente é realizado o *back up* diário do banco, através do comando de exportação de dados do mesmo: *export* (Loney & Kock, 2000). Na execução deste comando o banco de dados gera dois arquivos: o arquivo *agdump.dmp*, que contém os dados e informações de estruturas das tabelas, e o arquivo *aglog.log*, que contém a descrição do andamento do *back up*. Este último arquivo descreve se a realização do *back up* foi bem sucedida ou, caso contrário, indica os problemas ocorridos. Uma vez que os arquivos *agdump.dmp* e *aglog.log* são gerados, eles são renomeados e a data do dia é acrescentada em seu nome. Tem-se então os mesmos arquivos com os nomes *agdump<ddmmaaaa>.dmp* e *aglog<ddmmaaaa>.log*. Em seguida estes dois arquivos são compactados, em um único arquivo, cujo nome identifica o dia da semana como por exemplo, *agTer.zip* no caso de ser um *back up* realizado na terça-feira. A seguir o arquivo gerado é copiado para o diretório de *back up* diário do computador e é realizada uma cópia para o diretório de *back up* diário da estação de trabalho. Em um determinado dia da semana (no caso estabeleceu-se a segunda-feira) é feito o *back up* semanal onde os arquivos que se encontram no diretório de *back up* diário são compactados para formar o arquivo de *back up* semanal. Em seguida os arquivos de *back up* diário são eliminados e o arquivo de *back up* semanal resultante é armazenado em um diretório semanal no computador e é feita uma cópia no diretório de *back up* semanal na estação de trabalho. No dia 1 de cada mês os arquivos semanais (que são referentes ao mês anterior) são compactados, e o arquivo resultante recebe o nome do mês correspondente e é armazenado no diretório de *back up* mensal e é feita uma cópia no diretório mensal na estação de trabalho. Este algoritmo é executado automaticamente todos os dias.

Na estação de trabalho é executado o algoritmo que armazena os arquivos de *back up* na fita. Conforme apresentado no tópico esquema de *back up* proposto, existe uma fita para cada dia, semana e mês. O algoritmo da estação de trabalho é executado pelo responsável pelo *back up* pois ele deve alimentar as fitas na gravadora à medida em que são requisitadas. O algoritmo percorre os diretórios de *back up* diário, semanal e mensal da estação e verifica se existem arquivos nos respectivos diretórios. Como cada arquivo possui a identificação do dia, semana ou mês associado ao seu nome então o algoritmo identifica qual fita deve ser

inserida no dispositivo de leitura/gravação de fita. Uma vez identificada a fita, o algoritmo pede ao usuário para colocar a fita correspondente no dispositivo. depois que o algoritmo faz a cópia do arquivo em fita ele apaga o mesmo do diretório da estação de trabalho. os algoritmos são apresentados a seguir:

Algoritmo de *back up* diário, semanal e mensal no micro

Início do algoritmo

Pegar, no sistema, data do dia no formato (ddmmaaaa)

Pegar, no sistema, o mês atual no formato (mm)

Pegar, no sistema, o mês anterior no formato (mm)

Pegar, no sistema, o ano atual (aaaa)

Pegar, no sistema, o nome do dia da semana (seg, ter, qua, qui, sex, sab, dom)

Pegar, no sistema, o mês e o ano no formato (mmaaaa)

Pegar, no sistema, o número da semana no mês no formato (f)

Se for segunda-feira ou dia 01

Então {/*** Faz *backup* semanal ***/

 Ir para diretório de *back up* diário

 Se tiver arquivos de *back up* diário

 Então{

 Compactar arquivos de *back up* diário e gerar o arquivo de *back up* semanal com o valor da semana no mês (agsem< fmmaaaa> .zip)

 Copiar arquivo gerado para diretório de *back up* semanal no micro

 Copiar arquivo gerado para diretório de *back up* semanal na estação

 Apagar arquivos de *back up* diários do diretório corrente

 Fazer relatório do andamento do *back up* semanal

 }

Se for dia 01

Então {/*** Faz *backup* mensal ***/

 Ir para diretório de *back up* semanal

 Se tiver arquivos de *back up* semanal

 Então{

 Compactar arquivos de *back up* semanal com o nome do mês anterior

 (agmens< mmaaaaa> .zip)

 Copiar arquivo gerado para diretório de *back up* mensal no micro

 Copiar arquivo gerado para diretório de *back up* mensal na estação

 Apagar arquivos semanais do diretório corrente

 Fazer relatório do andamento do *back up* mensal

 }

}

```
/*** Faz backup diário ***/
```

Executar comando de geração de arquivo de *back up* do Oracle (resultando os arquivos agdmp.dmp e aglog.log)

Colocar a data do dia no nome dos arquivos (agdmp< ddmmaaaa> .dmp e aglog< ddmmaaaa> .log)

Compactar os arquivos de *back up* e colocar o dia da semana (ex: agdiaSeg.zip)

Apagar os arquivos de *back up* originais *.dmp e *.log

Armazenar o arquivo no diretório de *back up* diário

Enviar o arquivo de *back up* diário para a estação de trabalho

Fazer relatório do andamento do *back up* diário

Enviar uma mensagem com o resultado dos *back ups* para o responsável

Fim do algoritmo

Algoritmo de *Back up* na Estação de Trabalho

Início do algoritmo

```
/*** Backup mensal ***/
```

Ir para o diretório de *back up* mensal

Enquanto houver arquivos no diretório mensal {

 Verificar o mês do nome do próximo arquivo agmen(mmaaaa)

 Emitir mensagem de colocar fita de *back up* mensal correspondente ao mês do arquivo

 Gravar fita de *back up* mensal

 Apagar arquivo de *back up* mensal na estação de trabalho

 Fazer relatório do andamento do *back up* mensal

}

```
/*** Backup semanal ***/
```

Ir para diretório de *back up* semanal

Enquanto houver arquivos no diretório semanal {

 Verificar o número da semana e o nome do mês no nome do próximo arquivo

 Emitir mensagem de colocar fita semanal de mesmo número e mês do arquivo

 Gravar fita de *back up* semanal (tar c)

 Apagar arquivo de *back up* semanal na estação de trabalho

 Fazer relatório do andamento do *back up* semanal

}

```
/* ** Backup diário ** */  
Ir para diretório de back up diário  
Enquanto houver arquivos no diretório diário {  
    Verificar o nome do dia da semana no nome do próximo arquivo  
    Emitir mensagem de colocar fita diária de mesmo nome do arquivo  
    Gravar fita de back up diário  
    Apagar arquivo de back up diário na estação de trabalho  
    Fazer relatório do andamento do back up diário  
}  
Enviar mensagem com os relatórios dos back ups para o responsável  
Fim do algoritmo
```

Na descrição dos algoritmos deve-se notar que os arquivos de *back up* gerados no micro são armazenados em diretórios de *back up* diário, semanal e mensal no mesmo, e é feita uma cópia destes arquivos para os diretórios correspondentes na estação de trabalho. Quando o algoritmo de gravação em fita é executado os arquivos na estação de trabalho são apagados. Os diretórios armazenados no micronão são apagados. Assim existe um conjunto de arquivos de *back up* que fica armazenado em máquina bem como em fita. Desta maneira o projeto Agência possui duas formas de armazenamento de arquivos de *back up* o que aumenta a sua segurança. Esta decisão foi tomada em função de haver espaço suficiente no micro para armazenar os *back ups* realizados no banco de dados e por tratar-se de uma opção onde o acesso aos arquivos de *back up* é mais rápido no caso da necessidade de restaurá-los. No caso de haver falha do micro, então existem os *back ups* em fita que podem ser restaurados em outro micro. Caso seja necessário fazer um cópia dos arquivos de *back up* para um terceiro computador basta adicionar um novo comando de copiar arquivos de *back up* diário, semanal e mensal para o micro desejado no algoritmo.

Tecnologias Utilizadas na Implementação

Este trabalho apresentou o esquema de *back up* da Agência de Informação da Embrapa, que utiliza o banco de dados Oracle para armazenamento dos dados. Atualmente estes algoritmos encontram-se implementados e em funcionamento dentro do projeto. Um arquivo de *back up* diário possui um tamanho aproximado de 1,5 Mb, um arquivo de *back up* semanal possui aproximadamente 6 Mb e um arquivo de *back up* mensal possui aproximadamente 25 Mb. Na sua implementação utilizou-se o Ant (Apache Software Foundation, 2003), uma ferramenta de implementação de aplicativos de build baseada na linguagem Java (Sun Microsystems, 2003). A ferramenta Ant oferece uma série de funções para tratar arquivos. As tarefas a serem executadas são escritas em um arquivo na linguagem XML (Harold, 1999) e o aplicativo Ant interpreta este arquivo. Para as

operações de compactação de arquivos utilizou-se o WinZip (Winzip Computing, 2003). Dentro da estratégia proposta para o esquema de organização e armazenamento de *back ups* do projeto Agência, os algoritmos desenvolvidos estão satisfazendo às necessidades do projeto. O processo de *back up* está em constante acompanhamento por parte do responsável de *back ups* do projeto Agência para que sejam realizadas melhorias no procedimento, caso sejam necessárias.

Referências Bibliográficas

APACHE SOFTWARE FOUNDATION. The Apache Ant Project. Disponível em: < <http://ant.apache.org/> > . Acesso em: 21 jul. 2003.

EMBRAPA COMUNICAÇÃO PARA TRANSFERÊNCIA DE INFORMAÇÃO. Documento orientador para a organização da informação. Brasília, DF: Embrapa Comunicação para Transferência de Tecnologia: Embrapa Gado de Corte: Embrapa Informática Agropecuária, 2001. Disponível em < <Http://www.sct.embrapa.br/GOI/manuais/OrganizacaoInformacao.doc> > . Acesso em: 21 jul. 2003.

HAROLD, E. R. XML bible. Foster City: IDG Books Worldwide, 1999. 1015 p.

LONEY, K.; KOCK, G. Oracle 8i: the complete reference. Berkeley: Osborne McGraw Hill, 2000. 1308 p.

ORACLE CORPORATION. Internet developer suite CD Pack v 1.0.2.4.0 for Microsoft Windows. Redwood Shores, CA, 2000.

SAMBA opening windows to a wider world. Disponível em: < <http://www.samba.org/> > . Acesso em: 22 out. 2003.

SARMENTO, E. P. M. Organização da informação para comunicação e transferência tecnológica. Brasília, DF: Embrapa Comunicação para Transferência de Tecnologia, 2000. (Embrapa. Programa 18 Transferência de Tecnologia: Comunicação e Negócios. Projeto 18.2000.283). Projeto em andamento.

SUN MICROSYSTEMS. Java technology. Disponível em: < <http://java.sun.com/> > . Acesso em: 21 jul. 2003.

UNIVERSITY OF MARYLAND AT COLLEGE PARK. AMANDA - the Advanced Maryland Automatic Network Disk Archiver. Disponível em: < <http://www.amanda.org/> > . Acesso em: 22 out. 2003.

WINZIP COMPUTING. Winzip [home page]. Disponível em: < <http://www.winzip.com/> > . Acesso em: 21 jul. 2003.



Informática Agropecuária

Ministério da Agricultura, Pecuária e Abastecimento **Governo Federal**