



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS



**APERFEIÇOAMENTO E DESENVOLVIMENTO DE  
FERRAMENTAS DO CONTROLE DA AUTOMATIZAÇÃO DO  
BACKUP E RECUPERAÇÃO DOS DADOS DE SATÉLITES EM  
FITAS LTO E DLT**

**RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA  
(PIBIC/CNPq/INPE)**

**Carlos Alexandre de Oliveira Rodrigues (UNISAL, Bolsista  
PIBIC/CNPq)**

**Email: [carlos.rodrigues@cptec.inpe.br](mailto:carlos.rodrigues@cptec.inpe.br)**

**Wagner Flauber Araújo Lima (DSA/CPTEC/INPE, Orientador)**

**Email: [wagner.flauber@cptec.inpe.br](mailto:wagner.flauber@cptec.inpe.br)**

**Julho de 2011**

## **SUMÁRIO**

### **CAPÍTULO 1 – INTRODUÇÃO**

1.1 Objetivos gerais do presente projeto de iniciação científica

1.2 Metas cumpridas

### **CAPÍTULO 2 – ORIGEM E DESTINO DOS DADOS**

### **CAPÍTULO 3 – SATÉLITES**

3.1 Satélite GOES

3.2 Satélite MSG

3.3 Satélite NOAA

3.4 Satélite AQUA

3.5 Satélite TERRA

### **CAPÍTULO 4 – VOLUME TOTAL DE DADOS**

4.1 Tabela de volume

4.2 Gráfico de produtos

### **CAPÍTULO 5 – BANCO DE DADOS DE IMAGENS**

### **CAPÍTULO 6 – DESENVOLVIMENTO**

6.1 Backup

6.2 Recuperação de dados

### **CAPÍTULO 7 – PRINCIPAIS ATIVIDADES DESENVOLVIDAS**

### **CAPÍTULO 8 – CONSIDERAÇÕES FINAIS**

### **CAPÍTULO 9 – AGRADECIMENTOS**

# 1. INTRODUÇÃO

Com a grande e cada vez maior quantidade de dados e imagens de Satélites com grande número de aplicações ambientais nos campos da Meteorologia e Clima, Agricultura, Ciências Florestais, Hidrologia, Oceanografia, Defesa Civil etc., é importante disponibilizá-los continuamente e claro armazená-los. Entretanto, nem todas as instituições que trabalham com esses tipos de dados os disponibilizam continuamente.

Dados e imagens de Satélites são processados por sistemas especializados para tal e em seguida registrados em meio magnético para armazenamento e uso atual.

Com o surgimento de novos satélites, há a necessidade de automatização do processo de armazenamento dos dados, sabendo que não vem a ser possível tornar o processo 100% automatizado pela necessidade de colocar as fitas nas unidades de gravação e pela geração de etiquetas para as fitas. Pensando por essa parte, o meio mais cabível é automatizar tudo aqui que vem a ser possível de fazê-lo.

## 1.1. Objetivos gerais do presente projeto de iniciação científica

O projeto foi iniciado em Agosto de 2010 e o objetivo principal traçado deste projeto foi desenvolver uma ferramenta para agilizar e facilitar o procedimento de backup e recuperação de dados históricos, renomeando-os arquivos e passando para o novo formato em fitas LTO.

O trabalho foi desenvolvido utilizando apenas ferramentas livre que permitem aplicar a um ambiente Linux/Unix (Windows), mesmo com um grande volume de dados, permitindo uma grande redução nos custos. Diversas instituições, públicas e privadas, vêm adotando o uso de software livre, aumentando o investimento em tecnologia com baixos custos.

As principais ferramentas que foram utilizadas no desenvolvimento do sistema são: Java e Shell Scripts. Dentre os objetivos específicos do trabalho, pode-se destacar:

- 1- Aprofundar os conhecimentos no ambiente Linus/Unix;

2- Desenvolver Shell Scripts para armazenar e recuperar os dados armazenados em mídias DLT/LTO;

3- Aprofundar os conhecimentos em linguagem de programação Java Web e Shell Scripts;

4- Desenvolver interface para automatizar o processo de backup armazenamento de dados no banco de Dados de Imagens (BDI).

## **1.2. Metas cumpridas**

Foi desenvolvida uma ferramenta para agilizar esse processo de backup, mais especificamente no processo de disparar a gravação, no qual tornou possível definir as opções de gravação diretamente na tela do terminal, não precisando acessar o código do script para fazer a alteração, sendo essa uma forma mais segura.

Posteriormente foi possível automatizar a geração de listagens que mostravam o caminho dos dados a serem gravados, foi desenvolvido também um script que tinha um horário definido para ser disparado automaticamente (especificamente 00h01min), sendo que no momento de fazer o backup não ser mais necessário realizar esse procedimento, isso que trouxe grande ganho de tempo, visto que era preciso passar por varias etapas até ter a listagem estar pronta para gravação.

Outro ponto que foi passado, é a recuperação de dados das mídias antigas, esses que são dados importantes que precisavam ser disponibilizados.

Partindo desse princípio, os dados foram baixados das mídias antigas e colocados nos devidos diretórios, foram desenvolvidos scripts, o primeiro para pegar o caminho do arquivo no diretório e enviá-lo para outro de forma organizada e com nomenclatura e data do modelo atual, outro para gerar um log. de cada arquivo para serem enviados ao BDI (Banco de Dados de Imagens) para serem disponibilizados.

## 2. ORIGEM E DESTINO DOS DADOS

Resumo das etapas:

- 1- Recepção do satélite
- 2- Recepção do arquivo binário
- 3- Geração de log para a alimentação do banco.
- 4- Armazenamento em storage binário.
- 5- Armazenamento do produto em Storage web
- 6- Alimentação dos Metadados (Log. de Produtos)
- 7- Aplicação web busca imagens no banco de dados
- 8- BDI retorna link físico da imagem/binário
- 9- Aplicação acessa o local correto da imagem/binário selecionando o Storage Web ou Storage Binário

A próxima etapa é onde começa o backup manual em si. A máquina receptora é acessada remotamente, então é localizado o diretório em que se encontram os dados a serem armazenados. É preciso verificar o tamanho do diretório para designar também o tamanho da listagem e determinar a quantidade de arquivos (em GIGABYTES), que serão armazenados. Após isso a listagem é gerada e apendada ou gravada na fita LTO 2 ou LTO 3. Então é dado o comando para iniciar a gravação, através de script ou linha de comando dentro do terminal Linux.

Depois se a gravação for bem sucedida, vem a parte de apagar os dados do disco com a finalidade de gerar mais espaço para receber novas imagens, que são constantes, essas imagens caem de minutos em minutos então dá para se imaginar o volume de imagens.

Fora a lista de gravação é gerada uma listagem de saída mostrando todos os dados que foram gravados com o respectivo caminho dentro do diretório. Através de um Shell script é buscado esse caminho através da listagem de saída e apagado o dado do disco. Todos os passos de backup Manual são realizados através de scripts, desde a verificação do espaço em disco até apagar os arquivos.

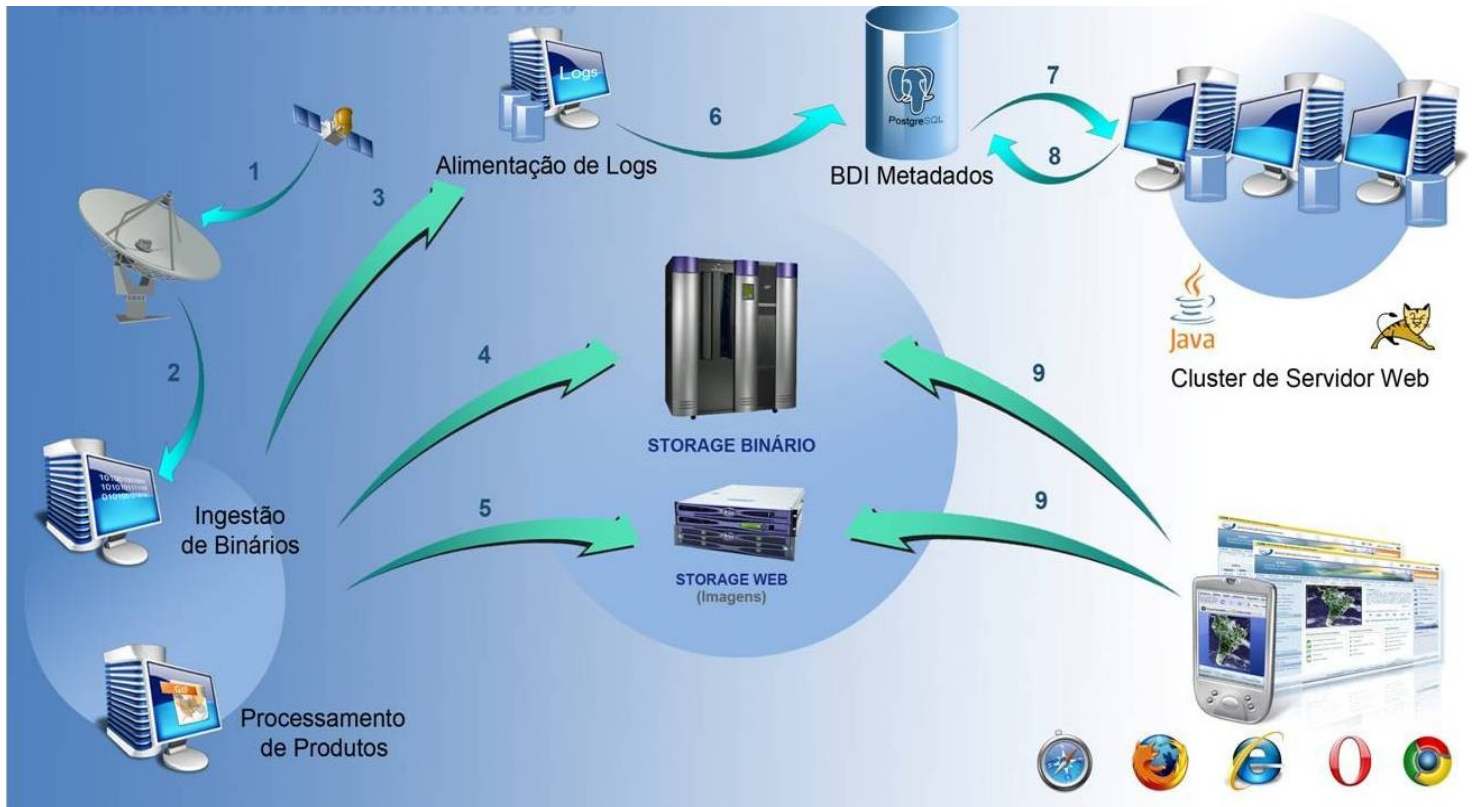


Figura1. Caminho das imagens de satélite, no site do DSA é disponibilizado o BDI com as imagens de diferentes satélites para serem consultadas, sendo essas sempre atualizadas

Esse backup, é muito importante e de grande responsabilidade, é preciso ter extrema atenção para realizar o processo, pois existem duas formas para disparar o backup, dizemos gravar (quando a fita ainda é virgem) e apendar (quando a fita já possui dados armazenados), pode ocorrer de o usuário gerar a listagem dos produtos e, por acaso o usuário venha a gravar numa fita que já tenha, por exemplo, cinco meses de dados, os mesmos serão simplesmente apagados, ou seja, esses dados estão perdidos e se alguém pedir os dados já não terá para disponibilizá-los.

Essa responsabilidade é claramente da pessoa que realiza o backup, para melhorar isso, foi desenvolvido o script para gravação que melhora bastante o problema do erro no comando e que o veremos mais a diante. Por outra parte ocorre o problema da vida útil que obviamente as fitas possuem, então se torna necessário uma mobilização e uma atenção maior dos responsáveis, de definir de período em período (talvez, de três em três anos), e providenciar mídias mais atuais para refazer as cópias dos dados para se ter mais segurança com fitas LTO 4 ou 5 (atualmente utilizamos 2 e 3).

### 3. SATÉLITES

Atualmente existem milhares de satélites presentes no espaço, claro incluindo funcionais e não-funcionais, no caso do INPE, a página da DAS/CPTEC disponibiliza uma análise e resumo dos principais satélites fonte de dados, que posteriormente são processados e acabam gerando as imagens:

#### 3.1. Satélite GOES

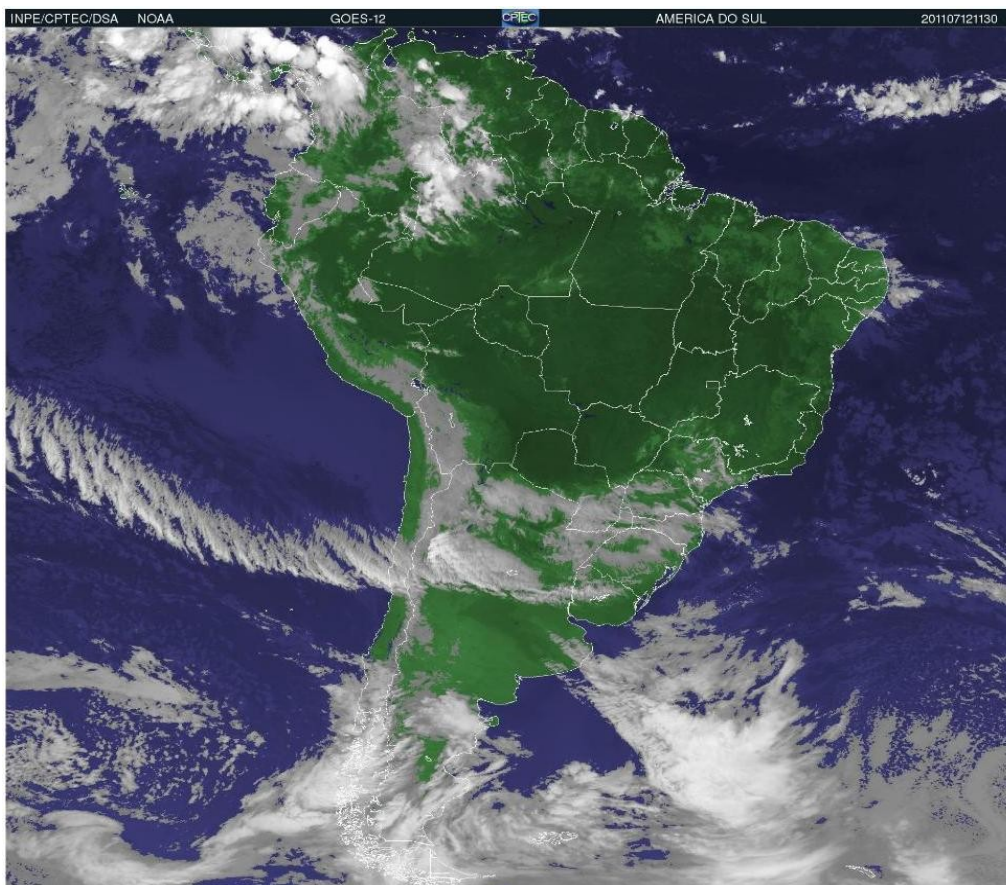


Figura2. Imagem do Satélite GOES

Em relação ao satélite GOES 8, sua órbita lhe permite ficar sempre sobre o mesmo ponto do equador. Em regime normal transmite uma imagem do globo terrestre a cada 3 horas, e de um fragmento do continente americano a cada meia hora, e seus sensores estão distribuídos em 5 canais, proveniente da Terra em 5 faixas de comprimento de onda. Cada elemento de informação, que vem a ser o pixel, referente ao canal visível corresponde a uma área de 1x1 km, se tratando de vapor d'água corresponde a 4x8 km, no ponto subsatélite.

### 3.2. Satélite MSG

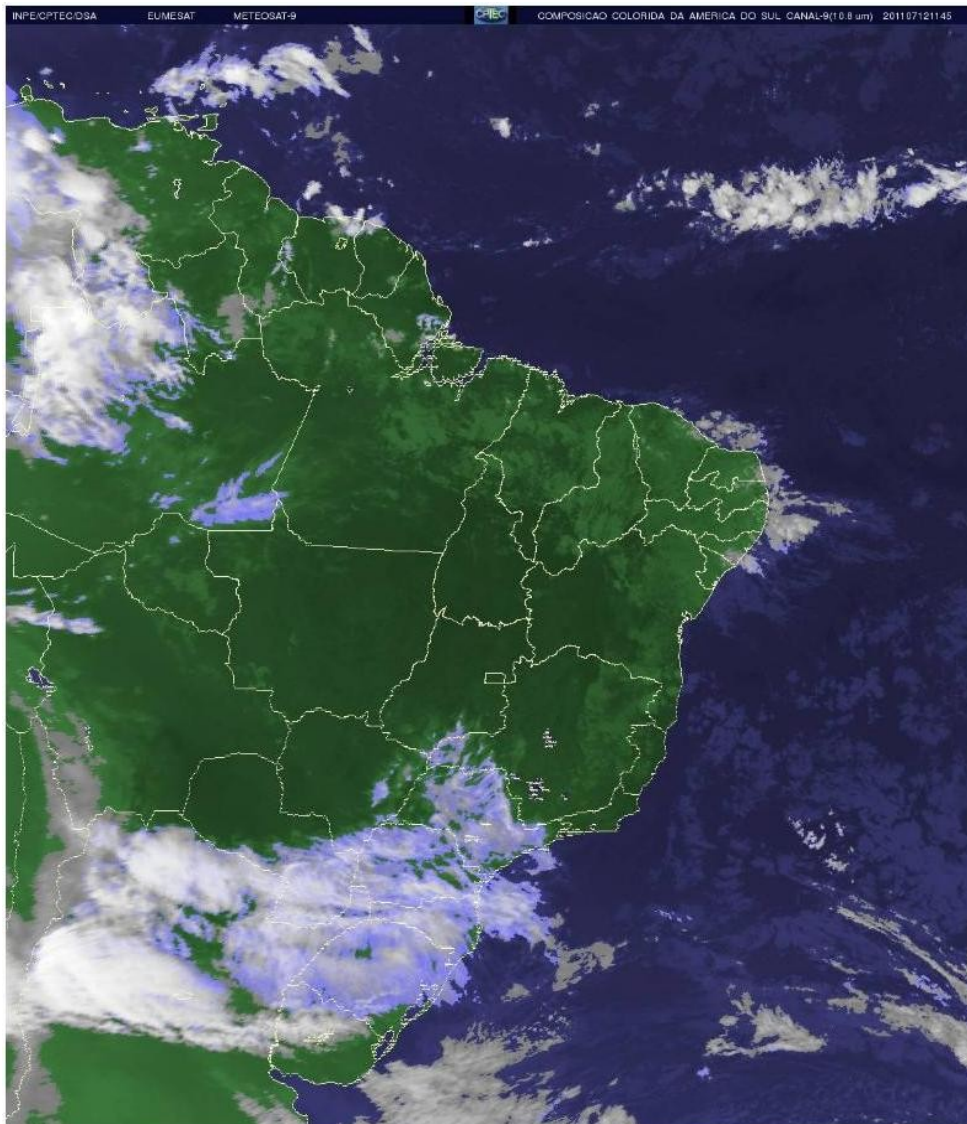


Figura3. Imagem do Satélite MSG

A plataforma espacial Meteosat de Segunda Geração (MSG) é localizada sobre o equador a 35.800km de altitude. Possui mesma velocidade angular da Terra e imageamento do disco planetário voltado ao satélite. O principal objetivo é a obtenção de imagens da atmosfera e da superfície dos continentes africanos e europeu e oceanos adjacentes. O Meteosat posicionado em 0° de longitude é importante para o Brasil pela cobertura do Oceano Atlântico e da região Nordeste.



### 3.3. Satélite NOAA

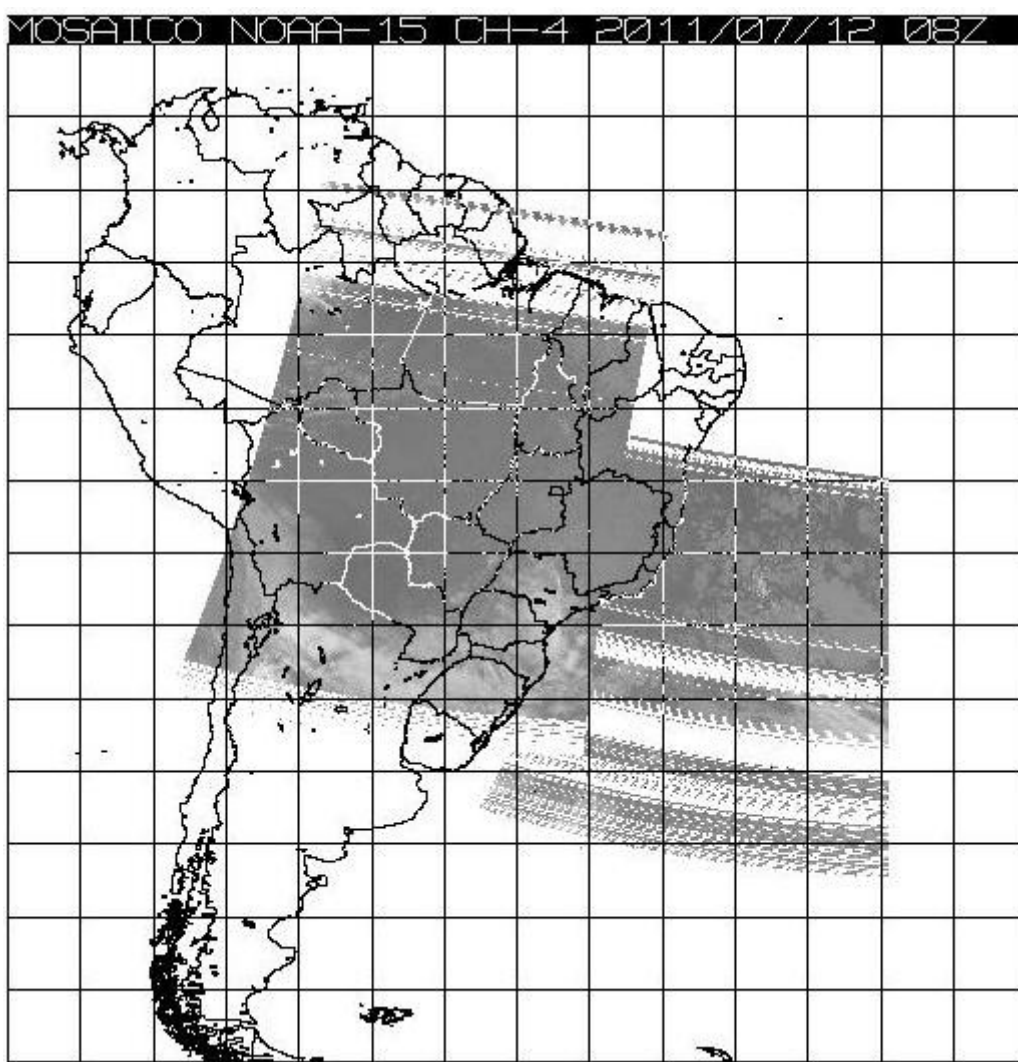


Figura4. Imagem do Satélite NOAA

Outro programa de satélites é gerido pela National Oceanic and Atmospheric Administration, a conhecida NOAA, através do NESDIS, e pela NASA, que é responsável pelo desenvolvimento e lançamento dos aparelhos. Este programa começou por se denominar TIROS (Television and Infrared Observation Satellite), e foi desenvolvido pela NASA e pelo Departamento de Defesa dos Estados Unidos, na tentativa de desenvolver um sistema de satélites meteorológicos. Entre 1960 e 1965 foram lançados 10 satélites TIROS. Entre 1966 e 1969, foram lançados 9 novos satélites, denominados TOS, ou seja, satélites operacionais TIROS, operados pela ESSA, pertencente à NOAA.

### 3.4. Satélite AQUA

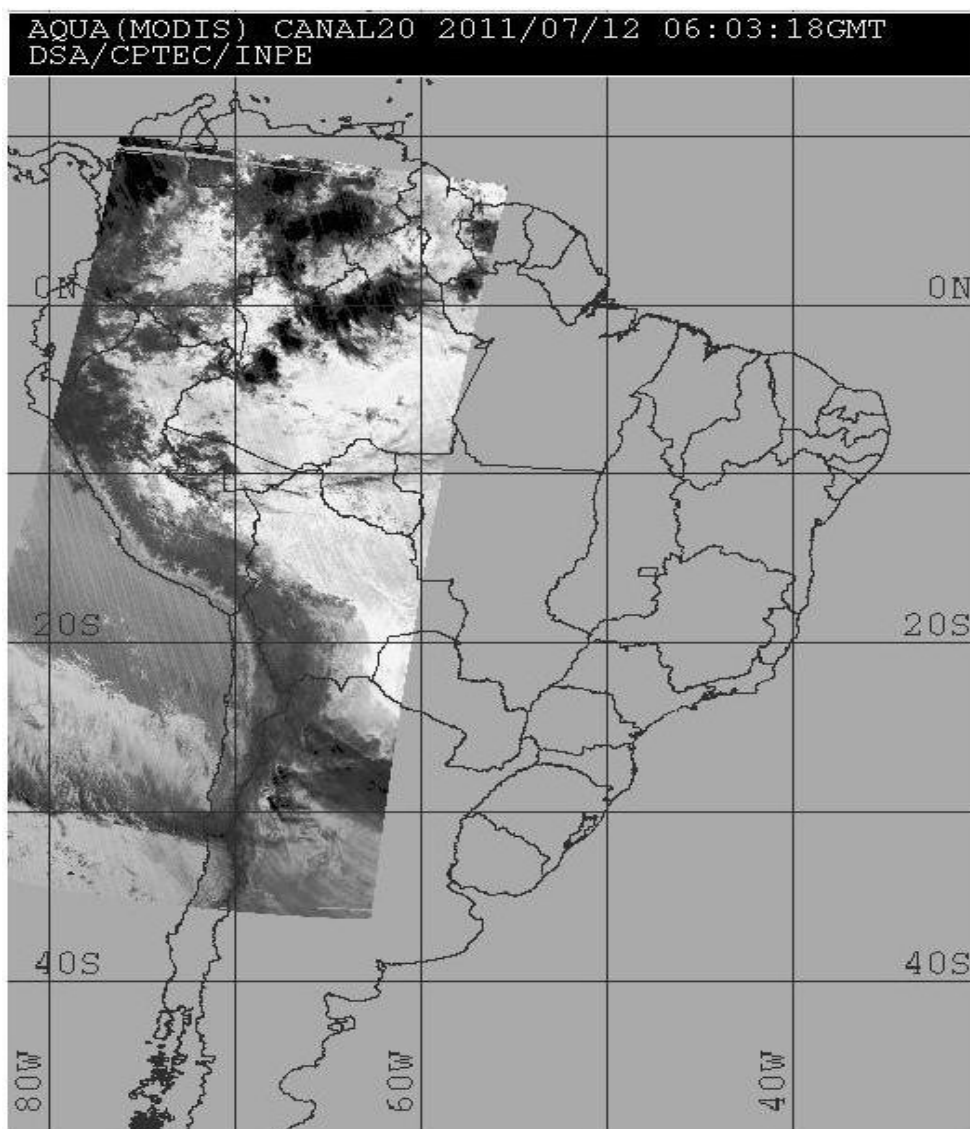


Figura5. Imagem do Satélite AQUA

Outro satélite é o Aqua, que é parte do programa Sistema de Observação da Terra (EOS/NASA) que estuda a Terra como parte do Sistema Global. O Aqua possui seis instrumentos de última geração para monitorar os oceanos, atmosfera, superfície, gelo, e cobertura de neve e vegetação, com alta precisão, detalhe espacial e frequência temporal. O Aqua é um projeto em conjunto com os Estados Unidos, Japão e Brasil e transporta o HSB, o sondador de umidade brasileiro. O principal sensor é o MODIS. O Aqua cruza o equador de sul para norte no período vespertino a uma altura de 700 km.

### 3.5. Satélite TERRA

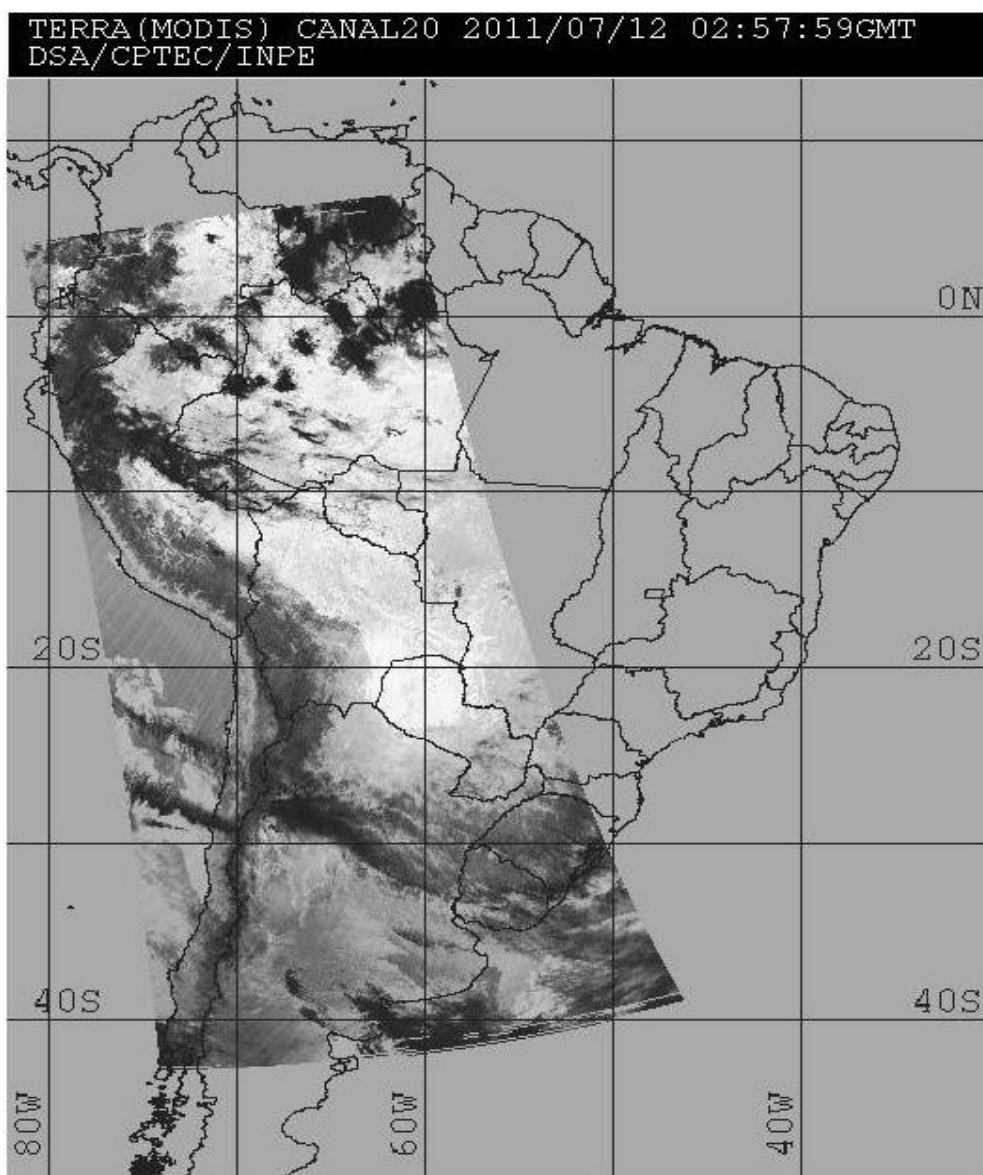


Figura6. Imagem do Satélite TERRA

Assim como o Aqua, o satélite Terra também faz parte do Sistema de Observação da Terra (EOS/NASA). Atua no período matutino cruzando o equador de norte para sul (movimento oposto ao Aqua). Esses satélites sondam a superfície total do planeta a cada 1-2 dias, utilizando 36 bandas espectrais.

Esses dados ajudam a compreender a dinâmica e os processos globais que ocorrem sobre a superfície, nos oceanos e na baixa atmosfera. O MODIS está desempenhando um papel vital na validação dos modelos globais do sistema Terra, estes capazes de prever as mudanças globais.

## 4. VOLUME TOTAL DE DADOS

As imagens desempenham um papel importante em aplicações geográficas. Sensores óticos, SAR e outros têm fornecido vários milhões de cenas que são sistematicamente coletadas, processadas e armazenadas. Anualmente são armazenadas dezenas de Tbytes de dados.

Esse grande volume de dados referentes aos satélites é recebido todos os dias, feito o trabalho e processamento, e podendo ser até mesmo imediatamente disponibilizados, caso o dado seja do tipo concorrente, ou seja, há a necessidade do dado estar disponível para zero(s) usuário(s) dentro das primeiras 24 horas após o início do trabalho. Inclusos nesses dados, estão aqueles que fazem parte das coletas realizadas rotineiramente, os quais são transmitidos de acordo com as normas existentes no código da Organização Meteorológica Mundial. Ainda possuem os dados que são classificados como recentes, esses abrangem um determinado, que pode variar bastante (até mesmo um mês), após esse período precisarão ser disponibilizados em no máximo dois dias.

O volume total de dados apurados referente ao mês de Julho de 2010, com os principais produtos, aqueles que apresentam maior volume e outros, foram de 5.047 GB, uma média de 166 GB por dia. Com isso dá para se ter a noção e perceber a importância de buscar novas formas e mecanismos com o objetivo de agilizar e dar mais segurança ao processo de backup. A seguir um gráfico e uma tabela apresentam esses números de forma mais detalhada:

### 4.1. Gráfico de produtos

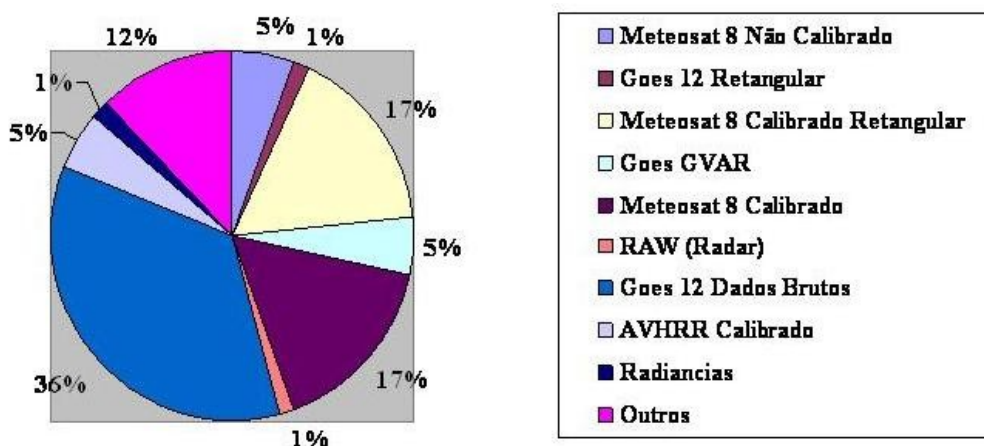


Figura8. Gráfico com a porcentagem dos produtos que possuem maiores volume de dados

## 4.2. Tabela de volume

Produtos	Subprodutos	Volume de Dados (GB)
Meteosat 8 não calibrado	36344	286
Goes 12 Retangular	11895	71
Meteosat 8 Cal. Retangular	32648	855
Goes GVAR	2571	265
Meteosat 8 Calibrado	32648	837
RAW (Radar)	45249	70
Goes 12 dados brutos	29528	1835
AVHRR Calibrado	3343	265
Radiancias	914	78

Figura7. Tabela que apresenta o número de subprodutos e o valor em GB dos produtos

## 5. BANCO DE DADOS DE IMAGENS

Se tratando de gerenciamento dos dados, Como permitir um acesso interessante a esse enorme volume de dados?

Infra		Vapor		Colorida		Realçada		Visível		Composta	
Baixa		Baixa		Baixa	Alta	Baixa	Alta	Baixa	Alta	Baixa	Alta
13:30		13:30		13:30	13:30	13:30	13:30	13:30	13:30	13:30	13:30
13:15		13:15		13:15	13:15	13:15	13:15	13:15	13:15	13:15	13:15
13:00		13:00		13:00	13:00	13:00	13:00	13:00	13:00	13:00	13:00
12:45		12:45		12:45	12:45	12:45	12:45	12:45	12:45	12:45	12:45
12:30		12:30		12:30	12:30	12:30	12:30	12:30	12:30	12:30	12:30
12:15		12:15		12:15	12:15	12:15	12:15	12:15	12:15	12:15	12:15
12:00		12:00		12:00	12:00	12:00	12:00	12:00	12:00	12:00	12:00
11:30		11:30		11:30	11:30	11:30	11:30	11:30	11:30	11:30	11:30
11:15		11:15		11:15	11:15	11:15	11:15	11:15	11:15	11:15	11:15
11:00		11:00		11:00	11:00	11:00	11:00	11:00	11:00	11:00	11:00
10:45		10:45		10:45	10:45	10:45	10:45	10:45	10:45	10:45	10:45
10:30		10:30		10:30	10:30	10:30	10:30	10:30	10:30	10:30	10:30
10:00		10:00		10:00	10:00	10:00	10:00	10:00	10:00	10:00	10:00
09:45		09:45		09:45	09:45	09:45	09:45	09:45	09:45	09:45	09:45
09:30		09:30		09:30	09:30	09:30	09:30	09:30	09:30	09:30	09:30
09:15		09:15		09:15	09:15	09:15	09:15	09:15	09:15	09:15	09:15

Figura9. Ilustração do Banco de Dados de Imagens disponibilizado na página do INPE, através do endereço: [http://satellite.cptec.inpe.br/acervo/goes\\_antiores.jsp](http://satellite.cptec.inpe.br/acervo/goes_antiores.jsp).

Bancos de Dados de Imagens são sistemas capazes de gerenciar grandes volumes de informação. Tipicamente: particionamento e multi-resolução.

Metadados são partes importantes do banco de dados visualização e/ou para distribuição.

Características necessárias para esses sistemas de bancos de dados de imagens:

- ✓ Navegação geográfica
- ✓ Navegação semântica
- ✓ Acesso através de serviços Web
- ✓ Consulta por conteúdo

As vantagens desse sistema gerenciador de banco de dados:

- ✓ Persistência
- ✓ Transações
- ✓ Controle de concorrência
- ✓ Recuperação em caso de falha
- ✓ Gerência de consultas
- ✓ Visões
- ✓ Segurança e proteção
- ✓ Eficiência
- ✓ Sistema distribuído
- ✓ Habilidade de tratar um grande número de usuários e um grande volume de dados

## 6. DESENVOLVIMENTO

### 6.1. Backup

A listagem apresenta o diretório e subdiretório que o dado a ser gravado em fita se encontra. O tamanho da listagem é designado pelo usuário, que na hora de gerar a listagem escolhe a data ou escolhe o volume em Gigabytes:



```
/samfs2/Lista_noaa_hrpt_standard_MES_11_bin - hsmdsa@bemtevi
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011010625.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011010802.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011012030.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011020738.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011022005.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011030714.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011030852.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011031940.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011032124.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011040830.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011041916.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011042100.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011050806.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011051852.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011051854.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011052035.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011060741.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011060743.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011062009.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011070718.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011070856.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011071946.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011072127.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011080833.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011090633.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011122108.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011130636.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011130814.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011131901.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011132042.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011132043.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011140749.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011140750.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011142017.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011142019.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011150727.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011151952.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011160703.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011160840.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011161928.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011162111.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011170640.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011170817.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011171905.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011172046.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011172047.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011180753.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011182019.gz
samfs2/DSA/noaal5/hrpt/standard/2010/11/$10322412_201011182022.gz
Line: 1/792      Column: 1      Character: 115 (0x73)
```

Figura10. Listagem já pronta para a gravação



O script para gravação desenvolvido visa o objetivo principal, mais segurança e agilidade. Como pode ser visto abaixo em execução, se o usuário escolher uma unidade de gravação que não exista, ao mesmo será informado que a opção é inválida e terá de tentar novamente:



```
bemtevi.cptec.inpe.br - PuTTY
-bash-3.00$ ./Script_Backup
*****
*
*      Escolha o tipo de fita a ser utilizado:      *
*
*          1) LTO2                                  *
*          2) LTO3                                  *
*          3) Sair                                   *
*
*****

  Opção: 2

*****
*
*      Escolha a unidade a ser utilizada:          *
*
*          7) Unidade 7                             *
*          10) Unidade 10                          *
*          3) Voltar e Corrigir                    *
*          4) Sair                                  *
*
*****

  Opção: 11

Opção inválida. Por favor, tente novamente...

*****
*
*      Escolha a unidade a ser utilizada:          *
*
*          7) Unidade 7                             *
*          10) Unidade 10                          *
*          3) Voltar e Corrigir                    *
*          4) Sair                                  *
*
*****

  Opção: 10
```

Figura11. Parte da execução do script para gravação

Na seqüência o script pedirá para o usuário escolher se deseja gravar (caso a fita seja virgem) ou apendar (caso a fita já contenha dados), depois é necessário digitar o nome da lista de saída na qual será escrito todo o processo

de gravação. Por ultimo por segurança é pedido para o usuário conferir se todos os dados estão corretos para continuar.

## 6.2. Recuperação de dados

Da forma que foi proposta a etapa, tornou-se necessário a criação de um script para realizar primeiramente a mudança da nomenclatura dos dados:

```
#!/bin/csh -x

set CANAL1 = '$10216950_'
set CANAL2 = '$10216952_'
set CANAL3 = '$10216954_'
set CANAL4 = '$10216956_'
set CANAL5 = '$10216958_'
set CANAL1CTL = '$10216951_'
set CANAL2CTL = '$10216953_'
set CANAL3CTL = '$10216955_'
set CANAL4CTL = '$10216957_'
set CANAL5CTL = '$10216959_'

cd /samfs5/Enrique/Goestr_Bin/REPRO/DLT0282

find . -type f > /samfs3/arq_ren4.txt

set NLINHAS = `cat /samfs3/arq_ren4.txt | wc -l`

while({(NLINHAS) > 0})

    set CAMINHO = `cat /samfs3/arq_ren4.txt | tail -${NLINHAS} | head -1`
    #chmod 755 ${CAMINHO}
    set ARQ = `basename ${CAMINHO}`
    set ANO = `echo ${ARQ} | cut -c 5-6`
    set MES = `echo ${ARQ} | cut -c 7-8`
    set DIA = `echo ${ARQ} | cut -c 9-10`
    set HORA = `echo ${ARQ} | cut -c 12-13`
    set MIN = `echo ${ARQ} | cut -c 14-15`

    set CTL = `echo ${ARQ} | grep ".ctl." | wc -l`
    if({(CTL) > 0}) then

        set CH = `echo ${ARQ} | grep "ch1." | wc -l`
        if({(CH) > 0}) then
            mkdir -p /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch1_bin/20${ANO}/${MES}/
            cp ${CAMINHO} /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch1_bin/20${ANO}/${MES}/${CANAL1CTL}20${ANO}${MES}${DIA}${HORA}${MIN}.ctl.gz
        else
            set CH = `echo ${ARQ} | grep "ch2." | wc -l`
            if({(CH) > 0}) then
                mkdir -p /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch2_bin/20${ANO}/${MES}/
                cp ${CAMINHO} /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch2_bin/20${ANO}/${MES}/${CANAL2CTL}20${ANO}${MES}${DIA}${HORA}${MIN}.ctl.gz
            else
                set CH = `echo ${ARQ} | grep "ch3." | wc -l`
                if({(CH) > 0}) then
                    mkdir -p /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch3_bin/20${ANO}/${MES}/
                    cp ${CAMINHO} /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch3_bin/20${ANO}/${MES}/${CANAL3CTL}20${ANO}${MES}${DIA}${HORA}${MIN}.ctl.gz
                else
                    set CH = `echo ${ARQ} | grep "ch4." | wc -l`
                    if({(CH) > 0}) then
                        mkdir -p /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch4_bin/20${ANO}/${MES}/
                        cp ${CAMINHO} /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch4_bin/20${ANO}/${MES}/${CANAL4CTL}20${ANO}${MES}${DIA}${HORA}${MIN}.ctl.gz
                    else
                        set CH = `echo ${ARQ} | grep "ch5." | wc -l`
                        if({(CH) > 0}) then
                            mkdir -p /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch5_bin/20${ANO}/${MES}/
                            cp ${CAMINHO} /samfs3/ENRIQUE_RENOMEADO_DLT0282/ch5_bin/20${ANO}/${MES}/${CANAL5CTL}20${ANO}${MES}${DIA}${HORA}${MIN}.ctl.gz
                        endif
                    endif
                endif
            endif
        endif
    endif
done

Line: 1/114      Column: 1      Character: 35 (0x23)
```

Figura12. Script para renomear os dados

Primeiramente, foi feito um levantamento das fitas DLT que seriam recuperadas, e foram escolhidas aquelas que apresentavam determinado produto, no caso, retangular 4km, então foi consultado no BDI a nomenclatura, visto que o produto apresenta 5 canais diferentes, do canal 1 ao canal 5.

As fitas DLT foram baixadas e armazenadas em um diretório específico no disco de RESTORE. Pelo script foi passado esse caminho dos dados, no qual foi feita uma varredura em todo o diretório e jogado para uma listagem todo seu conteúdo, posteriormente o script abriu essa listagem e foi verificando linha por linha para identificar os cinco diferentes canais do produto, sendo que cada um caía em uma diferente condição, cada tipo teria sua própria nomenclatura.

À medida que eram recuperados os dados, eles eram enviados para outro disco, e já não eram armazenados de forma confusa, cada nova fita tinha seu novo diretório com seus respectivos canais. Depois dessa etapa finalizada, fomos para a parte mais importante que foi a geração do log. de cada arquivo para serem enviados ao banco de dados de imagens, ali se concretizava de que o dado já estava recuperado e disponível para os usuários.

Claro, que também foi preciso fazer um backup desses dados, já que não seria possível deixá-los armazenados no disco pelo tamanho que ocupariam (temporariamente seria possível), então a fonte deles são as fitas LTO, aí vem a importância de tempo em tempo serem atualizadas para que não aja perda dos dados, isso precisa ser revisto, pois não possui espaço suficiente em disco para a armazenagem de todo esse volume, então a fita vem se tornando a única fonte.

Uma boa alternativa seria a existência de um disco espelho para cada disco operacional (disco onde os dados caem diretamente), no final de cada mês o disco operacional seria limpo e os dados jogados para o espelho, garantindo a integridade do principal.

Ao todo foi mais de 500 Gb de dados recuperados, um número significativo, tomando parte de que dados até mesmo de 2003 até 2007 foram recuperados. O caminho ainda é longo, visto que menos de 10% do total dados foram recuperados pelo fato de que a recuperação de dados é algo trabalhoso, e isso foi realizado somente nesse período, e com isso vamos evoluindo gradativamente.

## 7. PRINCIPAIS ATIVIDADES DESENVOLVIDAS

- Estudo do modelo de requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil), modelo que tem total influência no quesito gerenciamento de dados, o que nos fez parte;
- Leitura e exercício completo da apostila "Introdução ao LINUX e Programação em Script - Shell", da Universidade Federal Fluminense;
- Backup manual, gerando listagens e organizando-as para serem gravadas nas unidades LTO 3 e LTO 2, através de linhas de comandos disparando scripts para gravação utilizando acesso remoto;
- Desenvolvimento de um script para auxiliar no processo de armazenamento de dados em fitas LTO, visando mais agilidade e segurança.
- Desenvolvimento de scripts com a finalidade de passar todos os possíveis dados antigos, para o novo modelo padrão, incluindo dados reprocessados que apresentava horário errado, esses dados tiveram de ser corrigidos com suas devidas datas e horários;
- Recuperação dos dados históricos, referentes aos anos de 2003, 2004, 2005, 2006 e 2007;
- Geração do log. desses dados recuperados e envio dos mesmos ao BDI;

## **8. CONSIDERAÇÕES FINAIS**

Dentro do projeto que foi proposto e realizado, os principais objetivos foram realizados de forma eficiente, no plano inicial foi relatado a necessidade de auxiliar no desenvolvimento em Java script, até mesmo foi realizado um estudo do mesmo, por acaso podemos imaginar que faltou mais implementação de Java no projeto, mas isso se deu ao fato de que o Shell era o caminho mais simplificado e eficiente para a realização das principais metas traçadas, que vêm a ser, as ferramentas para agilizar o backup e a recuperação dos dados históricos.

Com isso tudo, consegui perceber que foi apenas o começo, pois vamos programar e finalizar um projeto Java web para fazer o cadastramento de todas as mídias, e também dar seqüência na recuperação de dados, que na realidade é muito maior do que podemos imaginar.

## **9. AGRADECIMENTOS**

Todo esse período de experiência foi muito produtivo e válido para minha carreira profissional. O meu objetivo pessoal agora é com certeza dar seqüência naquilo que pude começar através dessa oportunidade dada pelo CNPq, ao qual sou muito grato, através dessa bolsa, estou concluindo minha primeira experiência de trabalho.

Também gostaria de agradecer meu orientador Wagner, que foi uma pessoa muito prestativa quando necessitei, sempre atento ao calendário das atividades a serem cumpridas, ele realmente me orientou, por isso lhe digo muito obrigado.

Carlos Alexandre de Oliveira Rodrigues

Bolsista PIBIC

Wagner Flauber Araújo Lima

Orientador