

ORGANIZAÇÃO DO NOVO CLUSTER DE PRÉ-PROCESSAMENTO

O *cluster* de pré-processamento é denominado **preproc** e é composto pelas máquinas **coata.cptec.inpe.br** e **cambui.cptec.inpe.br** rodando o **Sistema Operacional Tru64**.

Esse cluster veio a substituir o antigo *cluster* de pré-processamento, o qual era composto pelas máquinas **igapy.cptec.inpe.br** e **aracapa.cptec.inpe.br** que também utilizavam o **Sistema Operacional Tru64**.

Com o novo *cluster*, espera-se ter um ganho maior tanto em **velocidade** quanto em **capacidade** de processamento, visto que as máquinas são mais novas, tendo portanto maior desempenho.

O novo *cluster* deve ser organizado segundo a estrutura ao lado.

Abaixo do diretório **Raiz (/)**, devem existir 2 diretórios, sendo eles: **data** e **brutos** – além dos diretórios normais instalados pelo sistema. Esses dois diretórios são *filesystems* diferentes. A estrutura proposta proporciona facilidade no acesso aos dados, tanto para pesquisa quanto para armazenamento (*storage*), manutenção e identificação de problemas, contudo uma parte dos dados brutos deve ser duplicada a fim de facilitar o acesso de outros processos operacionais. Futuramente a duplicação não será necessária, visto que um **Banco de Dados Brutos** será implementado. O desenvolvimento de tal aplicativo, já se encontra em fase de testes. Pode-se então afirmar que em pouco tempo teremos um Banco de Dados Brutos ativo e funcional.

ESTRUTURA DE DIRETÓRIOS

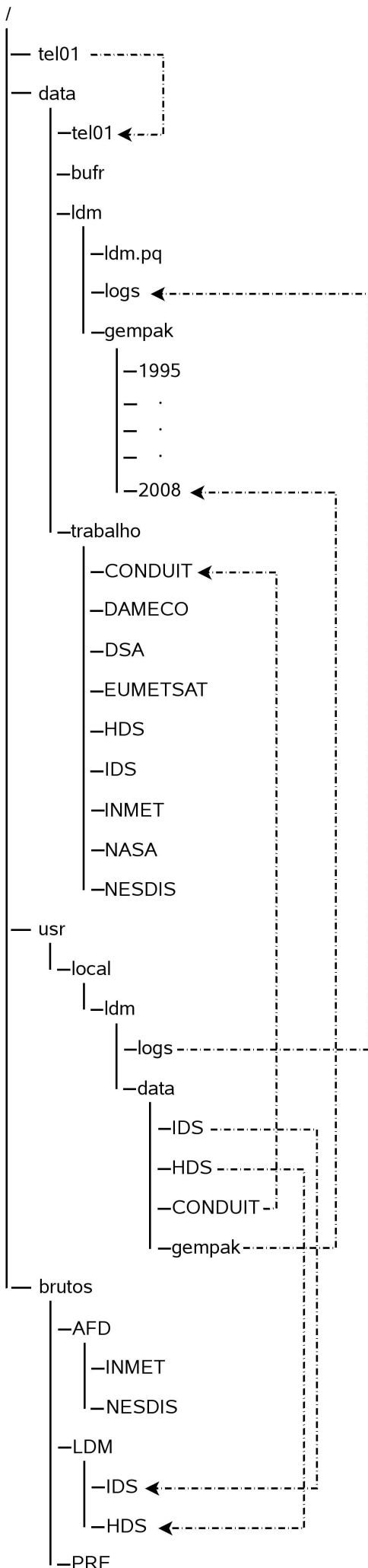
A estrutura de diretórios foi organizada tendo em mente 3 critérios:

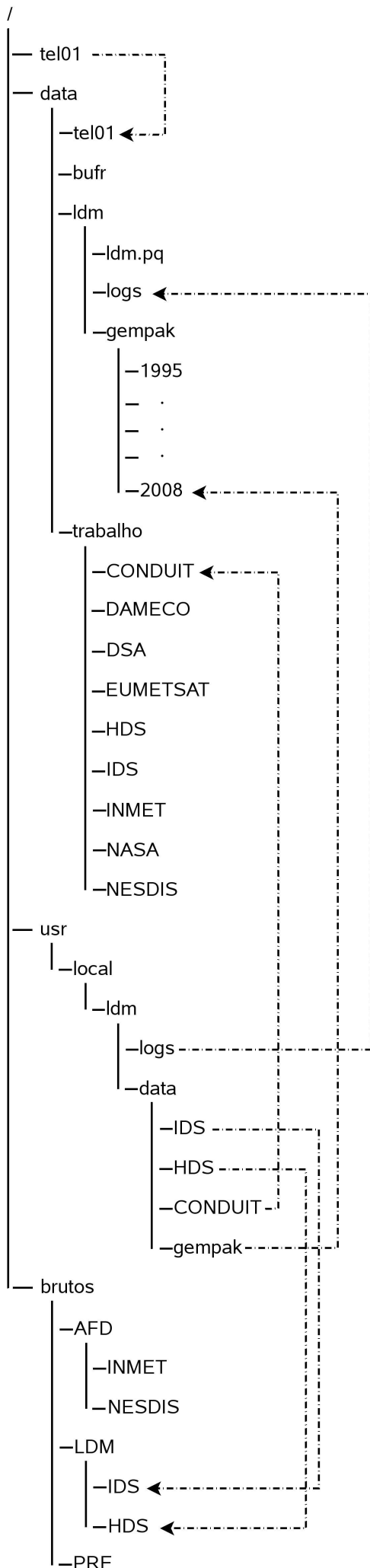
- Facilidade de Acesso
- Facilidade de *Backup*
- Compatibilidade com o Sistema Atual

Baseado nesses 3 critérios, a estrutura apresentada foi desenvolvida.

- Prós:

- Facilidade de Acesso por parte de pesquisadores, meteorologistas e climatologistas;
- Facilidade para *Backups* em fita ou qualquer outra mídia/sistema;
- Facilidade para identificação das origens de falhas no recebimento de dados;
- Maior flexibilidade que a estrutura anterior quanto a organização;
- Compatibilidade com os processos e procedimentos do pré-processamento antigo;
- Separação de dados de uma forma que possibilita melhor visão geral do sistema;





- Contras:

- Duplicidade de uma parte dos dados brutos;
- Dependência de certos diretórios para o ideal funcionamento de *softwares* e do antigo pré-processamento.

COMO FUNCIONA A ESTRUTURA DE DIRETÓRIOS

/tel01

O **/tel01** é o diretório de trabalho onde todas as rotinas do atual pré-processamento são rodadas. Ele é indispensável visto que todos o processos da atual rotina de pré-processamento procuram por esse diretório. No novo *cluster*, esse diretório é um *link* para o o diretório **/data/tel01**, onde todos os *scripts* do antigo pré estão armazenados e são executados.

/data

Esse diretório contém todos os dados que foram recebidos pelo centro e processados de alguma forma. É aqui que os dados trabalhados são armazenados. É aqui que fica armazenada a base de dados do **gempak** (**/data/ldm/gempak**) contendo todos os anos acessíveis e decodificados pelo mesmo. Aqui também se encontra o *queue file* do ldm e o diretório de **logs** do mesmo, bem como o diretório **tel01**, o **bufr** e o **predados**.

/data/bufr

Nesse diretório encontram-se os *scripts* para lidar com arquivos no formato bufr além de alguns dados nesse mesmo formato. Aqui encontram-se também as tabelas buf e vários

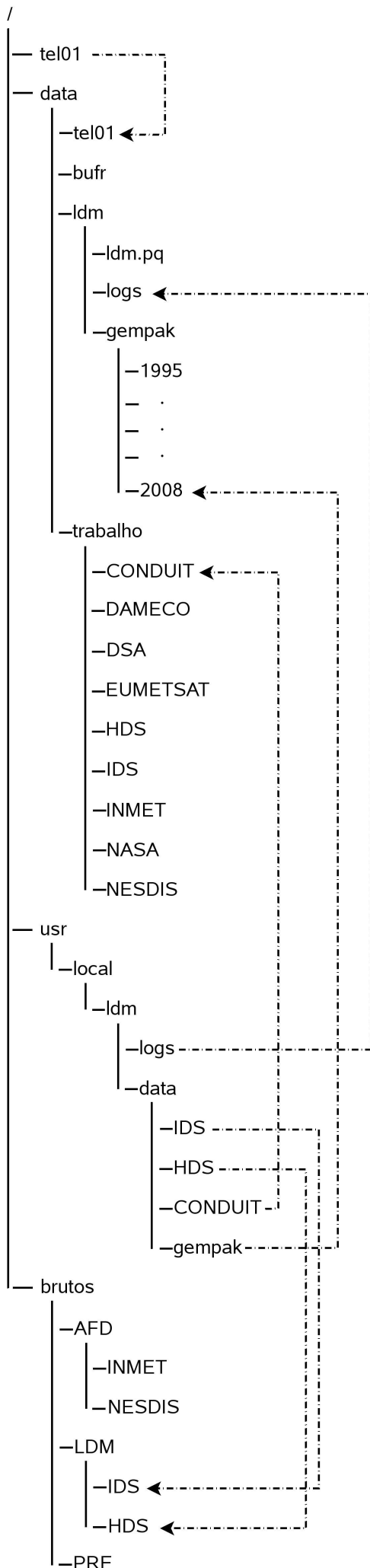
/data/ldm

Nesse diretório encontram-se o queue file do ldm, o diretório de logs do ldm e a base de dados gempak.

/data/trabalho

Esse diretório possui 9 sub-diretórios: **CONDUIT, DAMECO, DSA, EUMETSAT, HDS, IDS, INMET, NASA, NESDIS**. Nesses diretórios os dados são gravados utilizando a data dos boletins no nome do arquivo sem uma sub-divisão diretórios por ano ou seja, todos os dados são gravado no mesmo local sem seguir a estrutura **ano / mês / dia**. Possui como finalidade o acesso para as atividades operacionais. Ele contém uma replicação dos dados da área **/brutos/LDM** sendo, portanto, desnecessário possuir dados muito antigos, sendo mantidos apenas **3 dias de dados**. Os dados dentro dessa estrutura são, portanto, temporários.

OBS: Essa estrutura varia conforme a necessidade dos serviços operacionais.



/data/trabalho/CONDUIT

Nesse diretório são gravados os modelos de alta resolução do NCEP que são obtidos através do LDM. Possui 1 sub-diretório: **gfs**, que por sua vez possui 1 subdiretório: **prod**. Os arquivos são gravados aqui em diretórios divididos por horários. Aqui é mantido **1 dia de dados**.

/data/trabalho/DAMECO

Nesse diretório são gravados os Dados Meteorológico Consistidos (DAMECO) do INMET. São obtidos através do LDM e armazenado diretamente nesse diretório. Aqui são mantidos 600 dias de dados, uma vez que os arquivos são pequenos e constituem um volume de 1 arquivo por dia.

/data/trabalho/DSA

Nesse diretório são gravados os dados SATOB, SATEM e as PCDs do DSA recebidos através do AFD. Possui dois sub-diretórios: SATOB e PCD_TXT. São mantidos 3 dias de dados aqui.

/data/trabalho/EUMETSAT

Nesse diretório são gravados os mesmos dados SATOB que o DSA recebe. Esses dados chegam através do AFD. Possui um único sub-diretório chamado SATOB no qual os arquivos são gravados.

/data/trabalho/HDS

Nesse diretório são gravados os dados do fluxo HDS (que chegam através do LDM). São 3 tipos de dados: ATOVS, QUICKSCAT e SATOB. Cada um desses dados é armazenado em um sub-diretório separado.

/data/trabalho/IDS

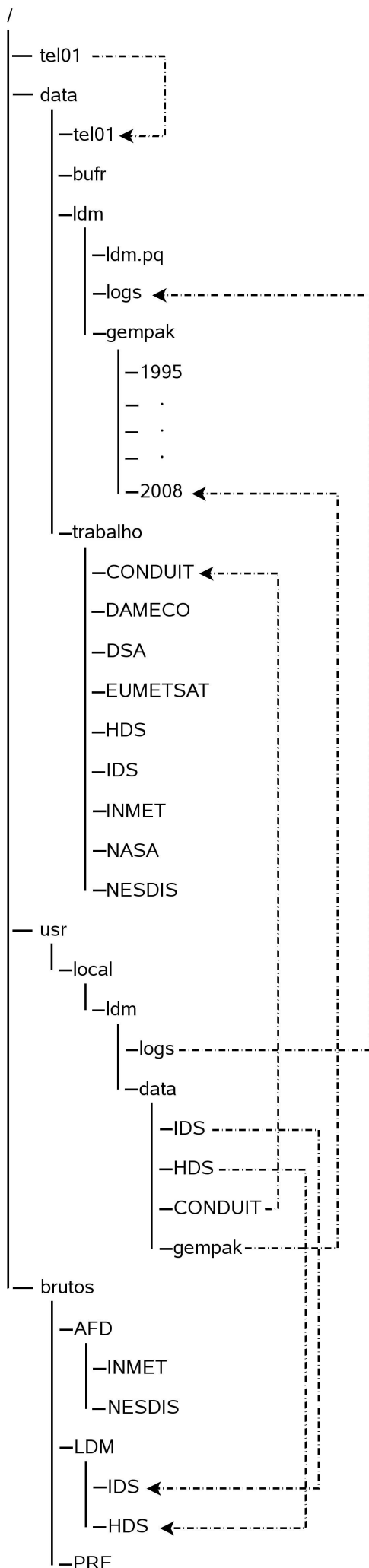
Nesse diretório são gravados os dados dos fluxos IDS e DDPLUS (que chegam através do LDM) e do INMET (obtidos através do AFD e do LDM). Possui 6 subdiretórios que organizam as mensagens por tipo (airep, buoy, metar, pilot, synop e temp).

/data/trabalho/INMET

Nesse diretório são gravados os dados do INMET. Possui 2 diretórios: OBS e BIN. No OBS são gravados os dados que chegam através do AFD. Esses dados são inseridos no LDM para serem processados pelo Pré-processamento. No diretório BIN são gravados os dados binários do INMET. Essa gravação é feita pelo LDM, que divide o diretório em vários sub-diretórios para separar os dados por tipo.

/data/trabalho/NASA

Nesse diretório são gravados os dados AIRS da NASA (que chegam através do AFD). Possui um sub-diretório para separar os dados por tipo.



/data/trabalho/NESDIS

Nesse diretório são gravados os ATOVS e QUICKSCAT que chegam através do AFD. Possui dois sub-diretórios para separar os dois tipos de dados que estão aqui armazenados.

/usr/local/ldm

Esse diretório é o diretório home do usuário ldm. É aqui que o ldm é instalado. Nesse diretório deve existir um outro diretório com o nome de **data** (**/usr/local/ldm/data**) e outro com o nome de **log** para que o ldm funcione corretamente. No novo cluster, dentro desse diretório, deverão existir 4 links simbólicos apenas. Esses links existem apenas para facilitar a manutenção do sistema. São 4 links **IDS**, **HDS**, **CONDUIT** e **gempak**. Os links IDS, HDS e CONDUIT apontam para os respectivos diretórios na partição **/brutos**, onde dados desse tipo são armazenados. O link **gempak** aponta para o diretório correspondente ao ano corrente no diretório **/data/ldm/gempak**. Esse link deve ser alterado todo início de ano.

/brutos

Tem como objetivo armazenar os dados brutos para **backup**. Existem 3 diretórios abaixo dele: **AFD**, **LDM** e **PRE**. Os diretórios são organizados da seguinte forma: <FONTE> / <FEED> onde, FONTE representa o software através do qual o dado chegou e FEED, a fonte que enviou os dados.

/brutos/AFD

Nesse diretório são salvos as mensagens brutas que chegam por meio do **AFD**. São dois fluxos (feeds) ao todo: **INMET** e **NESDIS**. Os dados gravados sob esse diretório utilizam o horário de chegada em sua nomenclatura.

/brutos/LDM

Nesse diretório são salvos as mensagens brutas que chegam por meio do **LDM**. São dois fluxos (feeds) ao todo: **IDS** e **HDS**. Os dados gravados sob esse diretório utilizam o horário de chegada em sua nomenclatura.

/brutos/PRE

Nesse diretório são salvos os dados que o pré-processamento utiliza para decodificar e carregar o BDM. Os scripts do diretório **tel01** buscam aqui os dados.

REDE E DIRETÓRIOS EXPORTADOS

O cluster de pré-processamento é uma máquina operacional que se encontra fora do servidor NIS e é classificada como **super crítica**. Ela possui um disco que é exportado para que outros setores tenham acesso aos dados, principalmente o setor de assimilação de dados e previsão de tempo.

O diretório **/data/trabalho** é diretório exportado para essa rede. É através dele que os outros setores do centro têm acesso aos dados.

PROCEDIMENTOS DE STORAGE

Os dados que devem passar pelos procedimentos de storage encontram-se todos no diretório **/brutos/AFD** e **/brutos/LDM**. A estrutura de dados nesses diretórios segue a seguinte regra:

/ brutos / <FONTE> / <FEED> / <PRODUTO> / <ANO> / <MES> / <DIA>

Exemplos:

/ brutos / AFD / NESDIS / ATOVS / 2007 / 01

/ brutos / LDM / IDS / synop / 2007 / 01

Será utilizado um software corporativo, o legatto, para efetuar armazenamento (storage) automático desses dados no robô de fitas do CPTEC / INPE (máquina bangu.cptec.inpe.br). Esse armazenamento deverá ser feito uma vez por mês, sendo que o volume mensal de dados é de aproximadamente, 10 GB.

O Anexo 1 contém informações mais detalhadas sobre o volume de dados e o processo de storage.

CONCLUSÃO E FUTURO

A idéia é que o novo pré-processamento que está sendo desenvolvido, seja colocado em operação nesse cluster novo. Com a reformulação da estrutura de diretórios, o funcionamento dos processos atuais não se altera porém, o acesso e a utilização desses dados torna-se mais padronizada.

Essa reformulação padroniza o local onde os dados devem ser colocados e estabelece uma função específica para cada diretório podendo eles, serem usados também, para identificar falhas no recebimento de dados.

Essa estrutura ainda não é a ideal, devendo sofrer algumas modificações ao longo do ano para atender a novas necessidades.