

# **Manipulação de Dados do HPCC**

**Equipe de documentação de Boca Raton**



## Manipulação de Dados do HPCC

Equipe de documentação de Boca Raton

Copyright © 2021 HPCC Systems®. All rights reserved

Sua opinião e comentários sobre este documento são muito bem-vindos e podem ser enviados por e-mail para <docfeedback@hpccsystems.com>

Inclua a frase **Feedback sobre documentação** na linha de assunto e indique o nome do documento, o número das páginas e número da versão atual no corpo da mensagem.

LexisNexis e o logotipo Knowledge Burst são marcas comerciais registradas da Reed Elsevier Properties Inc., usadas sob licença.

HPCC Systems® é uma marca registrada da LexisNexis Risk Data Management Inc.

Os demais produtos, logotipos e serviços podem ser marcas comerciais ou registradas de suas respectivas empresas.

Todos os nomes e dados de exemplo utilizados neste manual são fictícios. Qualquer semelhança com pessoas reais, vivas ou mortas, é mera coincidência.

2021 Version 8.2.80-1

<i>Manipulação de Dados do HPCC</i> .....	4
Introdução .....	4
Terminologia de Manipulação de Dados .....	5
Trabalhando com arquivo de dados .....	6
Métodos de Manipulação de Dados .....	9
Manipulação de Dados utilizando o ECL Watch .....	13
<i>Backup de Dados do HPCC</i> .....	25
Introdução .....	25
Dados do Dali .....	26
Arquivos de Configuração do Ambiente .....	27
Arquivos de Dados Thor .....	28
Arquivos de dados Roxie .....	30
Repositórios de atributos .....	31
Arquivos da Zona de entrada de arquivos .....	32

# ***Manipulação de Dados do HPCC***

## **Introdução**

Há diferentes formas pelas quais os dados podem ser transferidos de, para ou dentro de um HPCC system. Para cada uma dessas transferências de dados, há alguns parâmetros-chave que precisam ser conhecidos.

### **Pré-requisitos para a maior parte das manipulações dos arquivos:**

- Nome do arquivo lógico
- Nome do arquivo físico
- Tamanho do registro (fixo)
- Diretório de origem
- Diretório de destino
- Endereço IP do Dali (origem e/ou destino)
- Zona de entrada de arquivos

Os parâmetros acima são usados para esses principais métodos de manuseio de dados:

- Importar - Fazer o spray dos dados da Zona de entrada de arquivos para o Thor
- Exportar - Fazer o despray dos dados do Thor para a Zona de entrada de arquivos
- Copiar - Replicar os dados do Thor para o Thor (em um mesmo sistema de arquivos Dali)
- Copiar dados do Thor para o Thor (entre sistemas de arquivo Dali diferentes)

# Terminologia de Manipulação de Dados

Um *spray* ou *importação* é a transferência de um arquivo de dados de um local (como a zona de entrada de arquivos) para um cluster. O termo *spray* foi adotado devido à natureza da transferência dos arquivos – o arquivo é particionado entre todos os nós em um cluster.

Um *despray*, ou *exportação*, é a transferência de um arquivo de dados de um cluster para um único local da máquina (como uma zona de entrada de arquivo). O termo *despray* foi adotado em função da natureza da transferência dos arquivos – o arquivo é então remontado a partir de suas peças em todos os nós no cluster e colocado em um único arquivo no destino.

Uma *cópia* é a replicação de um arquivo de dados de um cluster para outro cluster no mesmo ambiente.

Uma *Cópia remota* é a replicação de um arquivo de dados de um cluster para outro cluster em um ambiente diferente.

Uma *Zona de entrada de arquivo* (ou zona de chegada) é um local de armazenamento físico definido no ambiente do seu sistema. É possível definir um ou mais desses locais. Um daemon (DaFileSrv) precisa estar em execução no servidor para possibilitar sprays e desprays do arquivo.

# Trabalhando com arquivo de dados

Depois de começar a trabalhar com seu HPCC System, é recomendado processar alguns dados reais; esta seção mostra como carregar dados no HPCC System.

## Antes de começarmos

Depois de começar a trabalhar com seu HPCC System, é recomendado processar alguns dados reais; esta seção mostra como carregar dados no HPCC System. Um HPCC System de produção típico teria uma capacidade de dados maior do que um sistema de desenvolvimento. O tamanho dos arquivos com os quais você deseja trabalhar é limitado pelo tamanho do seu sistema.

## Enviando um arquivo

Para arquivos de dados menores, com tamanho máximo de 2GB, você pode usar o utilitário upload/download file no ECL Watch.

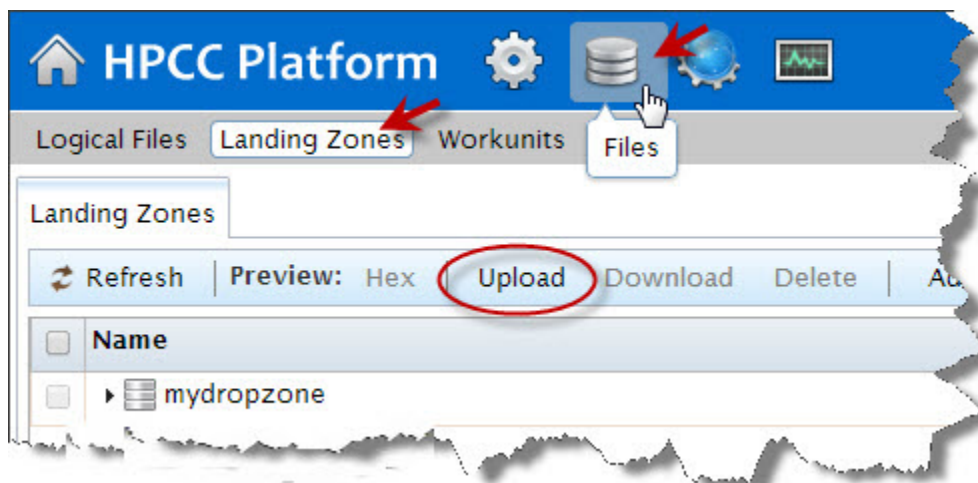
1. Em seu navegador, acesse o URL do **ECL Watch** de exemplo exibido, <http://nnn.nnn.nnn.nnn:8010>, onde nnn.nnn.nnn.nnn é o endereço IP do servidor ESP.



Seu endereço IP poderá ser diferente dos endereços fornecidos nas imagens de exemplo. Use o endereço IP fornecido pela **sua** instalação.

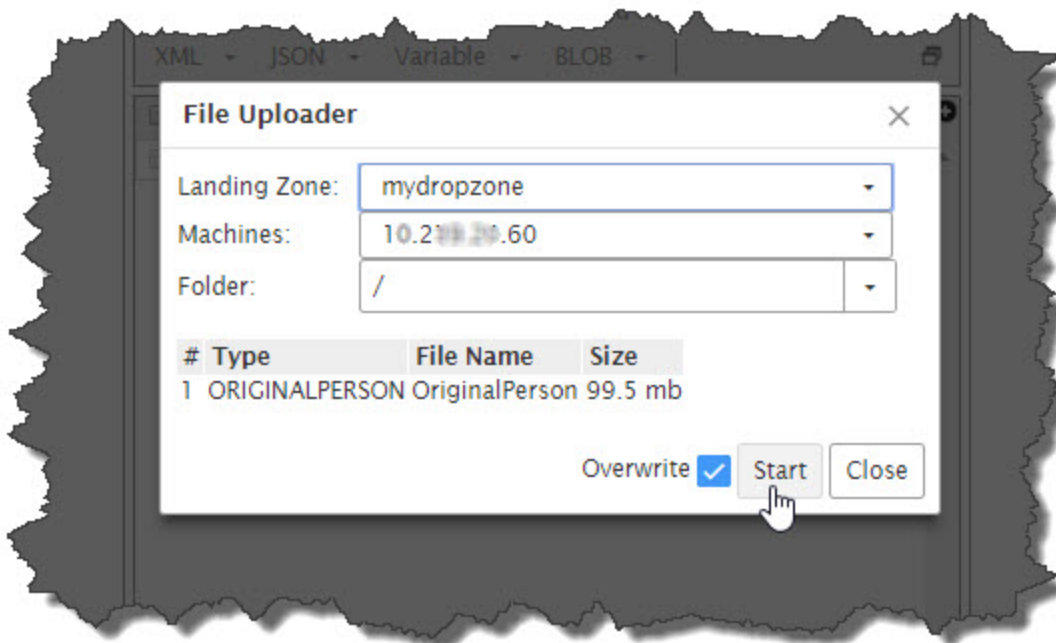
2. Na página do ECL Watch, clique no link **Upload/Download File** localizado no menu ao lado esquerdo.

**Figure 1. Upload/download**



Após clicar no link Upload/Download File, você será direcionado para a página de zona de entrada de arquivos; lá é possível **procurar** em seu computador o arquivo a ser enviado:

**Figure 2. Zonas de Entrada**



3. Pressione o botão **Browse** para navegar pelos arquivos em seu computador local, selecione o arquivo a ser enviado e clique no botão **Open**.

O arquivo selecionado deverá aparecer no campo “Selecione um arquivo para enviar!”.

4. Pressione o botão **Upload Now** para concluir o envio do arquivo.

## Enviando um arquivo com uma Cópia de Segurança

Para enviar um arquivo maior para processamento para sua máquina virtual, será necessária uma ferramenta com suporte para o protocolo de cópia de segurança. Nesta seção, abordaremos o uso do WinSCP. Há outras ferramentas disponíveis, mas os passos são similares.

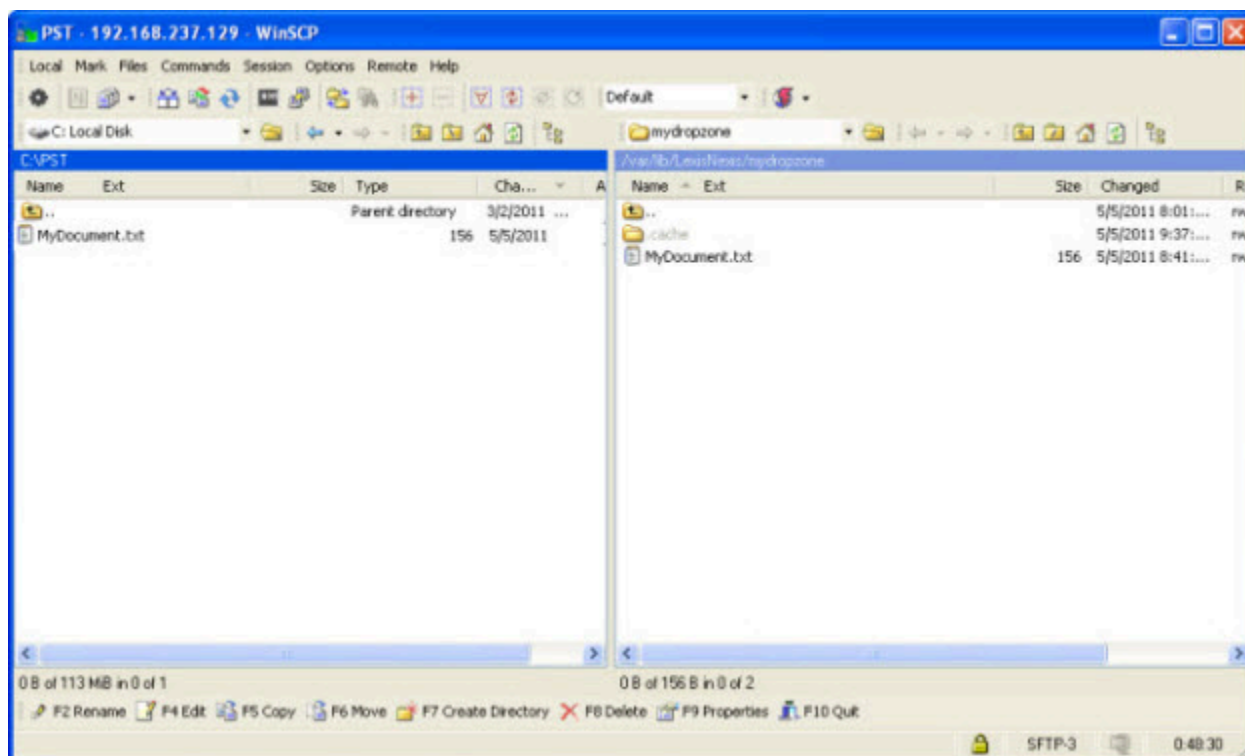
1. Abra a ferramenta WinSCP e faça login no nó da zona de entrada de arquivo usando o nome de usuário e senha fornecidos.

Login ID:	hpccdemo
Password:	hpccdemo

2. Depois de fazer login, o sistema irá navegar automaticamente até a pasta da zona de entrada de arquivo. (/var/lib/LexisNexis/mydropzone)

3. Navegue onde seu arquivo local está na parte esquerda da janela.

**Figure 3. WinSCP**



4. Selecione o arquivo de dados para envio e copie-o para a zona de entrada de arquivos usando o recurso de arrastar e soltar.



# **Métodos de Manipulação de Dados**

Há várias maneiras de fazer o spray, despray ou copiar arquivos de dados:

- A interface DFU no ECL Watch
- O utilitário de linha de comando DFU Plus

Consulte o manual *Ferramentas de cliente* para obter mais detalhes

- Como usar o código ECL e as funções de biblioteca FileServices.

Consulte a *Referência a linguagem ECL* para obter detalhes.

## Spray de Arquivos

Uma maneira de fazer o spray de arquivos para seus clusters é através da página **Zona de entrada de arquivo** no ECL Watch.

1. Selecione o arquivo na zona de entrada de arquivos marcando a caixa ao lado dele.
2. Selecione a opção adequada para o tipo de spray desejado no menu suspenso.

Por exemplo, para fazer o spray de um arquivo delimitado, selecione o botão de ação **Delimited** .

**Figure 4. Spray para a Zona de entrada de arquivos**

The screenshot shows the HPCC Target configuration window. At the top, there are tabs for 'Fixed', 'Delimited', 'XML', 'JSON', 'Variable', and 'BLOB'. The 'Target' tab is selected. Below the tabs, there are fields for 'Group' (set to 'mythor'), 'Queue' (set to 'dfuserver\_queue'), and 'Target Scope' (set to 'some:prefix'). A 'Target Name' field contains 'actors.list'. Below these fields is an 'Options' section with various settings: 'Format' is set to 'ASCII'; 'Max Record Length' is '8192'; 'Separators' is '\'; 'Omit Separator' is unchecked; 'Escape' is empty; 'Line Terminators' is '\n,\r\n'; 'Quote' is '"'; 'Overwrite' is checked; 'Replicate' is checked; 'Compress' is unchecked; 'Record Structure Present' is unchecked; 'Fail If No Source File' is unchecked; 'Quoted Terminator' is unchecked; and 'Expire in (days)' is empty. A 'Spray' button is located at the bottom right of the window.

3. Preencha os valores conforme necessário para o spray.
4. Pressione o botão **Spray** para realizar o spray do(s) arquivo(s).

## Spray de múltiplos arquivos

É possível optar por realizar o spray de mais de um arquivo com o recurso de spray de múltiplos arquivos. Isso é útil para realizar o spray de vários arquivos do mesmo tipo usando as mesmas opções de spray.

Arquivos (de comprimento) fixo podem ter diferentes tamanhos de registro e arquivos XML podem ter diferentes nomes de linha que precisam ser especificados individualmente para cada arquivo. Para especificar essas diferenças, selecione os arquivos para os quais você deseja fazer o spray e o tipo de spray. Você então verá os arquivos listados. Caso estiver utilizando tipos fixo de spray ou XML, insira as informações de tamanho de registro ou nome da linha. Marque todas as outras opções relevantes e faça o spray.

# Manipulação de Dados utilizando o ECL Watch

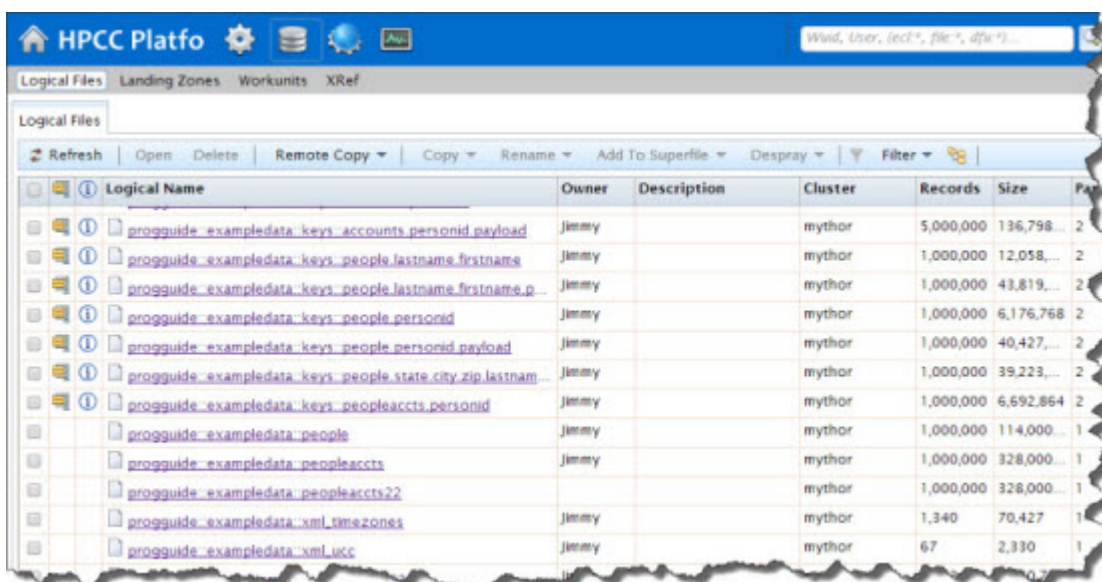
- Faça login no ECL Watch para o ambiente.

A URL é o endereço IP no qual o ESP Server é instalado seguido da porta à qual o serviço WsSMC está vinculado. A porta padrão é 8010. Por exemplo:

```
http://<ESPserverIP>:8010/
```

- Clique no ícone **Files** e selecione **Logical Files**.

A página Logical Files é aberta, mostrando todos os arquivos com registros lógicos no sistema de arquivos distribuídos do servidor Dali.



The screenshot shows the HPCC Platform interface with the 'Logical Files' tab selected. The table lists various logical files with columns for Logical Name, Owner, Description, Cluster, Records, Size, and Pages. The files are categorized by 'proguides' and 'people' data sets.

Logical Name	Owner	Description	Cluster	Records	Size	Pages
proguides_exampdata_keys_accounts_personid_payload	Jimmy		mythor	5,000,000	136,798...	2
proguides_exampdata_keys_people_lastname_firstname	Jimmy		mythor	1,000,000	12,058...	2
proguides_exampdata_keys_people_lastname_firstname_p...	Jimmy		mythor	1,000,000	43,819...	2
proguides_exampdata_keys_people_personid	Jimmy		mythor	1,000,000	6,176,768	2
proguides_exampdata_keys_people_personid_payload	Jimmy		mythor	1,000,000	40,427...	2
proguides_exampdata_keys_people_state_city_zip_lastname...	Jimmy		mythor	1,000,000	39,223...	2
proguides_exampdata_keys_peopleaccts_personid	Jimmy		mythor	1,000,000	6,692,864	2
proguides_exampdata_people	Jimmy		mythor	1,000,000	114,000...	1
proguides_exampdata_peopleaccts	Jimmy		mythor	1,000,000	328,000...	1
proguides_exampdata_peopleaccts22			mythor	1,000,000	328,000...	1
proguides_exampdata.xml_timezones	Jimmy		mythor	1,340	70,427	1
proguides_exampdata.xml_ucc	Jimmy		mythor	67	2,330	1

Nesta página, é possível realizar o despray ou copiar qualquer arquivo.

## Spray Fixo

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação Spray: **Botão de ação** Fixed

A caixa de diálogo **Fixed Spray** é mostrada.

- Preencha os detalhes necessários:

Destino	
<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Scope</b>	O prefixo do arquivo lógico
<b>Target Name</b>	O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.
<b>Record Length</b>	O tamanho de cada registro.
Opções:	
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>	
<b>Compress</b>	Marque esta caixa para compactar os arquivos.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.

- Pressione o botão **Spray**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Spray delimitado

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação **Spray Delimited**.

A página **Spray Delimited** será exibida.

- Preencha os detalhes necessários:

### Destino

<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Scope</b>	O prefixo do arquivo lógico
<b>Target Name</b>	O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.

### Opções:

<b>Format</b>	Selecione o formato na lista suspensa
<b>Max Record Length</b>	O tamanho do maior registro no arquivo.
<b>Separators</b>	O(s) caractere(s) usados como separadores no arquivo de origem.
<b>Omit Separator</b>	Marque essa caixa para omitir o separador.
<b>Escape</b>	Uma string terminada em nulo que contém caracteres escape do CSV.
<b>Line Terminators</b>	O(s) caractere(s) usados como terminadores de linhas no arquivo de origem.
<b>Quote</b>	O caractere usado como aspas no arquivo de origem.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).  <b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>
<b>Compress</b>	Marque esta caixa para compactar os arquivos.
<b>Quoted Terminator</b>	Marque essa caixa para indicar que o caractere delimitador pode ser incluído em campo em aspas. Caso não seja marcada, ela permite um particionamento mais rápido do arquivo (evitando uma varredura completa do arquivo).
<b>Record Structure Present</b>	Sinalizador que indica se a estrutura do registro deve ser derivada do cabeçalho do arquivo.

- Pressione o botão **Spray**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.



## Spray XML

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação **Spray XML**.

A caixa de diálogo **Spray XML** será exibida.

- Preencha os detalhes necessários:

Destino	
<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Scope</b>	O prefixo do arquivo lógico
<b>Target Name</b>	O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.
<b>Row Tag</b>	O nome do delimitador de linha. Requerido.
Opções:	
<b>Format</b>	Selecione o formato na lista suspensa
<b>Max Record Length</b>	O tamanho do maior registro no arquivo.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>	
<b>Compress</b>	Marque esta caixa para compactar os arquivos.

- Pressione o botão **Spray**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Spray JSON

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação **Spray JSON**.

A caixa de diálogo é exibida.

- Preencha os detalhes necessários:

Destino	
<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Scope</b>	O prefixo do arquivo lógico
<b>Target Name</b>	O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.
<b>Row Path</b>	O especificador de caminho para o conteúdo JSON. O padrão usa o conteúdo de nível de raiz como um conjunto de objetos a ser tratado como linhas.
Opções:	
<b>Format</b>	Selecione o formato na lista suspensa
<b>Max Record Length</b>	O tamanho do maior registro no arquivo.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>	
<b>Compress</b>	Marque esta caixa para compactar os arquivos.

- Pressione o botão **Spray**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Spray Variável

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação Spray: **Botão de ação Variável**.

A caixa de diálogo Spray **variável** é mostrada.

- Preencha os detalhes necessários:

Destino	
<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Scope</b>	O prefixo do arquivo lógico
<b>Target Name</b>	O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.

### Opções:

<b>Source Type</b>	Selecione o tipo de fonte na lista suspensa. Valores: recfmv, recfmvb, Variable, ou Variable Big-endian.n.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
	<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>
<b>Compress</b>	Marque esta caixa para compactar os arquivos.

- Pressione o botão **Spray**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Spray Blob

- Clique no ícone **Files** e no link **Landing Zones (Zona de entrada de arquivos)** localizados no submenu de navegação.

- Clique na seta ao lado da zona de entrada de arquivos para ampliar a lista.

Os arquivos na zona de entrada são exibidos.

- Marque as caixas de seleção dos arquivos para os quais você deseja fazer o spray e depois pressione o botão de ação **Spray BLOB** :

A caixa de diálogo **Spray BLOB** é mostrada.

- Preencha os detalhes necessários:

### Destino

<b>Group</b>	Selecione o nome do cluster para o qual você deseja fazer o spray. Só é possível selecionar um cluster no seu ambiente.
<b>Queue</b>	Selecione a fila.
<b>Target Name</b>	O nome do destino lógico a ser criado. Required. É necessário informar um nome de destino.
<b>Source Path</b>	O caminho até o arquivo. Ele é pré-preenchido com o nome dos arquivos de origem selecionados na zona de entrada de arquivos, mas pode ser alterado. Compatível com elementos curingas.

### Opções:

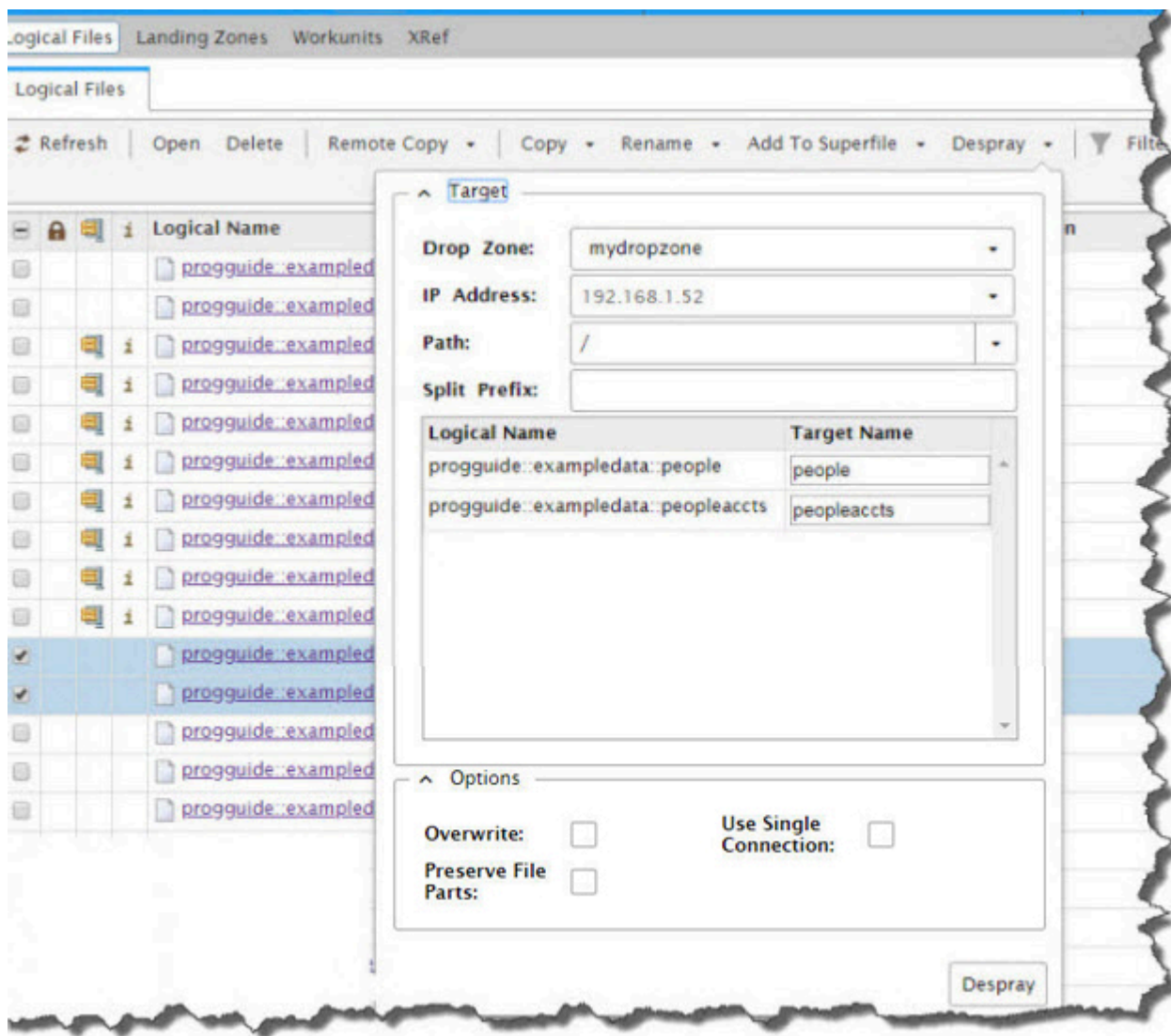
<b>Blob Prefix</b>	O prefixo do arquivo.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
	<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>
<b>Compress</b>	Marque esta caixa para compactar os arquivos.
<b>Fail if no source file</b>	Marque essa caixa para permitir que o spray falhe se nenhum arquivo de origem for encontrado.

- Pressione o botão **Spray** .

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de spray na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Desprays

- Localize o(s) arquivo(s) para o despray na lista de arquivos e depois pressione o botão de ação **Despray**.



- Insira informações sobre o **destino**.

### Drop Zone

Use a lista suspensa para selecionar a máquina para a qual será realizado o despray. Os itens na lista são zonas de entrada de arquivos definidos na configuração do sistema.

### IP Address

Isso é pré-preenchido com base na máquina selecionada.

### Path

Informe o caminho de arquivo completo do destino incluindo nome de arquivo e extensão.

### Split Prefix

Prefixo

### Overwrite

Marque essa caixa para substituir um arquivo com mesmo nome, se existir.

### Use Single Connection

Marque essa caixa para usar uma conexão de rede única para despray.

**Preservar partes do arquivo** Marque esta caixa para preservar as partes do arquivo original e gravar vários arquivos na landing zone.

- Pressione o botão **Despray** .

Uma aba DFU Workunit é exibida para cada job. É possível ver o andamento de cada operação de despray na guia. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Copiar

- Clique no ícone **Files** e no link Landing Zones (Zona de entrada de arquivos) localizados no submenu de navegação.
- Selecione o(s) arquivo(s) para copiar na lista de arquivos e depois clique no botão **Copy**.
- Preencha as **informações de Destino e Opções**

**Group** Seleccione o nome do cluster para o qual você vai fazer a cópia. Só é possível seleccionar um cluster no seu ambiente.

**Target Name** O nome de arquivo lógico a ser criado. Ele é pré-preenchido com o nome do arquivo de origem na zona de entrada de arquivos, mas pode ser alterado.

### Opções:

**Replicate** Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).

**Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.**

**Wrap** Marque essa caixa para manter o número de partes inalterado e encapsular se o cluster de destino for menor do que o original.

**No Split** Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.

**Overwrite** Marque esta caixa para substituir arquivos de mesmo nome.

**Compress** Marque esta caixa para compactar os arquivos.

**Retain Superfile Structure** Marque essa caixa para preservar a estrutura do super arquivo.

**Preserve Compression** Marque essa caixa para preservar a compactação do arquivo original ao copiar

- Pressione o botão **Copy**.

Uma aba **DFU Workunit** é exibida para cada job. É possível ver o andamento de cada operação de cópia na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

## Cópia Remota

A cópia remota permite copiar dados de um cluster fora do seu ambiente para um dentro do seu ambiente.

- Clique no ícone **Files** e no link Landing Zones (Zona de entrada de arquivos) localizados no submenu de navegação.
- Clique no link **Remote Copy**

A página **Copy File** é exibida.

- Preencha as informações de **Fonte**, **Destino** e **Opções**.

### Fonte:

<b>Dali</b>	O servidor Dali no ambiente remoto
<b>User ID</b>	O nome do usuário usado para ser autenticado no ambiente remoto (se necessário)
<b>Password</b>	A senha para autenticação no ambiente remoto (se necessária)
<b>Logical File</b>	O nome de arquivo lógico no ambiente remoto.

### Destino:

<b>Group</b>	Selecione o nome do cluster para o qual você vai fazer a cópia. Só é possível selecionar um cluster no seu ambiente.
<b>Logical Name</b>	O nome lógico do arquivo copiado.

### Opções:

<b>Replicate</b>	Marque essa caixa para criar cópias de backup de todas as partes do arquivo no diretório de backup (por convenção, na unidade secundária do nó seguindo no cluster).
	<b>Essa opção só está disponível em sistemas onde a replicação tiver sido ativada.</b>
<b>Wrap</b>	Marque essa caixa para manter o número de partes inalterado e encapsular se o cluster de destino for menor do que o original.
<b>Overwrite</b>	Marque esta caixa para substituir arquivos de mesmo nome.
<b>Compress</b>	Marque esta caixa para compactar os arquivos.
<b>No Split</b>	Marque esta caixa para impedir a divisão de partes do arquivo em várias partes de destino.
<b>Retain Superfile Structure</b>	Marque essa caixa para preservar a estrutura do super arquivo.

- Pressione o botão **Submit**.

Uma aba **DFU Workunit** é exibida. É possível ver o andamento das operações de cópia na aba. Se uma workunit falhar, as informações relacionadas à causa da falha também são exibidas.

- Pressione o botão **Refresh** periodicamente até o status da solicitação indicar **Finished** ou clique no hiperlink **View Progress** para ver um indicador de progresso.



# ***Backup de Dados do HPCC***

## **Introdução**

Esta seção abrange os dados críticos do sistema que requeiram procedimentos regulares de backup para impedir a perda de dados.

São eles:

- O armazenamento de dados do sistema (dados Dali)
- Arquivos de configuração do ambiente
- Arquivos de dados da refinaria de dados (Thor)
- Arquivos de dados do motor de entrega rápida de dados (Roxie)
- Repositórios de atributos
- Arquivos da Zona de entrada de arquivos

## **Dados do Dali**

Os arquivos do servidor Dali são normalmente espelhados em seu nó de backup. Essa localização está especificada no arquivo de configuração do ambiente através do Gerenciador de Configurações.

Uma vez que os dados são gravados simultaneamente em ambos os nós, não há necessidade de um procedimento de backup manual.

## **Arquivos de Configuração do Ambiente**

Existe apenas um arquivo de ambiente ativo, porém pode haver várias configurações alternativas.

O Gerenciador de Configurações funciona apenas em arquivos na pasta /etc/HPCCSystems/source. Para ativar a configuração, ela é copiada para /etc/HPCCSystems/environment.xml em todos os nós.

O Gerenciador de Configurações cria cópias de backup automaticamente na pasta /etc/HPCCSystems/source/backup/.

## Arquivos de Dados Thor

Os clusters Thor são normalmente configurados para replicar dados automaticamente para um segundo local, conhecido como localização espelhada. Normalmente, a localização espelhada está localizada na unidade secundária do nó subsequente.

Se os dados não forem encontrados no local principal (por exemplo, devido a uma falha de unidade ou porque um nó foi desativado), ele usará o diretório espelhado para ler os dados. Qualquer gravação vai para o local primário e depois para o espelhado. Isso oferece redundância contínua e um meio rápido de restaurar um sistema após a desativação de um nó.

O backup de dados do Thor deve ser realizado regularmente e sob demanda após a desativação de um nó.

## Backup Manual

Siga essas etapas para executar um backup manual:

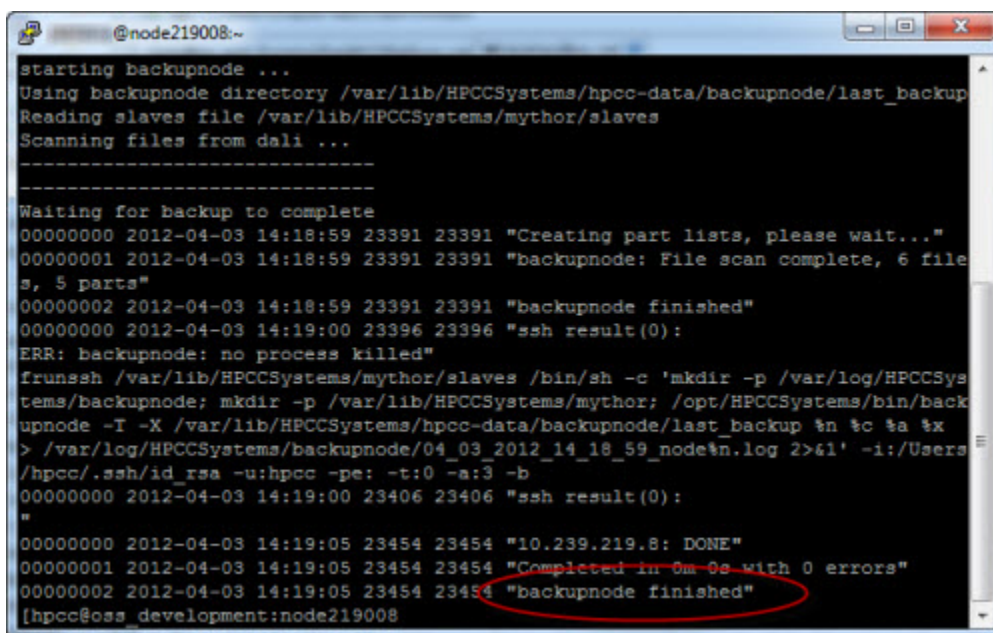
1. Faça o login no nó mestre do Thor.

Se não souber qual nó é considerado o nó mestre do Thor, pesquise usando o ECL Watch.

2. Execute o comando:

```
sudo su hpcc /opt/HPCCSystems/bin/start_backupnode <thor_cluster_name>
```

Isso iniciará o processo de backup.



```
@node219008:~
starting backupnode ...
Using backupnode directory /var/lib/HPCCSystems/hpcc-data/backupnode/last_backup
Reading slaves file /var/lib/HPCCSystems/mythor/slaves
Scanning files from dali ...
-----
Waiting for backup to complete
00000000 2012-04-03 14:18:59 23391 23391 "Creating part lists, please wait..."
00000001 2012-04-03 14:18:59 23391 23391 "backupnode: File scan complete, 6 file
s, 5 parts"
00000002 2012-04-03 14:18:59 23391 23391 "backupnode finished"
00000000 2012-04-03 14:19:00 23396 23396 "ssh result(0):
ERR: backupnode: no process killed"
frunssh /var/lib/HPCCSystems/mythor/slaves /bin/sh -c 'mkdir -p /var/log/HPCCSys
tems/backupnode; mkdir -p /var/lib/HPCCSystems/mythor; /opt/HPCCSystems/bin/back
upnode -I -X /var/lib/HPCCSystems/hpcc-data/backupnode/last_backup %n %c %a %x
> /var/log/HPCCSystems/backupnode/04_03_2012_14_18_59_node%n.log 2>&1' -i:/Users
/hpcc/.ssh/id_rsa -u:hpcc -pe: -t:0 -a:3 -b
00000000 2012-04-03 14:19:00 23406 23406 "ssh result(0):
"
00000000 2012-04-03 14:19:05 23454 23454 "10.239.219.8: DONE"
00000001 2012-04-03 14:19:05 23454 23454 "Completed in 0m 0s with 0 errors"
00000002 2012-04-03 14:19:05 23454 23454 "backupnode finished"
[hpcc@oss_development:node219008
```

Aguarde até a conclusão. Você verá a mensagem "backupnode finished" (backup do nó concluído) como mostrado acima.

3. Execute o utilitário XREF no ECL Watch para ter certeza de que não há arquivos órfãos ou perdidos.

## Backup programado

A forma mais fácil de agendar um processo de backup é criar um job cron. Cron é um daemon que serve como um agendador de tarefas.

Cron tab (abreviação de CRON TABLE) é um arquivo de texto que contém a lista de tarefas. Para editar com o editor padrão, use o comando:

```
sudo crontab -e
```

Aqui está um exemplo de entrada cron tab:

```
30 23 * * * /opt/HPCCSystems/bin/start_backupnode mythor
```

30 representa os minutos de hora.

23 representa a hora do dia

O asterisco (\*) representa todos os dias, meses ou dias da semana.

mythor é o nome do cluster.

Para listar as tarefas agendadas, use o comando:

```
sudo crontab -l
```

## **Arquivos de dados Roxie**

Os dados do Roxie são protegidos por três formas de redundância:

- Retenção de arquivo de dados de fonte original: Quando uma consulta é implementada, os dados são normalmente copiados dos discos rígidos de um cluster Thor. Dessa forma, os dados do Thor podem servir como backup, contanto que não sejam removidos nem alterados no Thor. Os dados do Thor normalmente são retidos por um período suficiente para servir como uma cópia de backup.
- Redundância de nó par: Cada nó secundário normalmente tem um ou mais nós pares em seu cluster. Cada par armazena uma cópia dos arquivos de dados que serão lidos.
- Redundância de cluster irmão: Embora não seja necessário, as implementações do Roxie podem ser executadas em múltiplos clusters Roxie configurados de modo idêntico. Quando dois clusters são implementados para produção, cada nó tem um gêmeo idêntico relacionado aos dados e/ou consultas armazenadas no nó no outro cluster.

Isso fornece múltiplas cópias redundantes dos arquivos de dados.

## **Repositórios de atributos**

Os repositórios de atributo são armazenados nos discos rígidos locais do desenvolvedor do ECL. Eles podem conter um número significativo de horas de trabalho; consequentemente, o backup deve ser feito regularmente. Além disso, sugerimos usar algumas formas de controle de versão de origem.

## **Arquivos da Zona de entrada de arquivos**

A Zona de entrada de arquivos contém os dados brutos de entrada. Também pode conter arquivos de saída (ou de resultados). Dependendo do tamanho ou complexidade desses arquivos, pode ser necessário manter cópias para redundância.