

Executando o HPCC em uma Máquina Virtual

Equipe de documentação de Boca Raton



HPCC em uma Máquina Virtual

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** no campo 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® 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 usados neste manual são fictícios. Qualquer semelhança com pessoas reais, vivas ou mortas, é mera coincidência.

2021 Version 8.2.80-1

Introdução	4
Requisitos do Sistema	5
Obtendo as Ferramentas e a imagem da VM	6
VM VirtualBox	7
Executando a VM HPCC	11
Executando o ECL IDE pela primeira vez	13
Executando o HPCC ECL IDE quando você já tem uma versão previamente instalada	15
Escrever código ECL	18
Trabalhando com ECL	23
Trabalhando com arquivo de dados	38
Próximos passos	42
Perguntas Frequentes	44

Introdução

Estas instruções servem como orientação durante a instalação e execução do HPCC.¹ System em um único nó na máquina virtual Linux executada em um host Windows.

Desenvolvida para ser executada em uma máquina virtual, essa versão oferece uma experiência prática com o HPCC System. É possível experimentar e até mesmo criar aplicações analíticas de dados reais em seu PC ou notebook.

Essa versão inclui as ferramentas e a funcionalidade do HPCC sem a necessidade de um cluster físico de servidores. A versão contém todos os recursos necessários para você avaliar o HPCC system e aprender a usar o ECL². Você, obviamente, não terá acesso ao poder do processamento paralelo com esta versão, mas poderá usá-la como uma ferramenta de avaliação, aprendizagem e experimentação.



Faça a leitura completa deste documento antes de começar. As etapas descritas neste documento podem levar de uma a duas horas para serem concluídas, dependendo da velocidade do seu download.

¹High Performance Computing Cluster (HPCC) é uma plataforma de computação de processamento massivamente paralelo que soluciona problemas de big data Consulte <http://hpccsystems.com/Why-HPCC/How-it-works> para obter mais detalhes.

²Enterprise Control Language (ECL) é uma linguagem de programação declarativa e centrada em dados usada para gerenciar todos os aspectos da junção, classificação e compilação de dados massivos que realmente diferenciam o HPCC das demais tecnologias na sua capacidade de fornecer análise de dados flexíveis em escala massiva

Requisitos do Sistema

A execução do HPCC em uma máquina virtual requer (no mínimo):

- Um computador com sistema operacional Windows XP, Vista, Windows 7 (de 32 ou 64 bits) ou mais recente.
- Memória RAM de, no mínimo, 2 GB com pelo menos 1.5 GB de memória livre disponível. Recomendamos 4 GB de memória ou mais.
- Processador Intel Pentium D (ou mais potente) ou AMD Athlon64/Opteron/Phenom
- Espaço em disco disponível de, no mínimo 5 GB – recomendamos 20 GB
- Oracle VM VirtualBox[®] (versão 4.0 ou posterior).
- Internet Explorer[®] 11, Google Chrome 10, ou Firefox 3.0 (ou mais recente)

Os usuários precisam estar familiarizados com os procedimentos de instalação e execução de aplicações no Windows.

Para executar o HPCC em uma máquina virtual, a virtualização deve estar habilitada no BIOS do seu computador. Essa configuração pode se chamar VT-x ou AMD-V.

Obtendo as Ferramentas e a imagem da VM

Para executar a versão da máquina virtual do HPCC System, você precisará de um software de virtualização. Esses pacotes permitem executar imagens virtuais dentro de um único host. Há vários fornecedores de software de virtualização; embora qualquer um desses, ou todos, poderiam funcionar, nós apoiamos o VM VirtualBox® da Oracle

VM VirtualBox

O software de virtualização da Oracle, VM VirtualBox, suporta a execução da imagem da máquina virtual do HPCC.

Nas próximas seção(ões) você irá:

- Fazer o download e instalar a VM VirtualBox
- Fazer o download da imagem da máquina virtual do HPCC direto do HPCC Systems.
- Abrir e importar a imagem na VM VirtualBox

Fazer o Download e Instalar a VM VirtualBox

Se você já tiver instalado a VM VirtualBox, pule para a etapa número 6 e verifique as definições da configuração de rede.

1. Acesse o site da VirtualBox: <https://www.virtualbox.org/wiki/Downloads>.
2. Clique no link apropriado para seu sistema operacional. Por exemplo: *VirtualBox para hosts do Windows*.
3. Faça o download do arquivo de instalação da VirtualBox. Salve-o em uma pasta no seu computador.
4. Clique duas vezes no arquivo de instalação para instalar a VirtualBox.
5. Siga as instruções na tela para concluir a instalação da VirtualBox.
6. Inicie a aplicação VM VirtualBox.

Importar o arquivo de imagem virtual do HPCC

1. Baixe o arquivo mais recente da imagem da máquina virtual do HPCC em:

<http://hpccsystems.com/download/hpcc-vm-image>

Observação: Pode ser necessário se registrar para fazer o login.

Selecione o arquivo de imagem da VM para VirtualBox. O arquivo *filename.ova* é apropriado para a VirtualBox.

2. Salve o arquivo em uma pasta no seu computador.
3. Abra a VM VirtualBox.
4. No menu **File**, selecione **Import Appliance... (Ctrl + I)**
5. Pressione o botão **Open appliance...** na caixa de diálogo para selecionar o dispositivo a ser importado.
6. Navegue para a pasta onde o arquivo foi salvo e selecione-o.

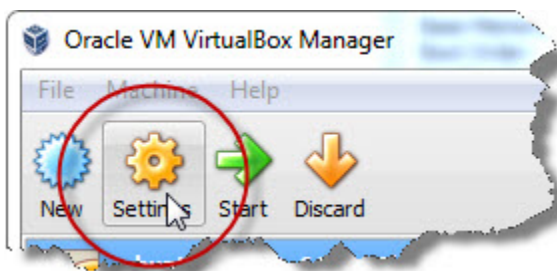
Por exemplo (**HPCCSystemsVM-n.n.n.n.ova**, onde n.n.n.n é o número da versão).

7. Pressione o botão **Open** para iniciar a importação. Siga os prompts durante o processo de importação e pressione **Next** como apropriado, e depois pressione **Import**.

Aguarde até que a máquina virtual do HPCC tenha concluído a importação. Esse procedimento pode levar alguns minutos.

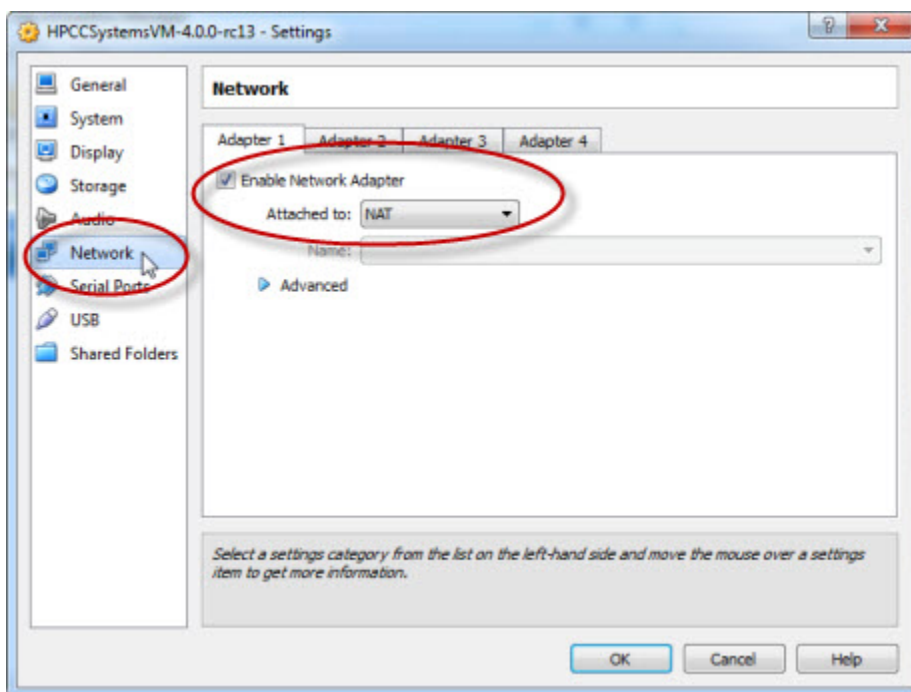
8. A máquina virtual recém-importada será listada na janela principal.
9. Selecione a nova máquina virtual.
10. Clique no ícone **Settings** para abrir a caixa de diálogo de configurações.

Figure 1. Ícone configurações do VirtualBox



11. No menu, selecione a opção **Network**. Na aba *Adapter 1*, marque a caixa *Enable Network Adapter* e defina a opção *Attached to:* para **NAT**

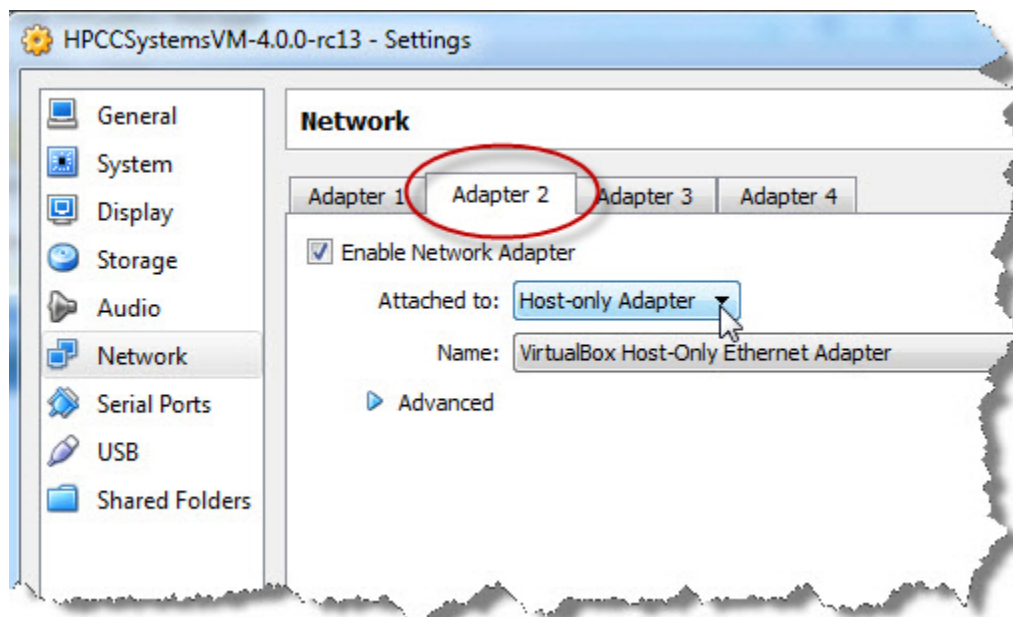
Figure 2. VMVirtualBox Network Adapter 1



Observação: Por padrão, essas configurações já podem estar definidas. Se for este o caso, basta verificar se estão corretas.

12. Na mesma janela *Network - Settings*, selecione a guia *Adapter 2*. Marque a caixa *Enable Network Adapter* e defina a opção *Attached to:* para **Host-only Adapter**.

Figure 3. VM VirtualBox Network Adapter 2



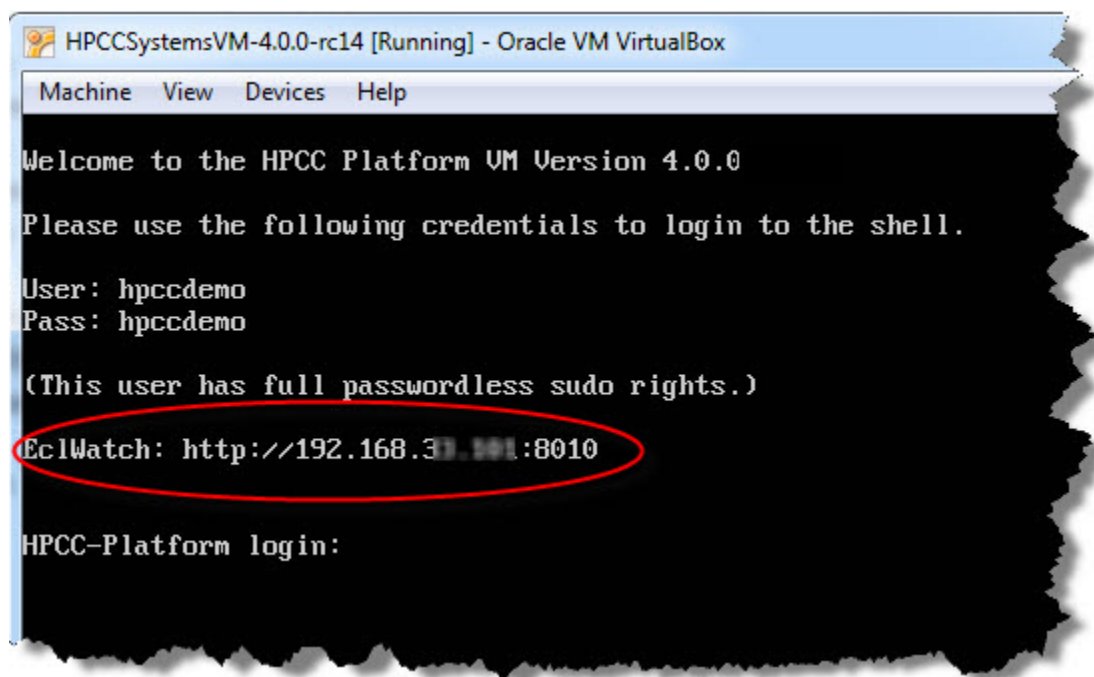
13. Pressione o botão **OK**.

14. Clique duas vezes na máquina virtual para inicializá-la.

Dica: Se aparecer qualquer mensagem de Erro de rede, revise as etapas the section called “Fazer o Download e Instalar a VM VirtualBox” e verifique se todas as configurações de rede foram definidas de maneira adequada.

15. Após a inicialização da VM ter sido concluída, aparecerá uma janela semelhante a seguinte:

Figure 4. Tela de boas-vindas da VirtualBox



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

Não redimensione essa janela, pois você não irá interagir com ela. Além disso, não há necessidade de fazer login.

Dica: Ao clicar na janela, ela captura seu teclado/mouse. Pressione a tecla **Ctrl do lado direito** para ter controle sobre seu mouse novamente.

Guest Additions

As imagens da VM do HPCC incluem o Guest Additions. Guest Additions são drivers de dispositivo e aplicações de sistema que otimizam o sistema operacional convidado (guest) para um melhor desempenho e usabilidade. Se desejar usar o ponteiro do mouse e/ou outras ferramentas da área de trabalho em sua máquina virtual, inicie a exibição do Windows X depois de efetuar o login em sua máquina virtual.

Faça o login com as credenciais fornecidas. (usuário: hpccdemo, senha: hpccdemo) No prompt de comando, digite: **startx.**

Executando a VM HPCC

Nesta seção, iremos acessar o HPCC através de uma interface Web: ECL Watch¹. Do ECL Watch, vamos fazer o download do ECL IDE de vários nós.² Caso já tenha instalado o ECL IDE, pule esta seção e continue em *Executar o ECL IDE no HPCC quando uma versão já tiver sido previamente instalada*.

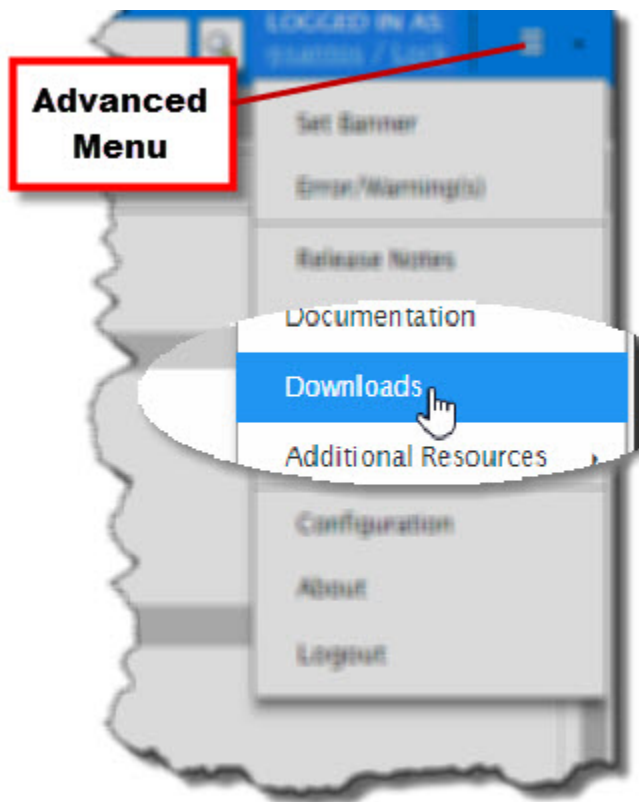
1. Em seu navegador, acesse o URL do **ECL Watch** (circulado em vermelho) na figura acima, *Tela de boas-vindas da VirtualBox*. Por exemplo, <http://nnn.nnn.nnn.nnn:8010>, onde nnn.nnn.nnn.nnn é o endereço IP da sua máquina virtual.



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

2. No menu Advanced do ECL Watch, selecione o link **Downloads**.

Figure 5. Menu Downloads



Siga o link para a página do portal de download do HPCC System.

3. Clique no link **ECL IDE** . (ao lado direito da coluna Download, abaixo do título Free Community Edition)

¹O ECL Watch é uma interface da Web para seu HPCC System. O ECL Watch permite examinar e gerenciar os vários aspectos do HPCC e permite ver informações sobre tarefas executadas, arquivos de dados e métricas de sistema.


²O ECL IDE (Ambiente de desenvolvimento integrado) é uma ferramenta usada para criar consultas em seus dados e arquivos ECL com os quais suas consultas serão compiladas. Essa é uma aplicação Windows.

4. Siga as instruções na página da Internet para instalar o ECL IDE.
5. Instale o ECL IDE, seguindo os prompts no programa de instalação. Após o ECL IDE ter sido instalado com sucesso, você pode prosseguir.

Executando o ECL IDE pela primeira vez

Nesta seção vamos configurar o ECL IDE.

1. Abra o ECL IDE, no menu iniciar. (Start >> All Programs >> HPCCSystems >> ECL IDE).

	Você pode criar um atalho em sua área de trabalho para acessar rapidamente o ECL IDE.
---	---

2. Insira o endereço IP para o servidor, exibido na tela de boas-vindas da Virtual Box , na caixa **Server** (como mostrado em Figure 6, “Preferências do ECL IDE”) e pressione o botão **OK** .

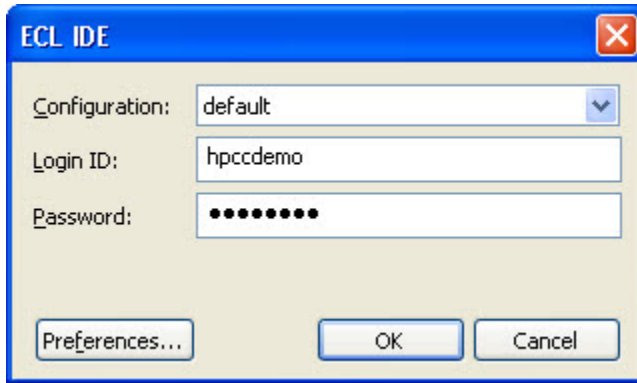
Figure 6. Preferências do ECL IDE



3. Insira o **Login ID** e a Senha (**Password**) fornecida na caixa de Login.

Login ID	hpccdemo
Senha	hpccdemo

Figure 7. Janela de Login



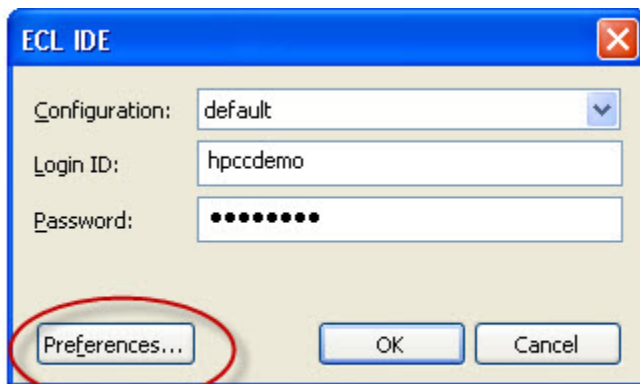
4. Pressione o botão **OK**.

Agora você está conectado e pronto para trabalhar com o HPCC!

Executando o HPCC ECL IDE quando você já tem uma versão previamente in- stalada

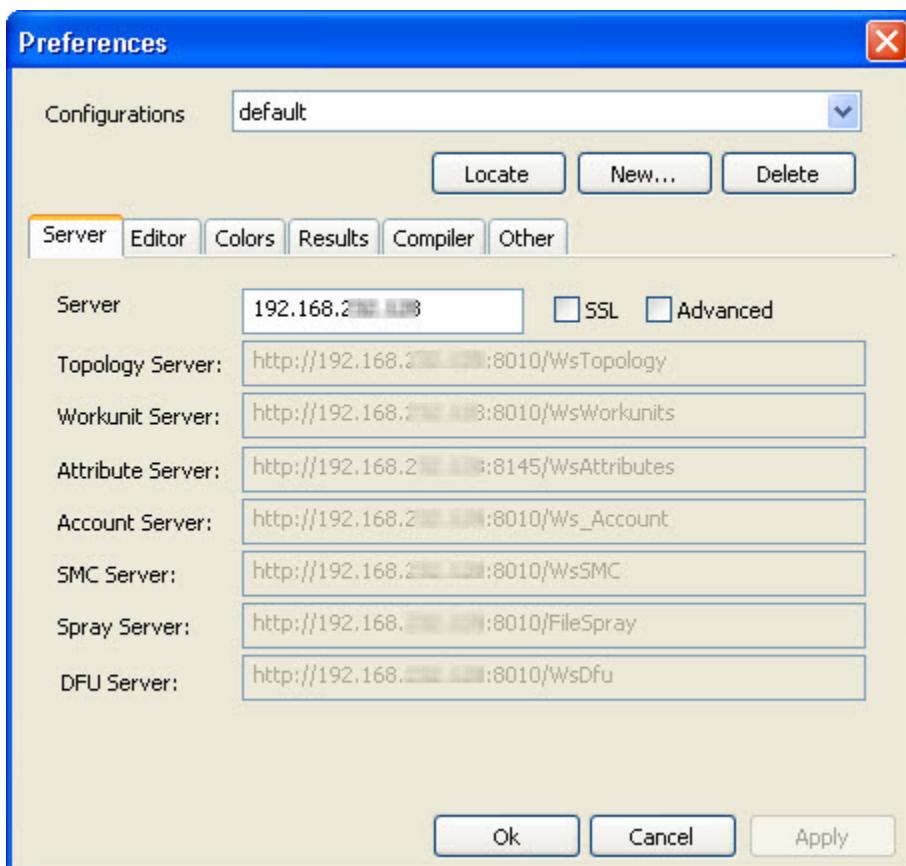
1. Abra o ECL IDE em sua estação de trabalho Windows, a partir do menu Iniciar. (Start: >> All Programs >> HPCC Systems >> ECL IDE >> ECL IDE)
2. Pressione o botão **Pressione o botão Preferences** na caixa de Login que é exibida na inicialização

Figure 8. Janela de Login



3. Insira o endereço IP exibido na tela de boas-vindas do Virtual Box para o servidor na **caixa** Server (como mostrado em Figure 9, “Preferências do ECL IDE”) e pressione o botão **OK**.

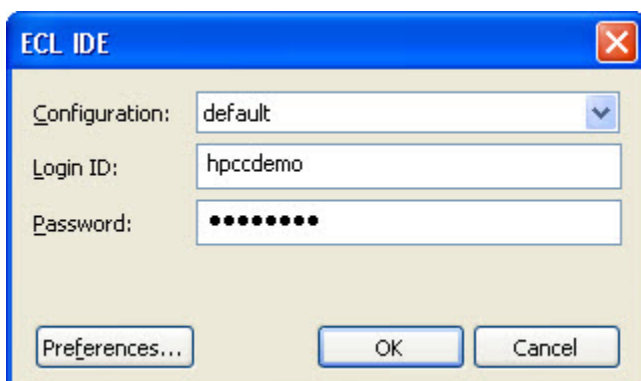
Figure 9. Preferências do ECL IDE



4. Insira a **Login Id** e a **Senha (Password)** fornecida na caixa de Login.

Login ID	hpccdemo
Senha	hpccdemo

Figure 10. Janela de Login



5. Pressione o botão **OK**.

Agora você está conectado e pronto para trabalhar com o HPCC!

Escrever código ECL

Vamos gravar, compilar e executar um programa simples denominado "Hello World" HPCC.

1. Abra o ECL IDE, no menu Iniciar. (Start >> All Programs >> HPCCSystems >> ECL IDE , ou use o atalho na área de trabalho caso tenha criado um.)

A janela de Login será exibida.

Figure 11. Janela de Login



2. Informe suas credenciais (hpccdemo) e pressione o botão OK . **OK** .
3. Abra uma nova **Janela do compilador** (CTRL+N) e escreva o seguinte código:

```
OUTPUT('Hello World');
```

Isso também poderia ser escrito como:

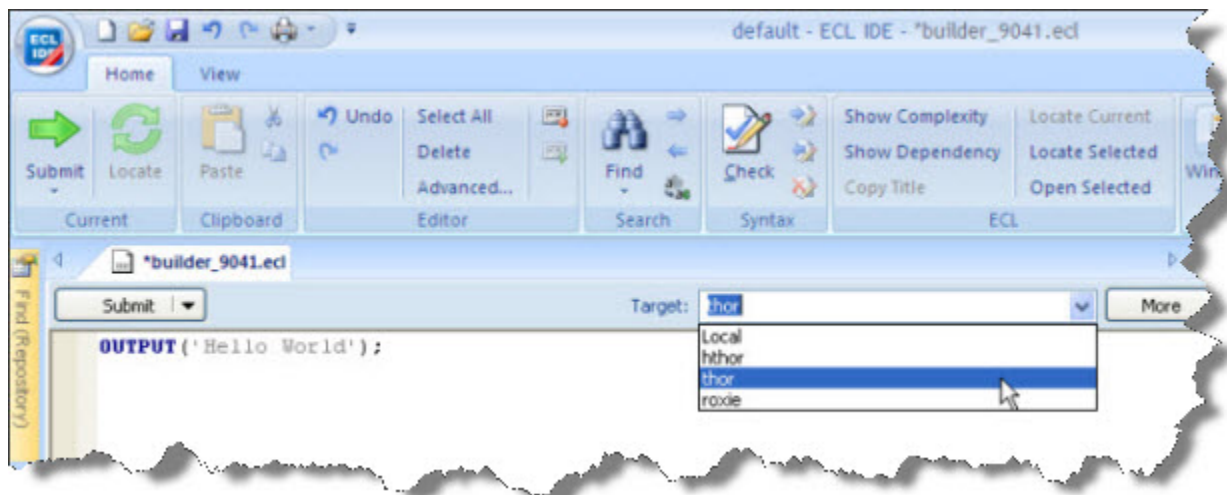
```
'Hello World';
```

Na segunda listagem de programa, a palavra-chave OUTPUT é ocultada. Isso ocorre possivelmente porque a linguagem é declarativa e a ação OUTPUT é implícita.

4. Selecione **thor** como seu cluster de destino.

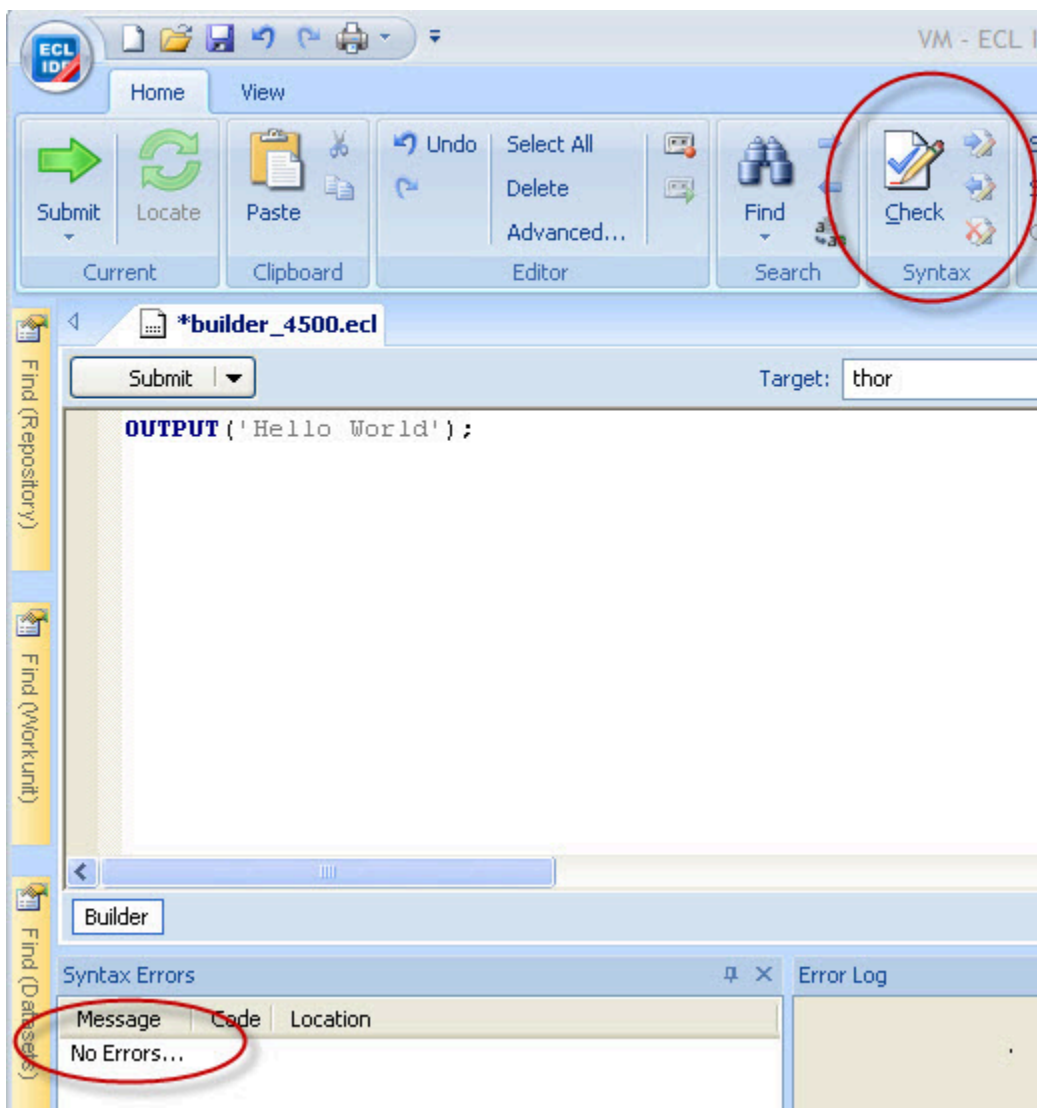
Thor é o componente da Refinaria de dados do seu HPCC. Trata-se de um cluster de computador massivamente paralelo, otimizado para classificar, manipular e transformar uma quantidade massiva de dados. Esse processo também é conhecido como ETL (Extrair, Transformar e Carregar)

Figure 12. Selecionar destino



5. Pressione o botão de verificação de sintaxe localizado na barra de ferramentas principal (ou pressione F7).

Figure 13. Verificação de sintaxe



Uma verificação de sintaxe bem-sucedida exibe a mensagem “Nenhum erro encontrado (no Errors)”.

6. Pressione o botão **Submit** (ou pressione as teclas CTRL+ENTER).

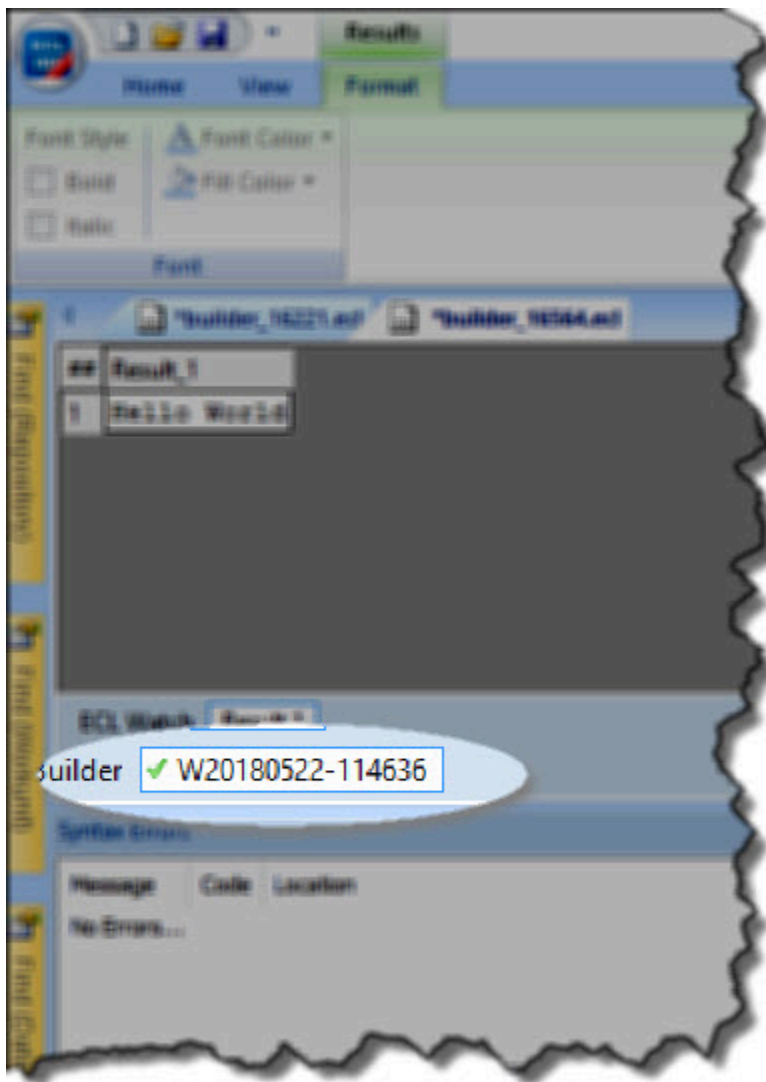
Figure 14. Job Concluído



A marcação na cor verde indica uma conclusão bem-sucedida.

7. Clique na guia do número da workunit e, em seguida, na guia Result 1 para ver os resultados.

Figure 15. Resultado do job concluído



Trabalhando com ECL

Agora que você enviou alguns códigos ECL, está na hora de tentar realizar algumas operações mais complexas.

Fornecemos os exemplos a seguir para que você possa começar.

Exemplo ECL: Anagram1

Este exemplo pega uma STRING e gera todos os anagramas possíveis a partir dela. Este código serve de base para um segundo exemplo que analisa quais destas são palavras reais usando um arquivo de dados da lista de palavras.

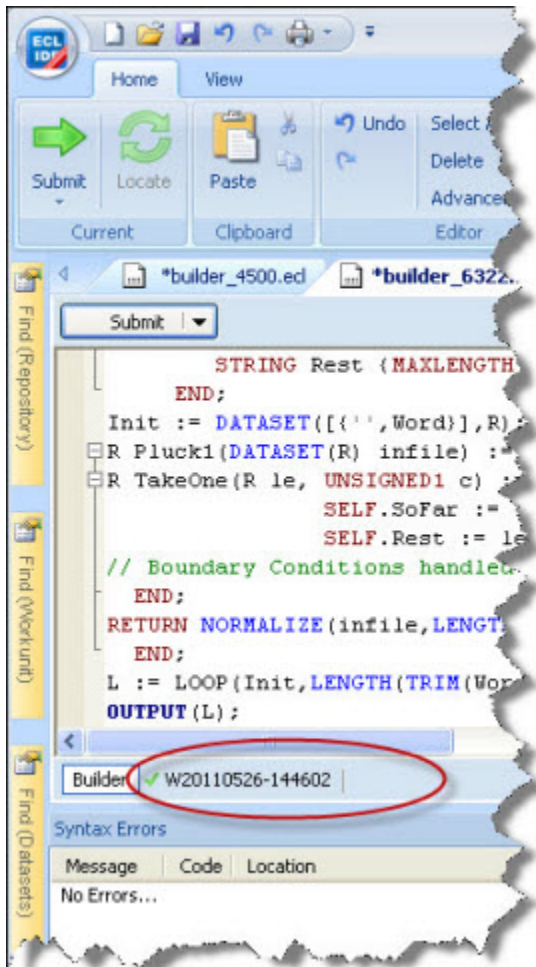
1. Abra uma nova **Janela do compilador (CTRL+N)** e escreva o seguinte código:

```
STRING Word := 'FRED' :STORED('Word');
R := RECORD
    STRING SoFar {MAXLENGTH(200)};
    STRING Rest {MAXLENGTH(200)};
END;
Init := DATASET([{'',Word}],R);
R Pluck1(DATASET(R) infile) := FUNCTION
R TakeOne(R le, UNSIGNED1 c) := TRANSFORM
    SELF.SoFar := le.SoFar + le.Rest[c];
    SELF.Rest := le.Rest[..c-1]+le.Rest[c+1..];
// Boundary Conditions handled automatically
END;
RETURN NORMALIZE(infile,LENGTH(LEFT.Rest),TakeOne(LEFT,COUNTER));
END;
L := LOOP(Init,LENGTH(TRIM(Word)),Pluck1(ROWS(LEFT)));
OUTPUT(L);
```

2. Selecione **thor** como seu cluster de destino.
3. Pressione o botão de verificação de sintaxe localizado na barra de ferramentas principal (ou pressione F7)

4. Pressione o botão **Pressione o botão Submit** (ou as teclas ctrl+enter).

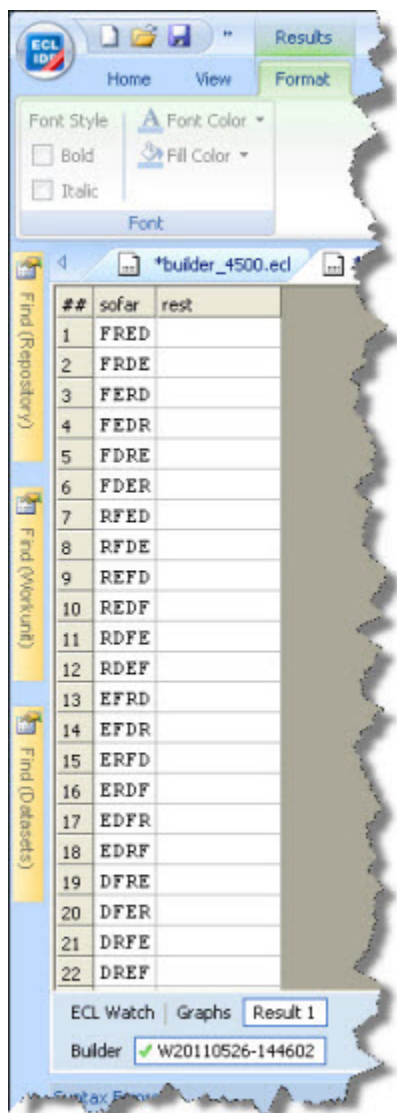
Figure 16. Job Concluído



A marcação na cor verde indica uma conclusão bem-sucedida.

5. Clique na guia do número da workunit e, em seguida, na guia Result 1 para ver os resultados.

Figure 17. Resultado do job concluído



##	sofar	rest
1	FRED	
2	FRDE	
3	FERD	
4	FEDR	
5	FDRE	
6	FDER	
7	RFED	
8	RFDE	
9	REFD	
10	REDF	
11	RDPE	
12	RDEF	
13	EFRD	
14	EFDR	
15	ERFD	
16	ERDF	
17	EDFR	
18	EDRF	
19	DFRE	
20	DFER	
21	DRFE	
22	DREF	

Exemplo Roxie: Anagram2

Neste exemplo, vamos baixar um arquivo de dados de código aberto com palavras do dicionário, fazer o spray (distribuir aos nós) desse arquivo para nosso cluster Thor e validar nossos anagramas em relação a esse arquivo para que possamos determinar quais palavras são válidas. A etapa de validação usa um JOIN da lista de anagramas para o arquivo do dicionário. O uso de um índice e de uma junção com chave seria mais eficiente, mas isso serve apenas como um simples exemplo.

Fazer o download da lista de palavras

Vamos fazer o download da lista de palavras em <http://wordlist.sourceforge.net/>

1. Faça o *download do* pacote Oficial 12 Dicts.
2. Extraia o arquivo **2of12.txt** para uma pasta em sua máquina local. Os arquivos estão disponíveis no formato tar.gz ou ZIP.

Carregar o arquivo de dicionário para sua Zona de Entrada de Arquivo

Nesta etapa, você copiará os arquivos de dados para um local onde eles possam ser distribuídos aos nós de seu cluster HPCC. Uma zona de entrada de arquivos é um local de armazenagem anexado ao seu HPCC. Ela possui um utilitário em execução para facilitar o spraying (processo de distribuir dados aos nós) para um cluster.

Para arquivos de dados menores, com tamanho máximo de 2GB, você pode usar o utilitário enviar/baixar arquivo no ECL Watch. Este arquivo de dados possui apenas 400 kb (aproximadamente).

Em seguida, você distribuirá (ou fará o spray) o dataset para todos os nós no cluster do HPCC. O poder do HPCC vem da sua capacidade de atribuir vários processadores para trabalhar nas diferentes partes do arquivo de dados em paralelo. Até mesmo na VM Edition, que possui apenas um nó único, os dados precisam passar pela operação de spray.

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

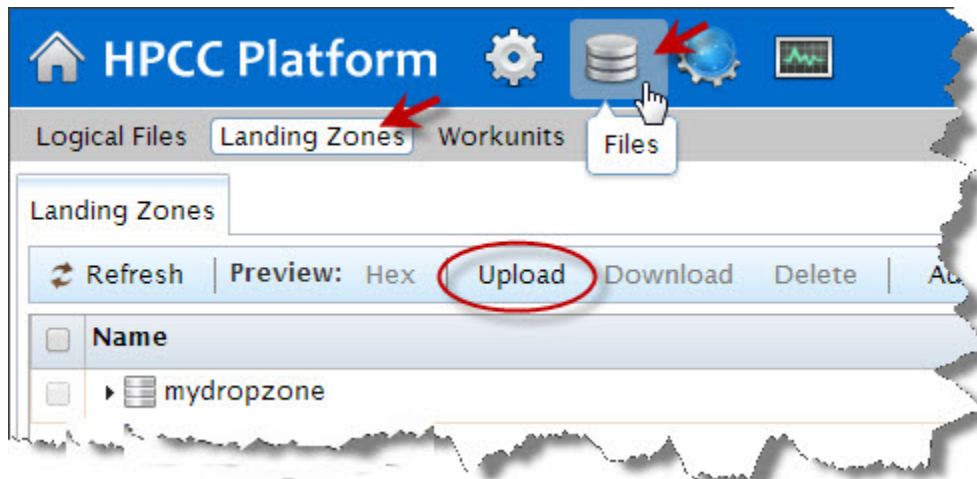


Seu endereço IP poderá ser diferente dos endereços fornecidos nas imagens de exemplo. Use o endereço IP fornecido pela **sua** instalação.

2. No ECL Watch, clique no ícone **Files** e no link **Zona de entrada de arquivos (Landing Zones)** localizados no submenu de navegação

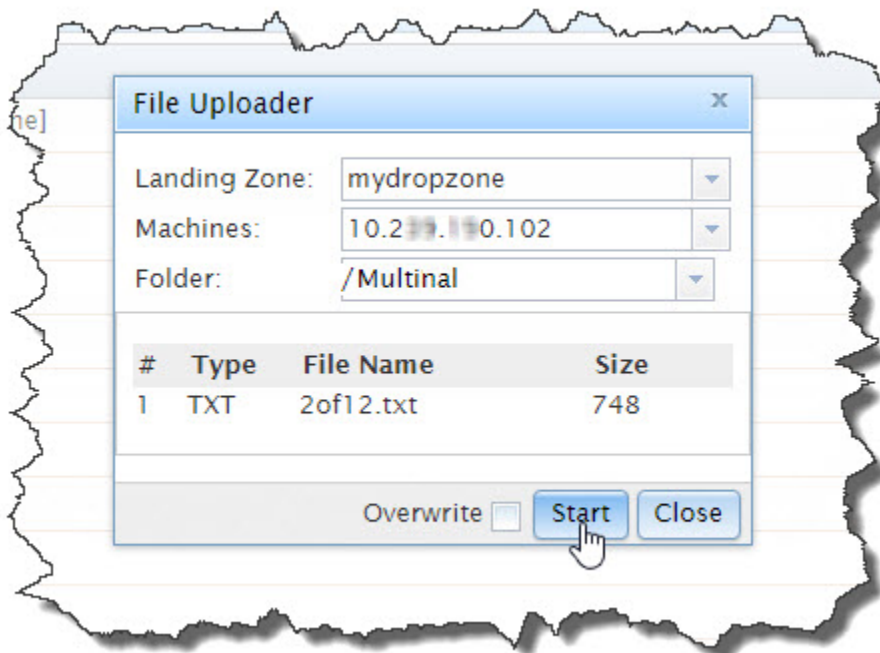
Pressione o botão de ação **Upload** .

Figure 18. Enviar



3. Uma caixa de diálogo será aberta. **Navegue** e selecione o arquivo a ser enviado e pressione o botão **Open** .

Figure 19. File Uploader



O arquivo selecionado deve aparecer no campo **File Name** . O arquivo de dados possui o seguinte nome: **2of12.txt**.

4. Pressione o botão **Start** para concluir o envio do arquivo.

Spray do arquivo de dados para o seu *Cluster da Refinaria de Dados (Thor)*

Para usar o arquivo de dados em seu HPCC System, é preciso fazer o "spray" (distribuir) desse arquivo para todos os nós. Um *spray* ou *importação* é a transferência de um arquivo de dados de um local (como a zona de entrada de arquivos) para diversas partes do arquivo ou nós em um cluster.

O arquivo distribuído recebeu um *nome-de-arquivo-lógico* como segue: **thor::word_list_csv**. O sistema mantém uma lista de arquivos lógicos e as localizações dos arquivos físicos correspondentes das partes do arquivo.

1. Abra o ECL Watch usando a URL:

http://nnn.nnn.nnn.nnn:pppp (onde **nnn.nnn.nnn.nnn** é o endereço IP do seu ESP Server e **pppp** é a porta. A porta padrão é 8010)

2. Clique no ícone **Files**, e em seguida no link **Landing Zones (Zona de entrada de arquivos)** no submenu de navegação. Selecione a zona de entrada de arquivos apropriada (caso haja mais de uma zona de entrada de arquivos). Clique na seta à esquerda da sua zona de entrada de arquivos para expandir.

3. Selecione o arquivo na zona de entrada de arquivos marcando a caixa ao lado dele.

4. Marque a caixa ao lado de 2of12.txt, e pressione o botão **Delimited**.

Figure 20. Spray delimitado

The screenshot shows the 'Spray delimited' dialog box in HPCC. The 'Target' section is expanded, showing 'Group' set to 'mythor', 'Queue' set to 'dfusever_queue', and 'Target Scope' set to '~thor'. The 'Target Name' list contains 'word_list_csv'. The 'Options' section is also expanded, showing 'Format' set to 'ASCII', 'Max Record Length' set to '8192', 'Separators' set to '\,', 'Line Terminators' set to '\n,\r\n', and 'Quote' set to ' '. The 'Overwrite' checkbox is checked, and 'Replicate' is also checked. The 'Spray' button is at the bottom right.

A página **Spray delimitado do DFU** será exibida.

5. Selecione "mythor" na lista suspensa do Target Group.
6. Preencha o Target Scope como *thor*.

7. Preencha os demais parâmetros (caso ainda não tenham sido preenchidos).

- Máximo tamanho do registro 8192
- Separador \,
- Terminador de linhas \n,\r\n
- Aspas: '

8. Preencha o Target Name usando o restante do nome do arquivo lógico desejado: word_list_csv

9. Não se esqueça de marcar a caixa **Overwrite** (Substituir).

Se disponível, certifique-se de que a caixa **Replicate** esteja marcada. (A opção replicar está disponível apenas em sistemas em que a replicação tenha sido ativada.)

10. Pressione o botão **Botão de ação Spray**.

A guia exibe a tarefa DFU onde é possível ver o progresso do spray (distribuição aos nós).

Execute a consulta no Thor

1. Abra uma nova **Janela do compilador** (CTRL+N) e escreva o seguinte código:

```
IMPORT Std;
layout_word_list := record
  string word;
end;
File_Word_List := dataset('~thor::word_list_csv', layout_word_list,
                        CSV(heading(1),separator(','),quote('')));
STRING Word := 'teacher' :STORED('Word');
STRING SortString(STRING input) := FUNCTION
  OneChar := RECORD
    STRING c;
  END;
  OneChar MakeSingle(OneChar L, unsigned pos) := TRANSFORM
    SELF.c := L.c[pos];
  END;
  Split := NORMALIZE(DATASET([input],OneChar), LENGTH(input),
    MakeSingle(LEFT,COUNTER));
  SortedSplit := SORT(Split, c);
  OneChar Recombine(OneChar L, OneChar R) := TRANSFORM
    SELF.c := L.c+R.c;
  END;
  Recombined := ROLLUP(SortedSplit, Recombine(LEFT, RIGHT),ALL);
  RETURN Recombined[1].c;
END;

STRING CleanedWord := SortString(TRIM(Std.Str.ToUpperCase(Word)));

R := RECORD
  STRING SoFar {MAXLENGTH(200)};
  STRING Rest {MAXLENGTH(200)};
END;
Init := DATASET([{'',CleanedWord}],R);
R Pluck1(DATASET(R) infile) := FUNCTION
  R TakeOne(R le, UNSIGNED c) := TRANSFORM
    SELF.SoFar := le.SoFar + le.Rest[c];
    SELF.Rest := le.Rest[..c-1]+le.Rest[c+1..];
    // Boundary Conditions
    // handled automatically
```

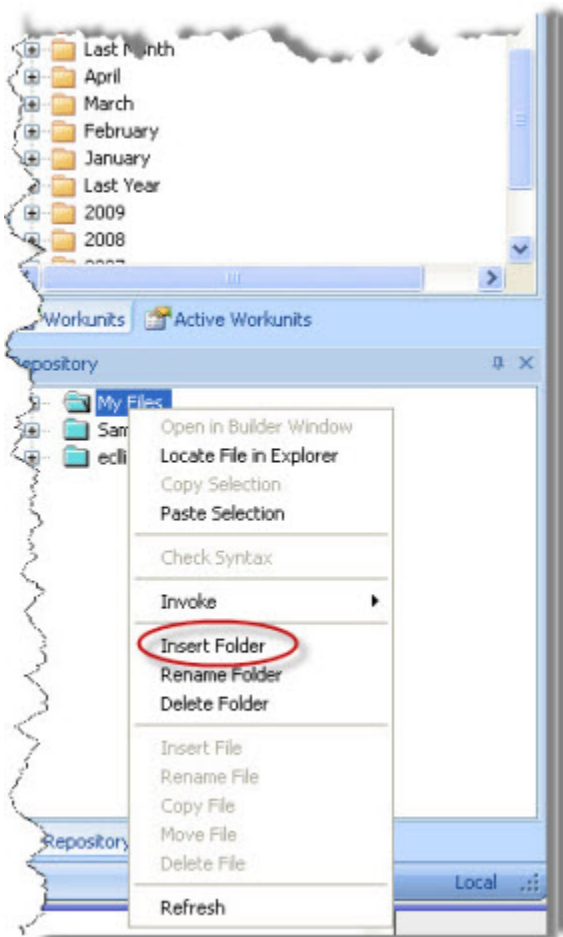
```
END;  
RETURN DEDUP(NORMALIZE(infile,LENGTH(LEFT.Rest),TakeOne(LEFT,COUNTER)) );  
END;  
L := LOOP(Init,LENGTH(CleanedWord),Pluck1(ROWS(LEFT)) );  
ValidWords := JOIN(L,File_Word_List,  
LEFT.SoFar=Std.Str.ToUpperCase(RIGHT.Word),TRANSFORM(LEFT));  
OUTPUT(CleanedWord);  
COUNT(ValidWords);  
OUTPUT(ValidWords)
```

2. Selecione **thor** como seu cluster de destino.
3. Pressione o botão de verificação de sintaxe localizado na barra de ferramentas principal (ou pressione F7)
4. Pressione o botão **Submit** .
5. Quando o envio estiver concluído, selecione a aba Workunit e em seguida a guia Results.
6. Examine o resultado.

Compilar e Publicar a consulta para o Roxie

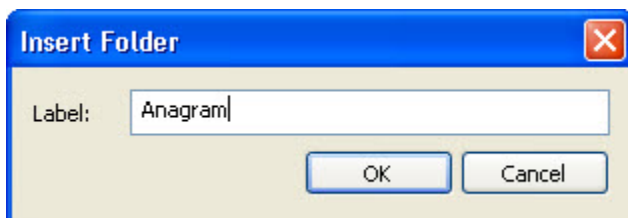
1. Clique com o botão direito na pasta **MyFiles** na janela Repositório , e selecione a opção **Insert Folder** no menu pop-up.

Figure 21. Inserir Pasta



2. Insira **Anagram** para o rótulo e depois pressione o botão OK.

Figure 22. Inserir o título da pasta

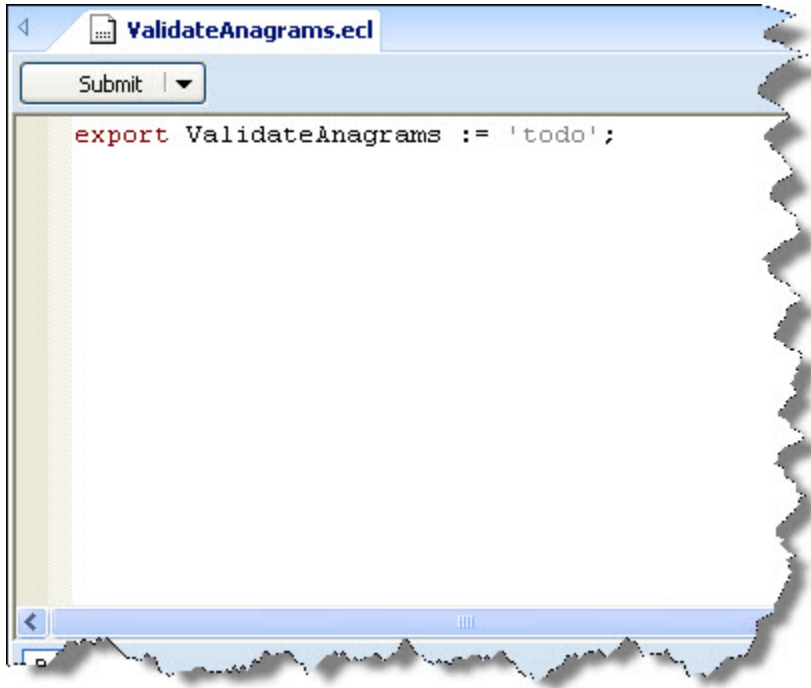


3. Clique com o botão direito na pasta **Anagram** e selecione **Insert File** no menu pop-up.

4. Insira **ValidateAnagrams** para o rótulo e depois pressione o botão OK.

Uma janela do compilador será aberta.

Figure 23. Janela do compilador



5. Escreva o seguinte código (ele pode ser copiado de outra janela do compilador):

```
IMPORT Std;
layout_word_list := record
  string word;
end;
File_Word_List := dataset('~thor::word_list_csv', layout_word_list,
                        CSV(heading(1),separator(','),quote('')));
STRING Word := 'teacher' :STORED('Word');
STRING SortString(STRING input) := FUNCTION
  OneChar := RECORD
    STRING c;
  END;
  OneChar MakeSingle(OneChar L, unsigned pos) := TRANSFORM
    SELF.c := L.c[pos];
  END;
  Split := NORMALIZE(DATASET([input],OneChar), LENGTH(input),
    MakeSingle(LEFT,COUNTER));
  SortedSplit := SORT(Split, c);
  OneChar Recombine(OneChar L, OneChar R) := TRANSFORM
    SELF.c := L.c+R.c;
  END;
  Recombined := ROLLUP(SortedSplit, Recombine(LEFT, RIGHT),ALL);
  RETURN Recombined[1].c;
END;

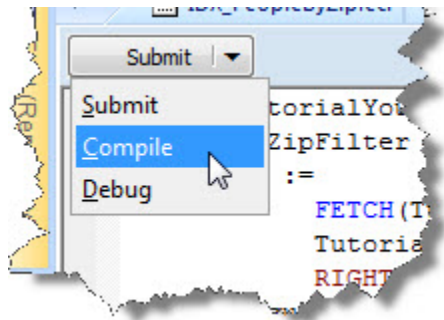
STRING CleanedWord := SortString(TRIM(Std.Str.ToUpperCase(Word)));

R := RECORD
  STRING SoFar {MAXLENGTH(200)};
```

```
STRING Rest {MAXLENGTH(200)};  
END;  
Init := DATASET(['',CleanedWord],R);  
R Pluck1(DATASET(R) infile) := FUNCTION  
  R TakeOne(R le, UNSIGNED1 c) := TRANSFORM  
    SELF.Sofar := le.Sofar + le.Rest[c];  
    SELF.Rest := le.Rest[..c-1]+le.Rest[c+1..];  
    // Boundary Conditions  
    // handled automatically  
  END;  
  RETURN DEDUP(NORMALIZE(infile,LENGTH(LEFT.Rest),TakeOne(LEFT,COUNTER)));  
END;  
L := LOOP(Init,LENGTH(CleanedWord),Pluck1(ROWS(LEFT)));  
ValidWords := JOIN(L,File_Word_List,  
LEFT.Sofar=Std.Str.ToUpperCase(RIGHT.Word),TRANSFORM(LEFT));  
OUTPUT(CleanedWord);  
COUNT(ValidWords);  
OUTPUT(ValidWords)
```

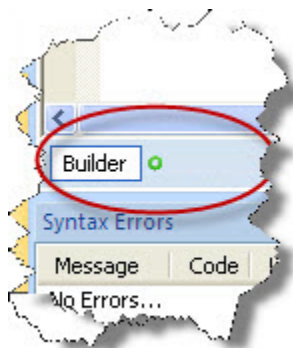
6. Selecione **Roxie** como seu cluster de destino.
7. Pressione o botão de verificação de sintaxe localizado na barra de ferramentas principal (ou pressione F7)
8. Na janela do compilador, no canto superior esquerdo, há uma seta suspensa ao lado do botão **Submit** . Selecione a seta para exibir a opção **Compile (Compilar)** .

Figure 24. Compilar



9. Selecione **Compile (Compilar)**
10. Quando o envio estiver concluído, selecione a guia Workunit e em seguida a guia Results.
11. Quando a tarefa for concluída, ela exibirá um círculo verde indicando que foi compilada.

Figure 25. Compilada

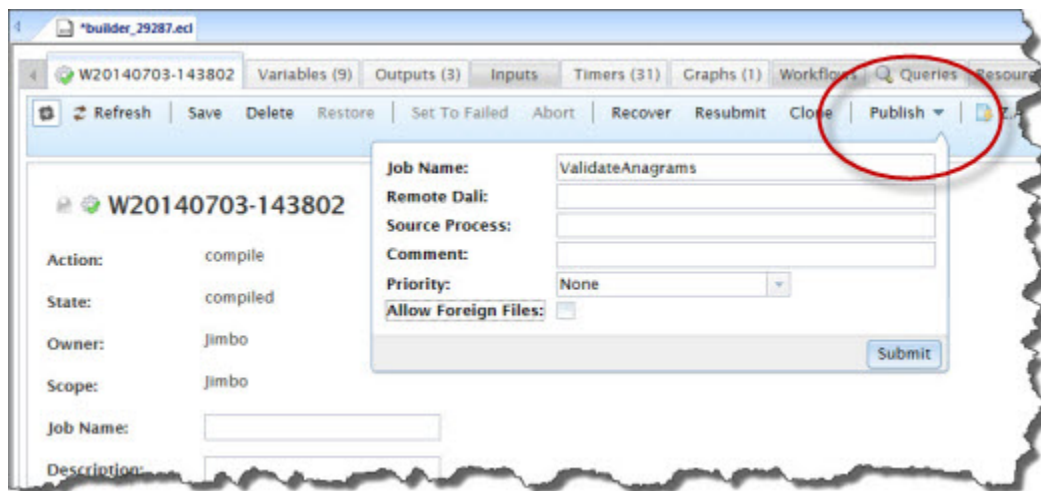


Publicar a consulta Roxie

Agora vamos publicar a consulta em um cluster Roxie.

1. Selecione a guia workunit para a ValidateAnagrams que você acabou de compilar.
2. Selecionar a aba ECLWatch.
3. Pressione o botão **Publish**, preencha a caixa de diálogo e pressione **Submit**.

Figure 26. Publicar Consulta



Uma mensagem de confirmação será exibida quando a consulta for publicada com sucesso.

Executar a consulta Roxie no WsECL

Agora que a consulta foi implementada a um cluster Roxie, podemos executá-la usando o serviço WsECL no URL a seguir:

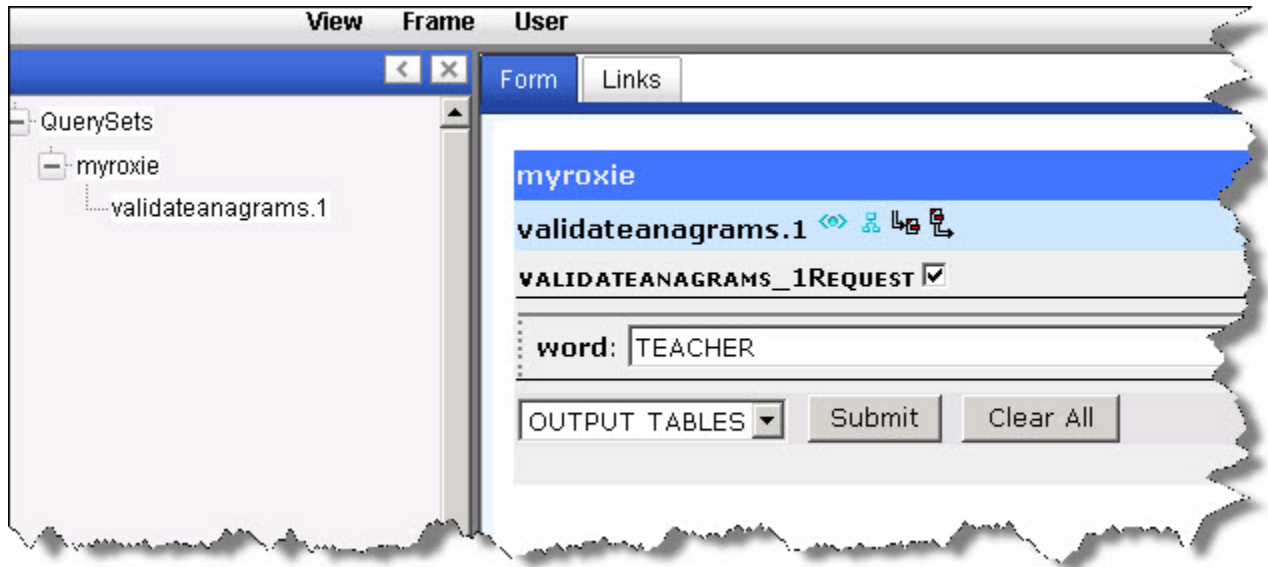
http://nnn.nnn.nnn.nnn:pppp (onde nnn.nnn.nnn.nnn é o endereço IP do seu ESP Server e pppp é a porta. A porta padrão é 8002)

1. Clique no sinal + ao lado de **myroxie** para expandir a árvore.

2. Clique no hyperlink **ValidateAnagrams.1** .

O formulário do serviço será exibido.

Figure 27. RoxieECL

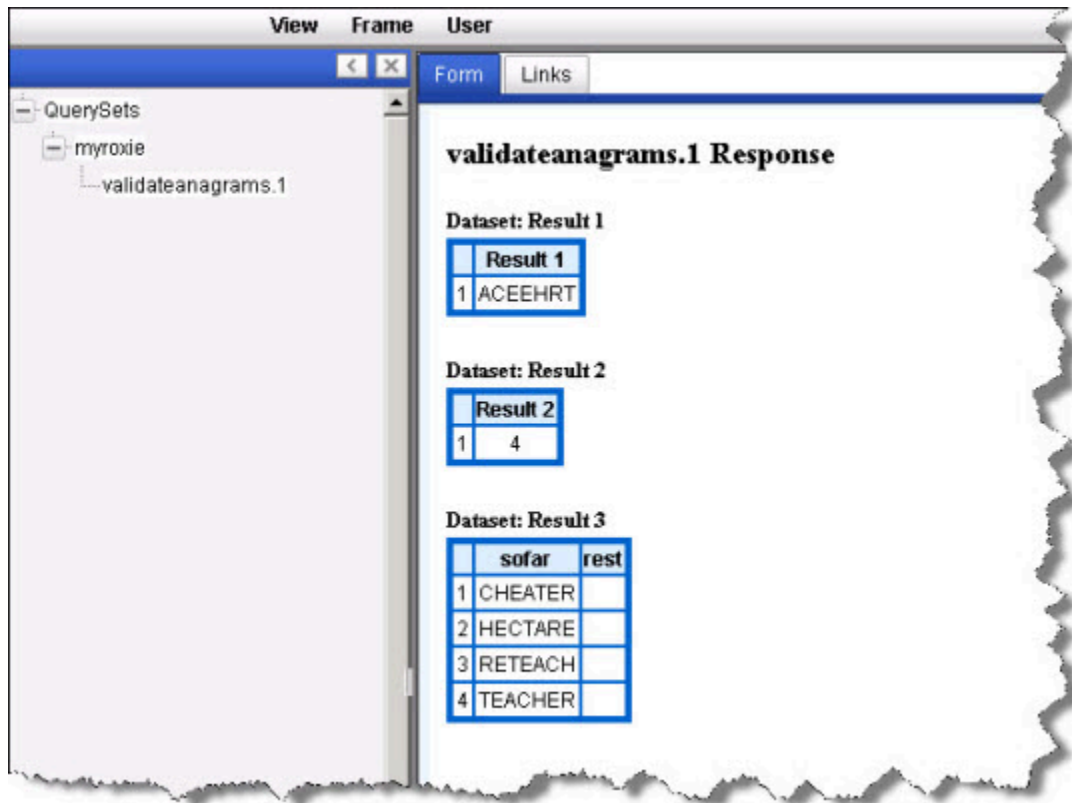


3. Selecione Tabelas de resultado na lista suspensa.

4. Forneça uma palavra para que seja feito o anagrama (p.ex., TEACHER) e pressione o botão Submit.

Os resultados serão exibidos.

Figure 28. RoxieResults



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 iniciar

Um típico HPCC System de produção teria uma capacidade maior de dados em comparação a um sistema virtual usado para realização de testes. O tamanho dos arquivos com os quais você deseja trabalhar é limitado pelo tamanho de sua máquina virtual.

- A máquina virtual possui um limite de 20GB.
- O tamanho do(s) arquivo(s) com os quais você deseja trabalhar é limitado pelo tamanho de sua máquina virtual. Certifique-se de ter espaço disponível adequado.



Se os limites de tamanho forem excedidos, o estado da sua VM não será salvo e você correrá o risco de perder seu trabalho. Lembre-se de reservar bastante espaço, não apenas para armazenar seus dados, mas também para que o “estado” da sua máquina virtual possa salvar seu trabalho.

Enviando um arquivo

Para arquivos de dados menores, com tamanho máximo de 2GB, você pode usar o utilitário enviar/baixar arquivo no ECL Watch.

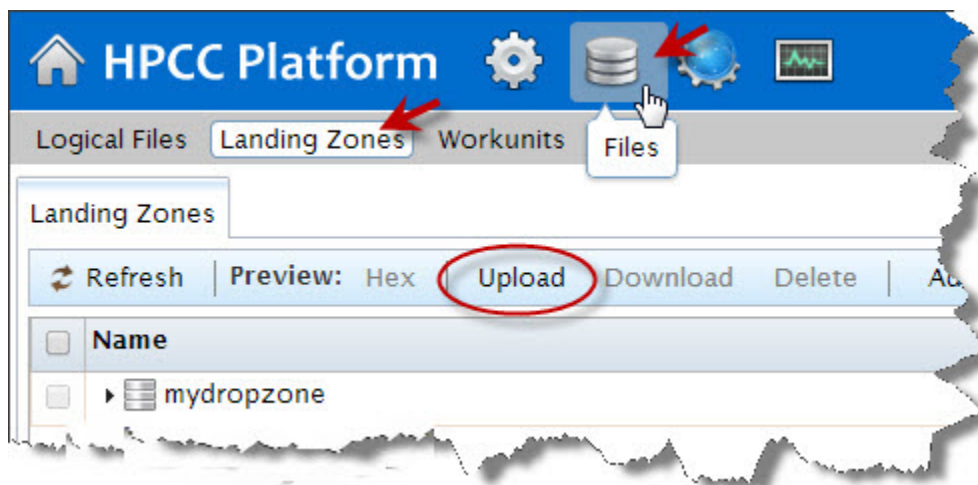
1. Em seu navegador, acesse a URL do **ECLWatch** (circulado em vermelho) na figura acima, tela de boas-vindas da Virtual Box. Por exemplo, <http://nnn.nnn.nnn.nnn:8010>, onde nnn.nnn.nnn.nnn é o endereço IP da sua máquina virtual.



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

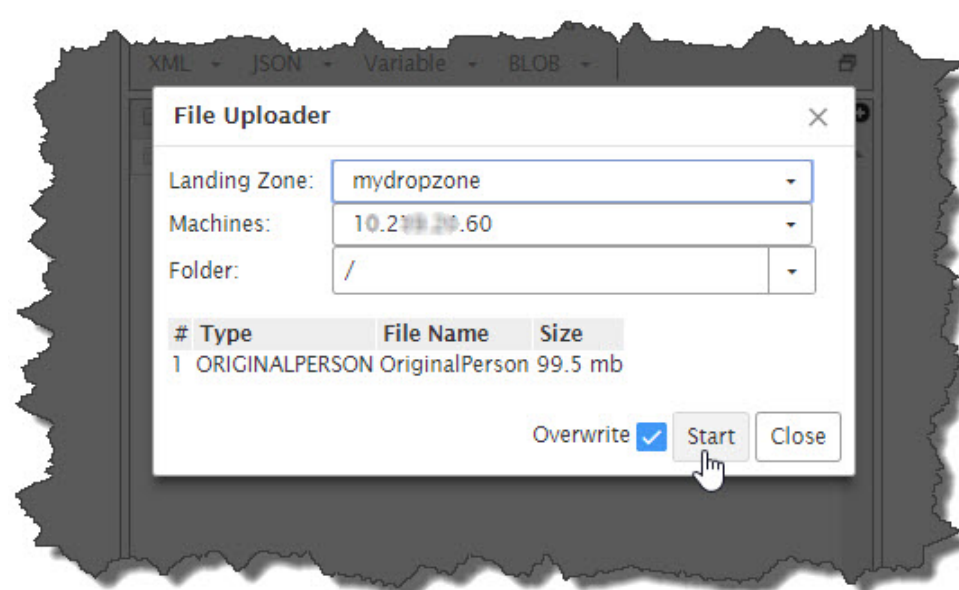
2. Na página do ECL Watch, clique no link **Upload/Download File** localizado no menu ao lado esquerdo.

Figure 29. Upload/download



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

Figure 30. Dropzones



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

O arquivo selecionado deve aparecer no campo **Select a file to upload** field.

4. Pressione o botão **Upload Now** para concluir o envio do arquivo.
5. Agora que o arquivo está em sua zona de entrada de arquivos, é possível fazer o spray do arquivo para seu cluster e gravar o código ECL para processá-lo.

Enviando arquivos com uma Cópia de Segurança.

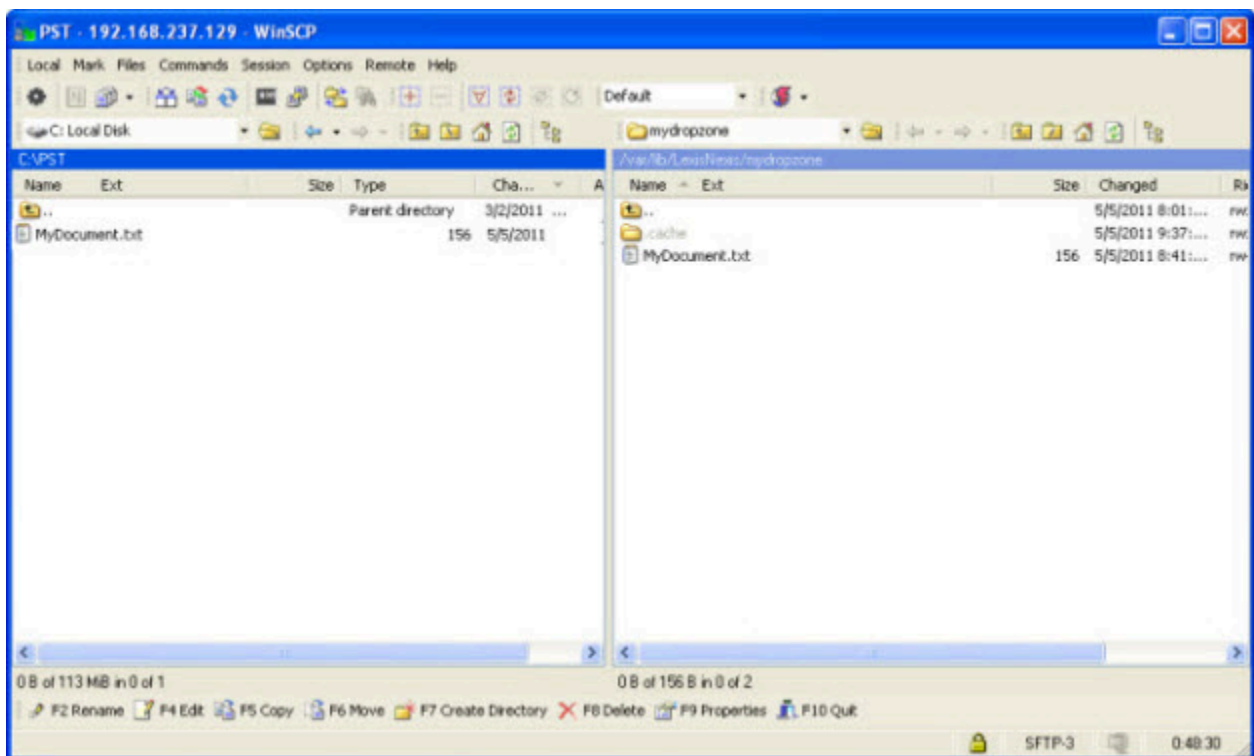
Para enviar um arquivo maior para processamento para sua máquina virtual, será necessária uma ferramenta com suporte para o protocolo de cópia de segurança. Nesta seção, abordaremos o uso do WinSCP. Há outras ferramentas disponíveis, mas os passos são similares.

1. Abra a ferramenta WinSCP e faça login no endereço IP da sua Máquina virtual usando o nome de usuário e senha fornecidos.

Login ID:	hpccdemo
Password:	hpccdemo

2. Depois de fazer login, ele deve navegar automaticamente até a pasta da zona de entrada de arquivo. (/var/lib/HPCCSystems/mydropzone)
3. Navegue onde seu arquivo local está na parte esquerda da janela.

Figure 31. 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.
5. Agora que o arquivo está em sua zona de entrada de arquivos, é possível fazer o spray do arquivo para seu cluster e gravar o código ECL para processá-lo.

Próximos passos

O menu no ECL Watch disponibiliza vários documentos que fornecem detalhes sobre os vários aspectos do HPCC.

Figure 32. Página Recursos do ECL Watch

The screenshot shows the ECL Watch HPCC Systems interface. On the left is a navigation menu with categories like Topology, DFU Workunits, DFU Files, Relationships, Roxie Queries, and Resources. The 'Resources' category is expanded, and the 'Browse' link is circled in red. The main content area has the title 'EclWatch' and 'Enterprise Serv Platform'. It contains a message about downloading a version from the installation, a link to <http://hpccsystems.com/download>, and three tables of resources: clienttools, documentation, and graphcontrol.

clienttools:

Name	Version	Description
ECL IDE Installer	ver-5_8_3_6_682_0	Windows installer for ECL IDE

documentation:

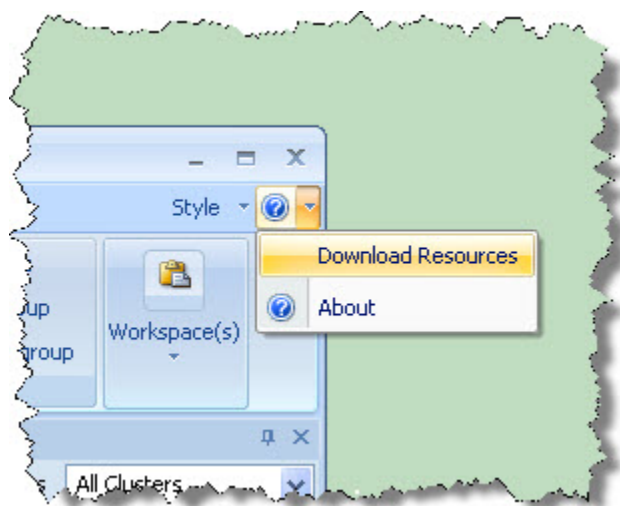
Name	Version	Description
HPCC Data Tutorial	B.4	HPCC Data Tutorial
ECL Language Reference	B.4	ECL Language Reference
HPCC Getting Started	B.4	HPCC Getting Started
ECL Programmers Guide	B.4.003	ECL Programmers Guide
IMDB Kevin Bacon Example	B.4	IMDB Kevin Bacon Example
HPCC Client Tools	B.4	HPCC Client Tools
Running HPCC in a Virtual Machine	B.4.06	Running HPCC in a Virtual Machine
HPCC Data Handling	B.4	HPCC Data Handling

graphcontrol:

Name	Version	Description
Graph Control Installer	20110523	Windows installer for Graph Control

Você pode acessar estes documentos a partir do menu Ajuda: Help >> Download Resources.

Figure 33. Menu Help



Para se familiarizar com o que o seu sistema é capaz de fazer, recomendamos:

- **O Tutorial de dados do HPCC**

Este é um tutorial simples, passo a passo, que mostra o processo completo, desde o recebimento de um arquivo de dados brutos até a publicação de uma consulta baseada na Web para realizar a pesquisa de dados. Durante o tutorial, você aprenderá como processar e indexar os dados, e depois como gravar e publicar uma consulta para pesquisar dados. O tutorial auto-conduzido e os dados que o acompanham estão disponíveis na página de Recursos do ECL Watch.

- **O Exemplo da teoria dos seis graus de separação (de Kevin bacon).**

Este é um exemplo mais complexo (*também disponível na página Recursos do ECL Watch) que usa uma base de dados para mover os dados a fim de localizar o grau de separação entre os atores em filmes.

- O portal do HPCC Systems® (<http://hpccsystems.com>) também é um recurso valioso para obter mais informações, incluindo:

- Vídeos tutoriais
- Exemplos adicionais
- Informe técnico
- Documentação
- Fóruns de suporte

Perguntas Frequentes

1. Posso executar a VM enquanto estiver conectado à rede através de uma Rede Privada Virtual (VPN)?

Não. A maioria das Redes Privadas Virtuais (VPN) controlam o seu dispositivo de rede e roteamento, não permitindo a divisão de túneis
2. Posso instalar essa máquina virtual em vários nós?

Não. Se desejar avaliar um sistema de vários nós, utilize a versão Community disponível no portal do HPCC Systems® : <http://hpccsystems.com>.
3. Quais são os limites desta versão?

O HPCC VM Edition é executado em um nó único, tem um limite de 20 GB em seu espaço de trabalho e não suporta configurações customizadas.
4. Posso executar a VM em meu computador Mac?

Você pode executar a VM do HPCC usando o Reprodutor de VM do Linux. O ECL IDE do HPCC é uma aplicação do Windows, porém pode ser executada no WINE. Consulte o manual Ferramentas de cliente para obter mais detalhes
5. Posso executar a VM em meu computador Linux?

Por enquanto não. Por enquanto não. Ainda não há uma versão do reprodutor da VM para Mac.
6. Posso executar a VM em meu Windows Server?

Sim, ela pode ser executada no Windows Server 2003 ou 2008, desde que você tenha acesso a ela através do Remote Desktop Protocol (RDP).
7. Eu preciso de um processador de 64 bits para executar a VM?

Não. O player da VM pode ser executado em ambientes de 32 ou de 64 bits e não exige um processador de 64 bits.
8. Eu preciso de um processador de 64 bits para executar o Community Edition?

Sim. Os binários do Community Edition são executados nativamente em servidores Linux de 64 bits. O HPCC pode ser acessado de qualquer estação de trabalho Windows (32 ou 64 bits) que seja capaz de executar o ECLIDE e um navegador compatível.
9. O que acontece com meu trabalho ao encerrar a máquina virtual?

O reprodutor da VM salva o status do seu sistema ao ser encerrado. Todas as informações da workunit, arquivos de dados e grupo de consultas publicadas são salvos, desde que haja espaço disponível. O tamanho limite de uma sessão salva é de 20 GB.GB.
10. Por que a minha VM não permite que eu acesse a minha interface de rede?

Verifique suas configurações de firewall. Pode ser necessário desabilitar o firewall da interface de rede da sua VM.
11. Esta versão usará meu processador de vários núcleos?

Esta VM foi projetada para usar um único núcleo.

- 12.** A VM não inicializa no Mac OSx após a atualização para o OSx 10.9 Mavericks. Como posso corrigir esse problema?

Para corrigir o problema, execute o seguinte comando:

```
sudo launchctl load /Library/LaunchDaemons/org.virtualbox.startup.plist
```

- 13.** Onde posso buscar mais informações?

Visite o portal da Web do HPCC Systems® no portal, através do endereço <http://HPCCsystems.com>.