

Aplicações MapReduce – Manipulando dados no HDFS

PROFESSORA

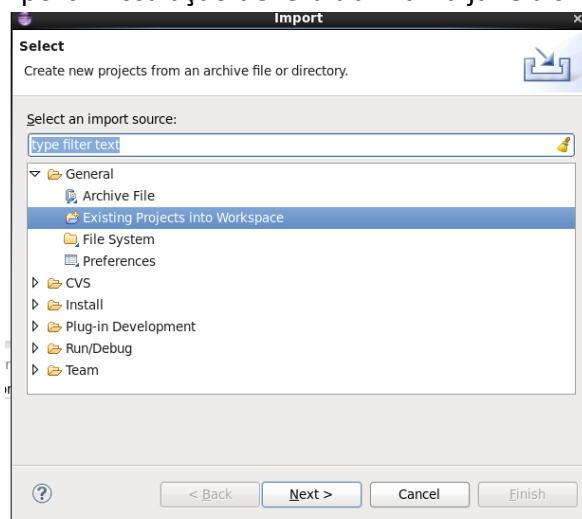
Rosângela de Fátima Pereira – rpereira@larc.usp.br

DESCRIÇÃO

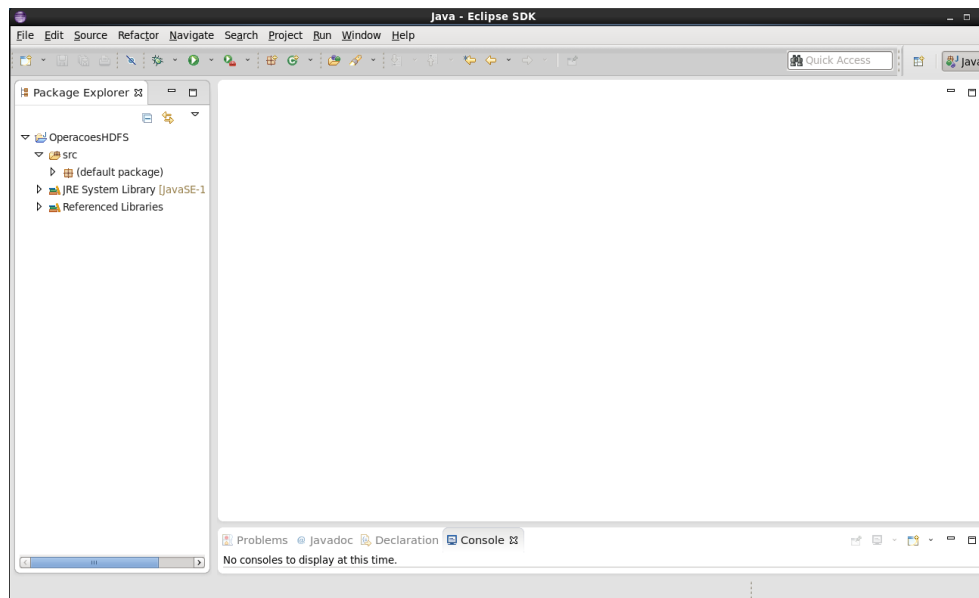
Esse documento descreve os passos necessários para a implementação e execução de uma aplicação que permite a manipulação de dados no HDFS utilizando a biblioteca Java do Hadoop.

ATIVIDADES

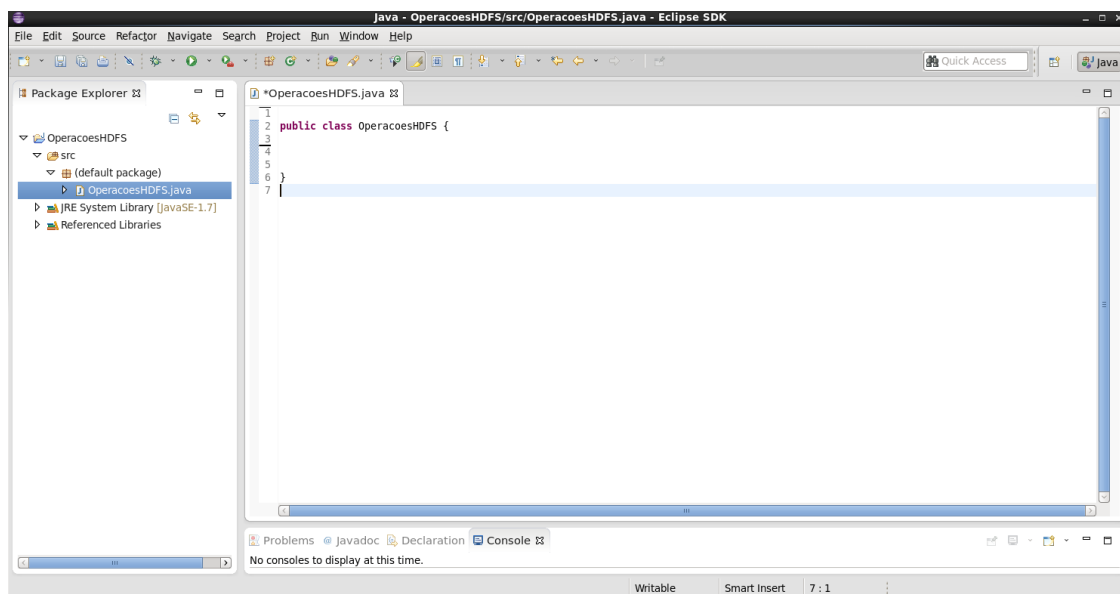
1. Abra a IDE Eclipse
2. Selecione a opção “File>>Import”. Essa ação deverá abrir uma janela similar à Figura a seguir:



3. Selecione a opção “Existing Projects into Workspace” e pressione o botão Next.
4. No campo “Select root directory” indique o seguinte caminho:
/home/training/workspace/OperacoesHDFS
Feito isso, pressione o botão Finish. O projeto deverá aparecer conforme imagem a seguir:



5. Agora deverá ser criada a classe OperacoesHDFS. Para isso, clique com o botão direito do mouse sobre o nome do projeto e selecione a opção: New >> Class.
6. No campo "Name", indique a classe OperacoesHDFS. Após isso, pressione o botão Finish. Deverá aparecer a seguinte imagem:



7. Na classe OperacoesHDFS, inserir o código a seguir:

```
import java.io.IOException;
```

```
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FSDataInputStream;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
```

```
public class OperacoesHDFS {
```

```

public static void main(String[] args) throws IOException{

    Configuration config = new Configuration();

    config.set("fs.defaultFS","hdfs://localhost:8020/");

    FileSystem dfs = FileSystem.get(config);

    System.out.println("Criando diretorio entrada");
    String nomeDir = "entrada";
    Path caminhoDir = new Path(dfs.getWorkingDirectory()+"/"+nomeDir);
    dfs.mkdirs(caminhoDir);

    System.out.println("Criando sub-diretorio amazon");
    String nomeSubDir = "amazon";
    Path caminhoSubDir = new Path(dfs.getWorkingDirectory()+"/"+nomeDir+"/"+ nomeSubDir);
    dfs.mkdirs(caminhoSubDir);

    System.out.println("Enviando arquivo da base local para o HDFS");
    Path caminhoLocal = new Path("/home/training/bases/amazon-meta.txt");
    Path destino = new Path(dfs.getWorkingDirectory()+"/"+nomeDir+"/"+nomeSubDir+"/");
    dfs.copyFromLocalFile(caminhoLocal, destino);

    System.out.println("Verificando o fator de replicaçao");
    System.out.println(dfs.getDefaultReplication(destino));

    System.out.println("Verificando o tamanho do bloco");
    System.out.println(dfs.getDefaultBlockSize(destino));

    dfs.close();
}
}

```

8. Repita o mesmo procedimento da Etapa 5, para criar a classe LeituraHDFS, inserindo o código a seguir:

```

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FSDataInputStream;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;

public class LeituraHDFS {

    public static void main(String[] args) throws IOException{

        Configuration config = new Configuration();

        config.set("fs.defaultFS","hdfs://localhost:8020/");

        FileSystem dfs = FileSystem.get(config);

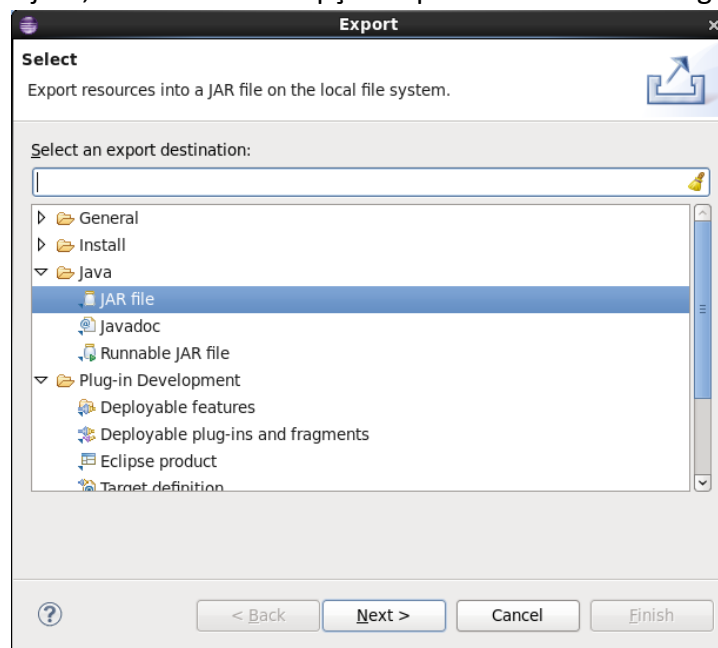
```

```

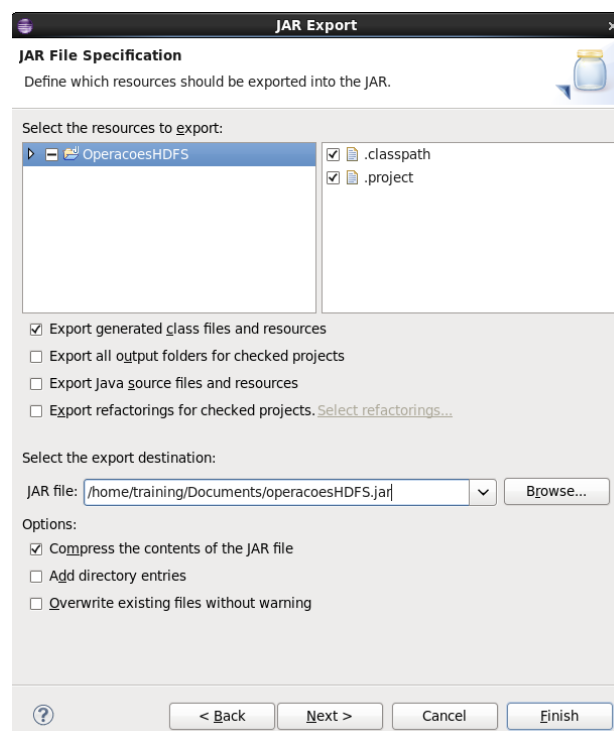
        System.out.println("Visualizando o conteudo do arquivo");
        Path arquivo = new Path(dfs.getWorkingDirectory()+"/entrada/amazon/amazon-meta.txt");
        FSDataInputStream inStream = dfs.open(arquivo);
        String data = inStream.readUTF();
        System.out.println(data);
        inStream.close();
    }
}

```

9. Tendo implementado todas as classes, gere um arquivo JAR clicando com o botão direito do mouse no nome do projeto, selecionando a opção export. Será aberto a seguinte janela:



10. Selecione a opção JAR file e pressione o botão “Next”. Deverá aparecer a seguinte janela:



11. No campo JAR file, indique o seguinte caminho: /home/training/Documents/operacoesHDFS.jar e pressione o botão Finish. Essa ação criará o jar do job implementado.

12. Para executar a classe OperacoesHDFS do job, abra um terminal e execute o seguinte comando:

```
$ hadoop jar ~/Documents/operacoesHDFS.jar OperacoesHDFS
```

13. Após o job ter sido executado, verifique se de fato o arquivo amazon-meta.txt foi de fato enviado ao HDFS por meio do seguinte comando:

```
$ hadoop fs -ls entrada/amazon
```

14. Deverá aparecer um resultado similar ao exemplo a seguir:

```
Found 1 items
-rw-rw-rw- 1 training supergroup 7547104 2016-06-26 20:15 entrada/amazon/amazon-meta.txt
```

15. Após o arquivo foi transferido corretamente, execute a classe LeituraHDFS do job, por meio do seguinte comando:

```
$ hadoop jar ~/Documents/operacoesHDFS.jar LeituraHDFS
```

16. Essa ação deverá aparecer o conteúdo do arquivo similar ao exemplo a seguir:

```
...
Id: 547021
ASIN: B0000694WF
title: Maggie and the Ferocious Beast - Adventures in Nowhere Land
group: DVD
salesrank: 7552
similar: 5 B00007KK53 B0000A7W12 B0002J4X0C B0002V70IS B0000UJLGY
categories: 11
...
Entertainment[516494]|All Sony Pictures Titles[573628]
|[139452]|DVD[130]|Special Features[408328]|Titles[579510]|( M )[579536]
|[265523]|Amazon.com Stores[285080]|Amazon.com Outlet[517808]|Categories[526028]|DVD
Outlet[734006]|Deals Under $20[734044]
|[139452]|DVD[130]|Special Features[408328]|Today's Deals in DVD[409298]|All Deals[882766]
|[139452]|DVD[130]|Special Features[408328]|Today's Deals in DVD[409298]|Kids &
Family[531554]|General[883074]
|[265523]|Amazon.com Stores[285080]|Amazon.com Outlet[517808]|Categories[526028]|DVD
Outlet[734006]|Kids & Family[734018]|General[883756]
|[265523]|Amazon.com Stores[285080]|Amazon.com Outlet[517808]|Categories[526028]|DVD Outlet[734006]|All
Deals[883766]
|[139452]|DVD[130]|Genres[404276]|Kids & Family[163414]|Characters & Series[292345]|Maggie and the
Ferocious Beast[1063886]
reviews: total: 2 downloaded: 2 avg rating: 5
2002-8-5 cutomer: A14OJS0VWMOSWO rating: 5 votes: 2 helpful: 1
2003-3-21 cutomer: A2C27IQUH9N1Z rating: 5 votes: 4 helpful: 4
...
```

Parabéns! Você concluiu as etapas de implementação e execução de uma aplicação para realizar operações no HDFS utilizando a biblioteca Java do Hadoop.