**Professor Eduardo Kugler Viegas**

**Big Data**

**Prática HDFS**

***Teste 1***

1. Copie o arquivo para o HDFS utilizando o fator de replicação = 1.
   1. Arquivo: ‘/home/dados/sensores/base\_parcial.csv’
   2. Definir o fator de replicação: hadoop fs -D dfs.replication=1 -copyFromLocal ...
2. Marque com um ‘*X*’ na tabela a localização de cada bloco criado

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Datanode1** | **Datanode2** | **Datanode3** |
| Bloco 1 |  |  | X |
| Bloco 2 | X |  |  |
| Bloco 3 |  |  |  |
| Bloco 4 |  |  |  |
| Bloco 5 |  |  |  |
| Bloco 6 |  |  |  |

1. Qual o tamanho total utilizado pelo HDFS para armazenamento do arquivo (verifique nos datanodes)?
   1. Localização dos blocos em cada datanode:

/hadoop/dfs/data/current/BP-\*/current/finalized/subdir0/subdir0/

* 1. Para visualizar o tamanho utilize o comando na pasta dos blocos

ls -lah \*

***RESPOSTA:***

*Bloco 1 -> Datanode 3 -> 128 M*

*Bloco 2 -> Datanode 1 -> 87 M*

*Tamanho total -> 215M*

1. Qual o conteúdo da primeira e última linha de cada bloco gerado?
   1. Localização dos blocos em cada datanode:

/hadoop/dfs/data/current/BP-\*/current/finalized/subdir0/subdir0/

* 1. Para visualizar a última linha utilize o comando na pasta dos blocos

tail -1 ARQUIVO

* 1. Para visualizar a primeira linha utilize o comando na pasta dos blocos

head -1 ARQUIVO

|  |  |  |
| --- | --- | --- |
| **#** | **Primeira linha completa?** | **Última linha completa?** |
| Bloco 1 | FC0754BD6E2EEA2C0BAA80A4CADCACDF;17;11;1743;11;01;5.968;57.059998;10.34(SIM) | 503FED8CBDA6E45A80AEFBCFD0054F5;1;40;2013;06;01;22.733;3  4.56(NÃO) |
| Bloco 2 | ;134.25002(NÃO) | ABA9D24F91E9C036BB4289A4E54B996F;19;2;1903;11;01;26.259998;18.51;72.62999  (SIM) |
| Bloco 3 |  |  |
| Bloco 4 |  |  |
| Bloco 5 |  |  |
| Bloco 6 |  |  |

1. O que aconteceu com entradas no arquivo com a divisão em blocos?

***RESPOSTA:***

*Foram quebradas em blocos de acordo com o tamanho máximo do bloco.*

***Teste 2***

1. Antes de começar este experimento limpe o seu HDFS, para tanto no namenode execute:
   1. hadoop fs -rmr /
2. Copie o arquivo para o HDFS utilizando o fator de replicação = 2.
   1. Arquivo: ‘/home/dados/sensores/base\_parcial.csv’
   2. Definir o fator de replicação: hadoop fs -D dfs.replication=2 -copyFromLocal ...
3. Marque com um ‘*X*’ na tabela a localização de cada bloco criado

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Datanode1** | **Datanode2** | **Datanode3** |
| Bloco 1 | X | X |  |
| Bloco 2 | X |  | X |
| Bloco 3 |  |  |  |
| Bloco 4 |  |  |  |
| Bloco 5 |  |  |  |
| Bloco 6 |  |  |  |

1. Qual o tamanho total utilizado pelo HDFS para armazenamento do arquivo (verifique nos datanodes)?
   1. Localização dos blocos em cada datanode:

/hadoop/dfs/data/current/BP-\*/current/finalized/subdir0/subdir0/

* 1. Para visualizar o tamanho utilize o comando na pasta dos blocos

ls -lah \*

***RESPOSTA:***

*Datanode 1 -> 128M(Bloco 1) + 87M(Bloco 2)*

*Datanode 2 -> 128M(Bloco 1)*

*Datanode 3 -> 87M(Bloco 2)*

*Tamanho total -> 430M*

1. Qual o conteúdo da primeira e última linha de cada bloco gerado?
   1. Localização dos blocos em cada datanode:

/hadoop/dfs/data/current/BP-\*/current/finalized/subdir0/subdir0/

* 1. Para visualizar a última linha utilize o comando na pasta dos blocos

tail -1 ARQUIVO

* 1. Para visualizar a primeira linha utilize o comando na pasta dos blocos

head -1 ARQUIVO

|  |  |  |
| --- | --- | --- |
| **#** | **Primeira linha completa?** | **Última linha completa?** |
| Bloco 1 | FC0754BD6E2EEA2C0BAA80A4CADCACDF;17;11;1743;11;01;5.968;57.059998;10.34(SIM) | 503FED8CBDA6E45A80AEFBCFD0054F5;1;40;2013;06;01;22.733;3  4.56(NÃO) |
| Bloco 2 | ;134.25002(NÃO) | ABA9D24F91E9C036BB4289A4E54B996F;19;2;1903;11;01;26.259998;18.51;72.62999  (SIM) |
| Bloco 3 |  |  |
| Bloco 4 |  |  |
| Bloco 5 |  |  |
| Bloco 6 |  |  |

1. Como a replicação influenciou nos testes?

***RESPOSTA:***

*Ela duplicou o espaço utilizado pelo arquivo e, também, permitiu que o tail e head de vários blocos fossem acessíveis em um datanode.*

***Teste 3***

1. Antes de começar este experimento limpe o seu HDFS, para tanto no namenode execute:
   1. hadoop fs -rmr /
2. Copie o arquivo para o HDFS utilizando o fator de replicação = 3 e tamanho do bloco como 64MB.
   1. Arquivo: ‘*/home/dados/sensores/base\_parcial.csv’*
   2. Definir o fator de replicação: hadoop fs -D dfs.block.size=67108864 -D dfs.replication=2 -copyFromLocal ...
3. Marque com um ‘*X*’ na tabela a localização de cada bloco criado

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Datanode1** | **Datanode2** | **Datanode3** |
| Bloco 1 | X | X | X |
| Bloco 2 | X | X | X |
| Bloco 3 | X | X | X |
| Bloco 4 | X | X | X |
| Bloco 5 |  |  |  |
| Bloco 6 |  |  |  |

1. Qual o tamanho total utilizado pelo HDFS para armazenamento do arquivo (verifique nos datanodes)?
   1. Localização dos blocos em cada datanode:

/hadoop/dfs/data/current/BP-\*/current/finalized/subdir0/subdir0/

* 1. Para visualizar o tamanho utilize o comando na pasta dos blocos

ls -lah \*

***RESPOSTA:***

*Tamanho Total -> 215 MB\* 3 = 645 MB*

1. Como o tamanho do bloco influenciou nos testes?

***RESPOSTA:***

*O tamanho dos blocos influenciou no número de blocos criados pelo HDFS para dividir o arquivo e no número total de blocos com o fator de replicação igual a 2.*

***Teste 4***

1. Antes de começar este teste espere o professor!
2. Antes de começar este experimento limpe o seu HDFS, para tanto no namenode execute:
   1. hadoop fs -rmr /
3. Copie o arquivo para o HDFS utilizando o fator de replicação = 3 e tamanho do bloco como 64MB.
   1. Arquivo: ‘*/home/dados/sensores/base\_parcial.csv’*
   2. Definir o fator de replicação: hadoop fs -D dfs.block.size=67108864 -D dfs.replication=3 -copyFromLocal ...
4. Espere o professor avisar para ir ao próximo passo
5. Copie o arquivo para o HDFS para o sistema local
   1. Comando, no namenode, hadoop fs -copyToLocal ARQUIVOHDFS ARQUIVOLOCAL

O arquivo foi copiado?

***RESPOSTA:***

1. Espere o professor avisar para ir ao próximo passo
2. Copie novamente o arquivo para o HDFS para o sistema local
   1. Comando, no namenode, hadoop fs -copyToLocal ARQUIVOHDFS ARQUIVOLOCAL

O arquivo foi copiado?

***RESPOSTA:***