**Exercício**:

* Repita todos os passos da demonstração, utilizando como base

**arquivo de entrada no linux**: /var/log/ambari-agent/ambari-agent.log

**diretório de entrada no HDFS**: /user/hdpadmin/wordcountlab2

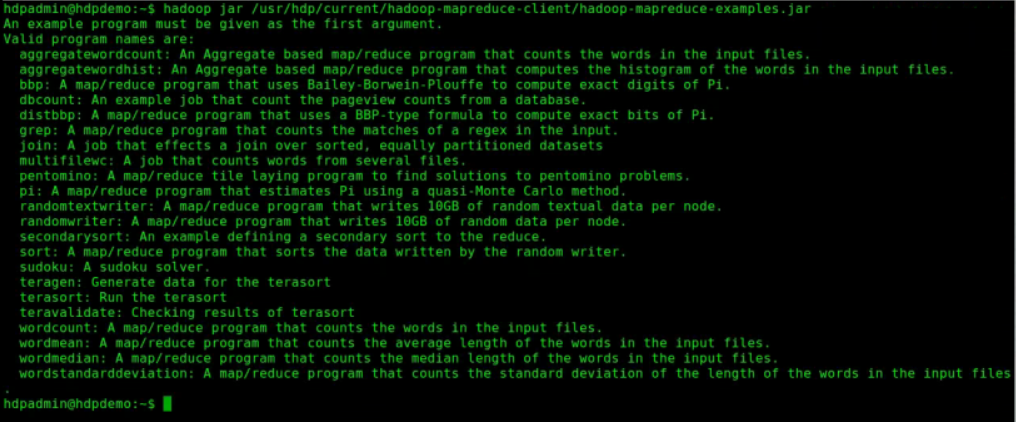
**diretório de saida no HDFS**: /user/hdpadmin/wordcountlab2/output

**Objetivo:** Familiarizar-se com os exemplos de MapReduce e acompanhamento de Jobs no YARN

***Resolução do Exercício:***

1. Inicie a máquina virtual e dê um start nos serviços HDFS, YARN e MapReduce2 e aguarde o termino do carregamento.
2. Siga as instruções em sala para criar o arquivo center\_earth.txt
3. Abra um terminal novo terminal
4. Rode o comando a seguir para ver quais são os exemplos disponíveis:

hadoop jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.jar



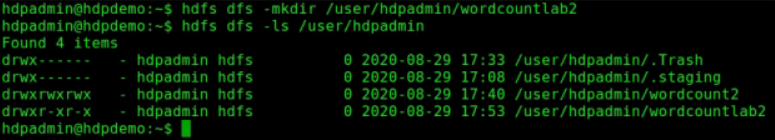
1. Rode o exemplo wordcount para ver quais parâmetros devem ser utilizados:

hadoop jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.jar wordcount



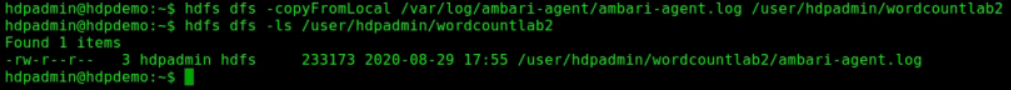
1. Crie no hdfs o diretório /user/hdpadmin/wordcountlab2

hdfs dfs -mkdir /user/hdpadmin/wordcountlab2



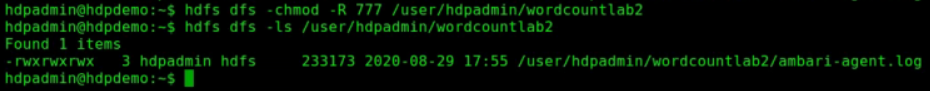
1. Copie o arquivo ambari-agent.log para o diretório criado

hdfs dfs -copyFromLocal /var/log/ambari-agent/ambari-agent.log /user/hdpadmin/wordcountlab2



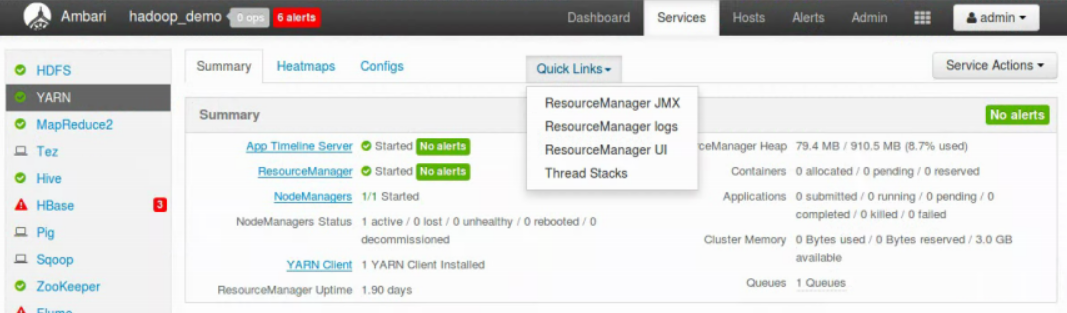
1. Altere a permissão do diretório criado recursivamente para 777

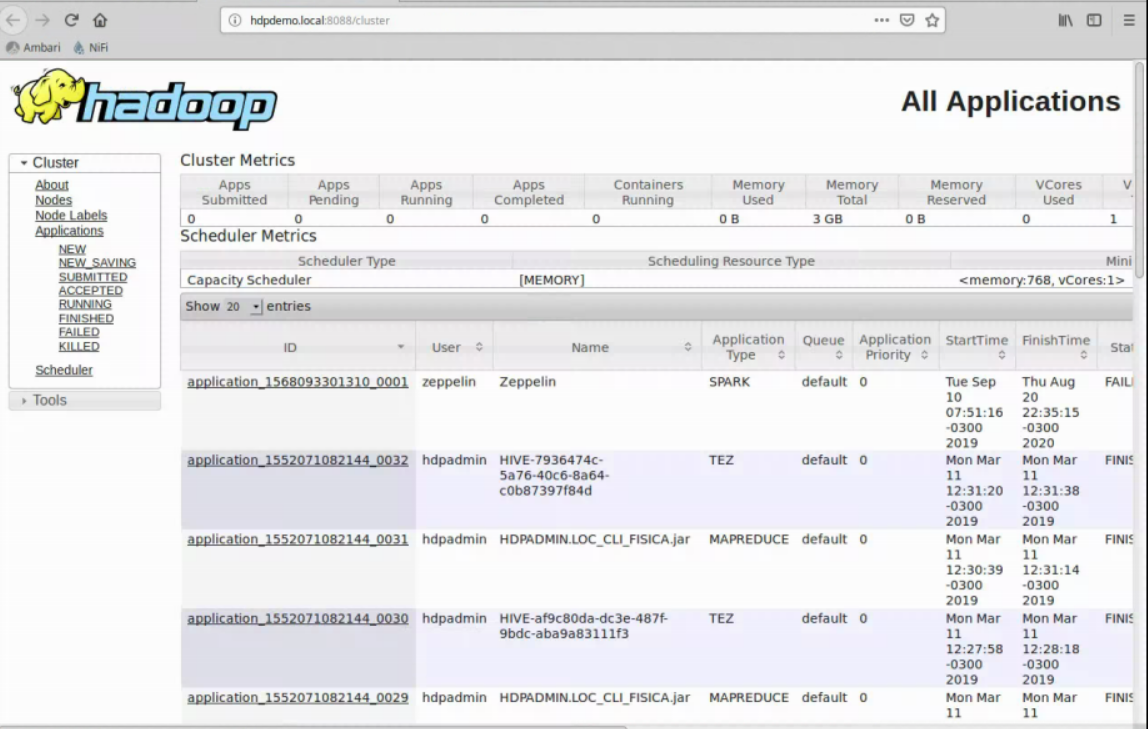
hdfs dfs -chmod -R 777 /user/hdpadmin/wordcountlab2



1. Abra a página de monitoramento de processos do YARN e faça uma rápida análise dos processos listados

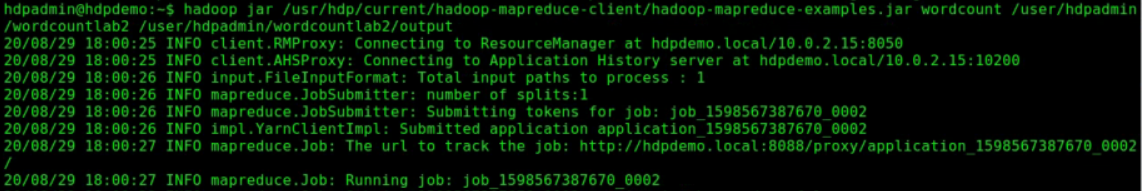
Ambari -> YARN -> Quick Links -> ResourceManager UI



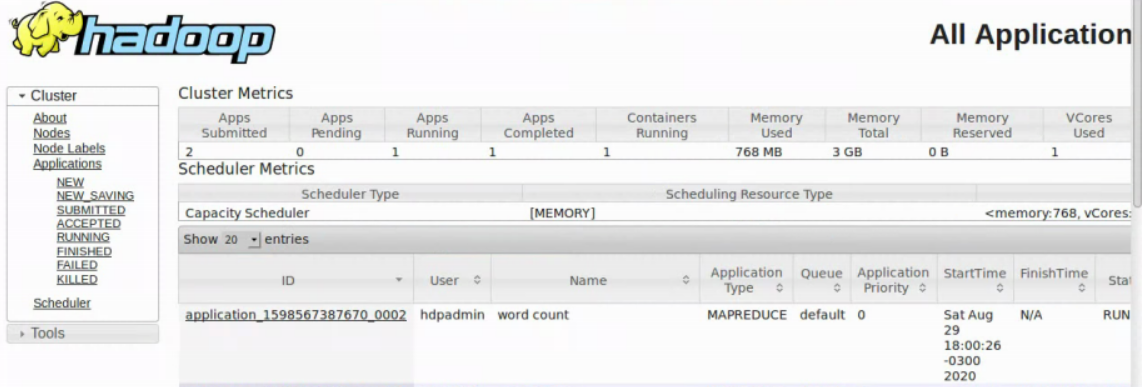


1. Submeta o programa wordcount ao YARN

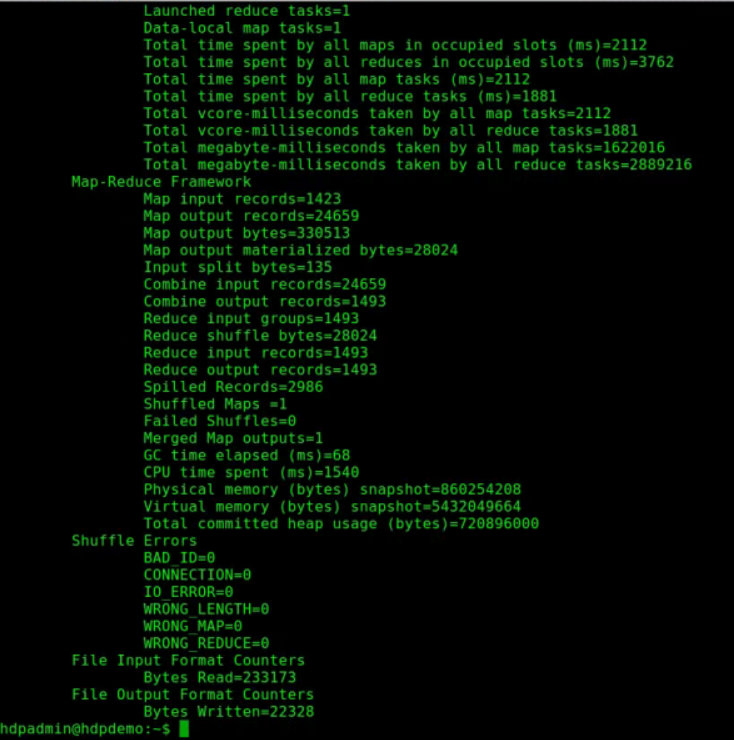
hadoop jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples.jar wordcount /user/hdpadmin/wordcountlab2 /user/hdpadmin/wordcountlab2/output



1. Acompanhe a execução do processo na UI do YARN e evidencie que o processo foi finalizado com sucesso

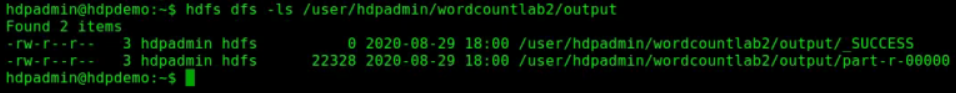






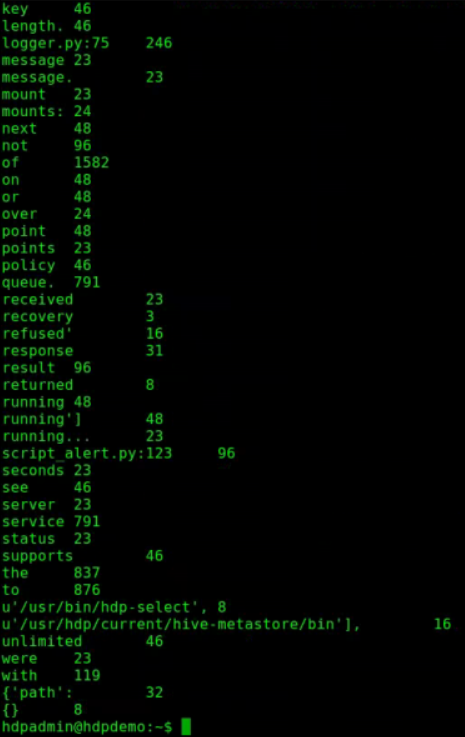
1. Liste os arquivos presentes no diretório de saída do comando:

hdfs dfs -ls /user/hdpadmin/wordcountlab2/output



1. Faça uma análise dos arquivos presentes no diretório, explicando cada um deles do ponto de vista dos conceitos de MapReduce
2. Verifique o resultado do processo de wordcount imprimindo o conteúdo do arquivo no terminal:

hdfs dfs -cat /user/hdpadmin/wordcount/output/part-r-00000



**Desafio**:

* Você deve ter notado na demonstração que a presença de pontuação no texto original alterou o resultado da contagem. Procure gerar um novo arquivo sem pontuação a partir do center\_earth.txt e utilize o programa de contagem de palavras nesse arquivo.

**Dica**: O comando Linux abaixo abre o arquivo **inputfile**, remove todas as pontuações e salva no arquivo **outputfile**

cat **inputfile** | tr -d '[:punct:]' > **outputfile**

**Dica2**: Procure evidenciar com cuidado cada um dos passos, incluindo a evidência do arquivo sem pontuação.

***Apresentando a execução do desafio:***

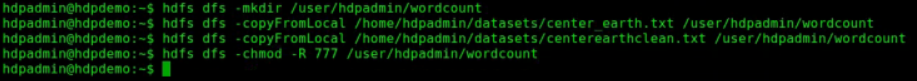
* Aqui criamos o arquivo centerearthclean.txt onde não contém pontuação, através do comando do Linux citado acima;



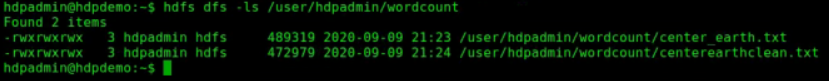
* Aqui demonstramos os arquivos utilizados para o desafio, onde podemos observar que o center\_earth.txt que tem a pontuação normal está com um tamanho maior que o arquivo gerado sem pontuação;



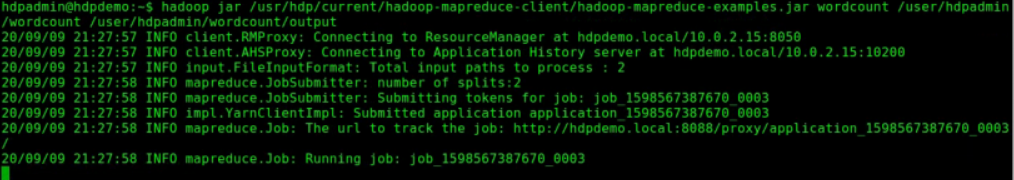
* Abaixo criamos a pasta wordcount para realizar a execução do desafio, onde copiamos os arquivos do ambiente local para dentro do hadoop, através do comando -copyFromLocal e ajustamos as permissões para acesso a pasta e seu objetos;



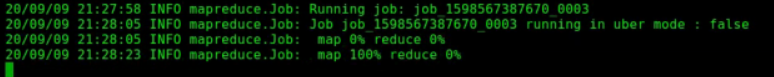
* Segue apresentação dos arquivos dentro da pasta criada;



* Executamos a wordcount para os arquivos montados;



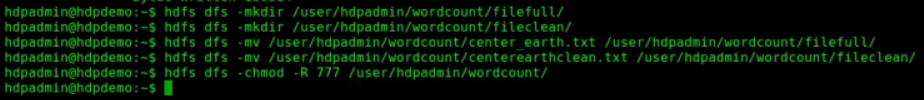
* Apresentamos o job sendo executado;



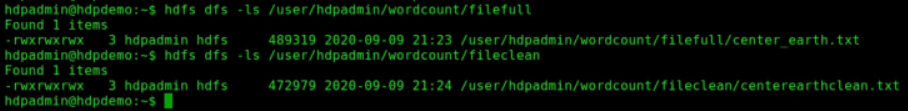
* Apresentamos o resultado do wordcount, onde ele fez a contagem de ambos os arquivos juntos;



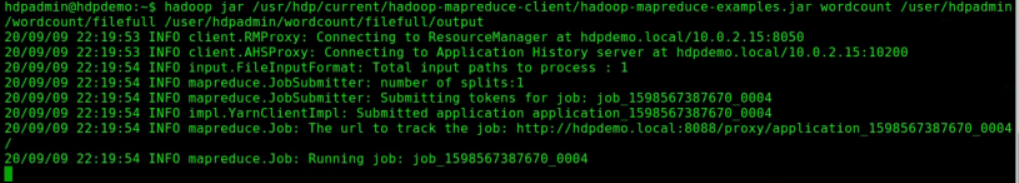
* Para separarmos a contagem dos arquivos, criamos uma pasta específica para cada arquivo, denominadas filefull e fileclean, para os arquivos center\_earth.txt e centerearthclean.txt, respectivamente;



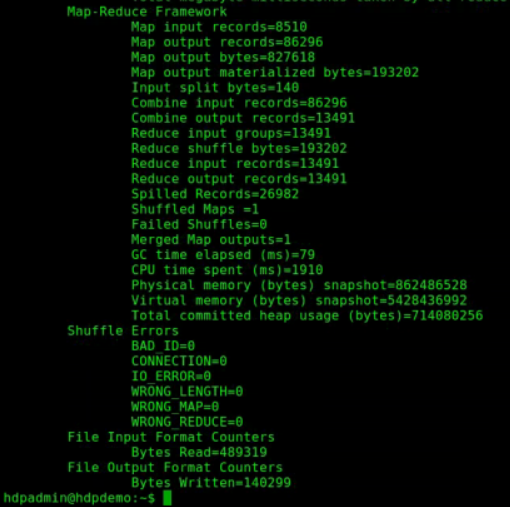
* Apresentamos os arquivos dentro de suas respectivas pastas;



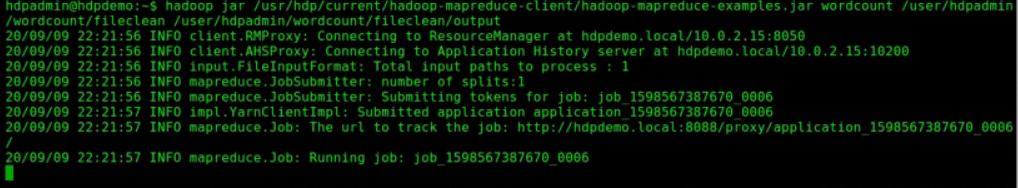
* Realizamos a contagem do arquivo center\_earth.txt, onde apresentamos o mapreduce executando o job de contagem;



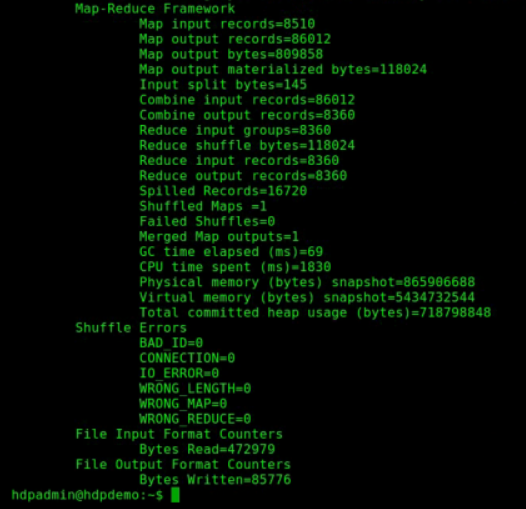
* Apresentamos o resultado da contagem do arquivo completo;



* Realizamos a contagem do arquivo centerearthclean.txt (arquivo que teve pontuação retirada), onde apresentamos o mapreduce executando o job de contagem;



Apresentamos o resultado da contagem do arquivo centerearthclean.txt, onde apresenta uma contagem de caracteres menor de que o arquivo full, devido a retirada da pontuação;

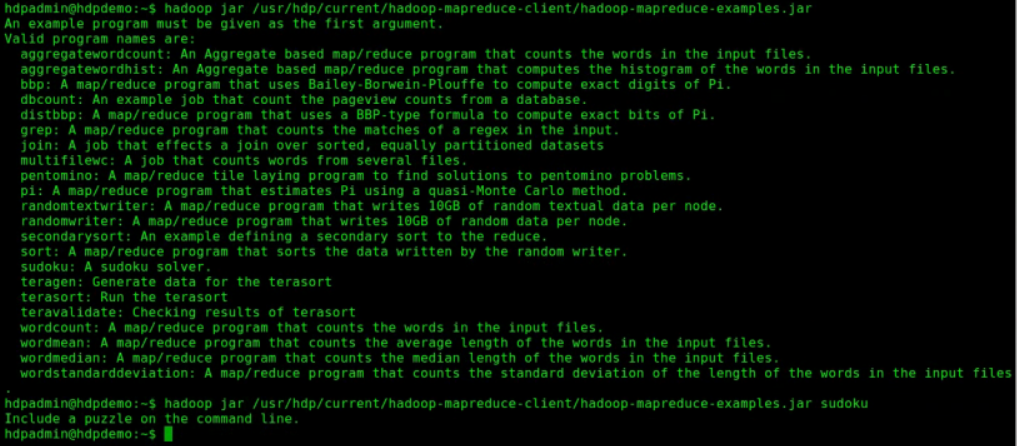


**Desafio 2**:

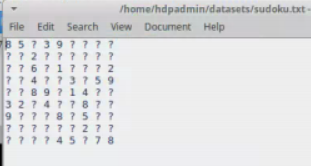
* Utilize o MapReduce para resolver o Sudoku enviado em arquivo no portal

***Apresentando a execução do Desafio2***

Verificamos a disponibilidade do comando sudoku e quais as necessidades para sua execução, onde contatamos a necessidade de informar o dezenho do sudoku a ser resolvido;



Aqui apresentamos o arquivo (sudoku.txt) utilizado para que o mapreduce resolva



Aqui temos o comando de execução do job e o resultado do mapreduce para solução do sudoku;



***Integrantes:***

*Carlos Eduardo Zanchetta – RM 335232*

*Thiago Silva Veiga – RM 336295*

*Victor Hugo R. de Oliveira – RM 335798*

*Wellington Rodrigo Nonato – RM 335826*

***Turma 35SCJ***