

Endereços úteis do ambiente

Container	Endereço
JobTracker	http://IP_CONTAINER:8088
TaskTracker1	http://IP_CONTAINER:8042
TaskTracker2	http://IP_CONTAINER:8042
TaskTracker3	http://IP_CONTAINER:8042
NameNode	http://IP_CONTAINER:9870
DataNode1	http://IP_CONTAINER:9864
DataNode2	http://IP_CONTAINER:9864
DataNode3	http://IP_CONTAINER:9864
SparkMaster	http://IP_CONTAINER:8080
SparkWorker1	http://IP_CONTAINER:8081
SparkWorker1	http://IP_CONTAINER:8081
SparkWorker1	http://IP_CONTAINER:8081
SparkJupyter	Acesse o terminal do container, execute (jupyter notebook list), altere o IP e acesse http://IP_CONTAINER:8888

Lista Comandos Docker

Descrição	Comando
Iniciar containers do hadoop	<code>docker-compose up</code>
Listar containers em execução	<code>docker container list</code>
Exibir IP do container	<code>docker inspect ID</code>
Acessar terminal do container	<code>docker exec -it NOME_CONTAINER /bin/bash</code>
Encerrar todos os containers	<code>docker stop \$(docker ps -a -q)</code>
Encerrar container	<code>docker stop ID</code>
Limpar containers	<code>docker system prune</code>

Para replicar o nosso ambiente você deve apenas:

- Dentro da pasta /home/dockers/, execute “./script.sh”
- Caso queira iniciar o ambiente novamente, repita o passo anterior =)

Lista Comandos HDFS

Os comandos devem ser executados no terminal do namenode!

Para visualizar via navegador acesse http://IP_NAMENODE:9870

Descrição	Comando
Exibir conteúdo diretório	<code>hadoop fs -ls /</code>
Copiar arquivo local para HDFS	<code>hadoop fs -copyFromLocal ARQUIVO LOCAL CAMINHO REMOTO</code>
Copiar arquivo do HDFS para local	<code>hadoop fs -copyToLocal CAMINHO_REMOTO ARQUIVO_LOCAL</code>
Criar diretório	<code>hadoop fs -mkdir /DIRETORIO</code>
Excluir arquivo	<code>hadoop fs -rm ARQUIVO</code>
Excluir pasta	<code>hadoop fs -rmr PASTA</code>
Copiar arquivo local para HDFS, como fator de replicação	<code>hadoop fs -D dfs.replication=X -copyFromLocal ARQUIVO_LOCAL CAMINHO_REMOTO</code>

Lista Comandos MapReduce

Checklist:

- O arquivo a ser processado deve estar armazenado no HDFS
- Os arquivos de map.py e reduce.py devem estar no JobTracker
- A submissão deve ser efetuada no JobTracker

Descrição	Comando
Submissão da tarefa (execute no JobTracker)	<code>hadoop jar /opt/hadoop-3.1.3/share/hadoop/tools/lib/hadoop-streaming-3.1.3.jar -file map.py -mapper map.py -file reduce.py -reducer reduce.py -input ARQUIVO_HDFS -output CAMINHO_HDFS</code>