

Objetivos

- Perceber os fundamentos do modelo de programação **MapReduce**, concretizados na infraestrutura Apache Hadoop, nomeadamente:
 - Funções **map** e **reduce**
 - Dados de entrada e de saída
 - Tipos de sistemas de ficheiros suportados pelo Apache Hadoop (**file://** e **hdfs://**)
 - Acesso ao sistema de ficheiros **HDFS** utilizando a interface de linha de comandos (**CLI**) e a API Java.
 - Configuração e parametrização de aplicações MapReduce
1. Estude os exemplos disponíveis na plataforma Moodle;
 2. No contexto do exemplo **MapReduce** de Contagem de palavras (Ex10-WordCount-01) execute diferentes execuções modificando:
 - a) O *input* e o *output* de dados de modo que o mesmo possa ser local (**file://**) ou no sistema de ficheiros distribuído (**hdfs://**).
 - b) O número de *reducers*.
 - c) Modifique o exemplo para poder reconfigurar os dados da aplicação utilizando configurações passadas na linha de comando. Neste contexto modifique por exemplo os *codecs* de compressão utilizado para guardar o resultado do processamento.
 3. Desenvolva um novo *job* **MapReduce** capaz de realizar o *merge* de diferentes *jobs* **MapReduce** do tipo Ex10-WordCount-01.
 4. Desenvolva dois programas em Java que possam ser utilizados para copiar dados entre o sistema de ficheiros local (tipo de sistema de ficheiros **file://**) e o sistema de ficheiros HDFS (tipo de sistema de ficheiros **hdfs://**).
 5. Compare o desempenho dos programas que desenvolveu na alínea anterior com as ferramentas disponibilizadas de raiz pela infraestrutura Hadoop que possibilitam a cópia de dados entre diferentes tipos de sistemas de ficheiros.