

# BIG DATA

Lucas Azevedo Dias

- 1) O *Map* faz o processamento inicial sobre os dados apenas verificando ocorrências e criando os pares chaves-valores. Enquanto isso, o *Reduce* faz o processamento final sobre os pares chaves-valores já concatenados.
- 2) O *Map* e o *Reduce*.
- 3) Em um *DataNode* serão executadas uma quantidade de funções de *Map* igual a de blocos do arquivo naquele nodo.
- 4) Há o paralelismo a nível de cluster (entre máquinas) e a nível de nodo (entre os processos de cada função de *Map*).
- 5) Entre elas, há dependência do *Reduce* sobre o *Map*, porém, individualmente, elas podem ser executadas de forma paralela (o que dá um grande ganho de performance).