

Bancos de Dados não Convencionais  
2o. Trabalho – Bancos de Dados NoSQL e MapReduce

Favor responder às questões em um zip de arquivos e enviar por email para [valmeida@gmail.com](mailto:valmeida@gmail.com).

### MapReduce

**Questão 1 (1,0 ponto):** Carregar os dados do Twitter disponíveis no Dropbox, pasta 2o\_Trabalho, no Hadoop HDFS. Os dados do Twitter estão em um formato texto com duas colunas de números inteiros identificadores de usuários. Cada linha neste arquivo representa que o usuário da esquerda segue o usuário da direita.

Recomendações:

- Utilizar a máquina virtual Cloudera QuickStart VM ([http://www.cloudera.com/downloads/quickstart\\_vms/5-8.html](http://www.cloudera.com/downloads/quickstart_vms/5-8.html)), pois já vem com o hadoop instalado e configurado.

**Questão 2 (3,0 pontos):** Fazer um programa MapReduce em Java que calcule, para cada usuário da direita, quantos seguidores este possui.

Recomendações:

- Os dados do arquivo estão separados por espaços simples
- Este programa é similar à operação de agregado (*group by*) da álgebra relacional que discutimos em sala de aula

**Questão 3 (1,0 ponto):** Recuperar o arquivo de resultado do job MapReduce da etapa anterior. Este será o resultado a ser enviado

### Banco de Dados NoSQL - MongoDB

**Questão 4 (1,0 ponto):** Importar no MongoDB o arquivo CPGF.json no Dropbox, pasta 2o\_Trabalho, contendo os dados do Portal da Transparência relativo a Despesas - Gastos Diretos - Cartão de Pagamento do Governo Federal (CPGF), referente ao ano de 2016 (janeiro a julho), mesmos dados utilizados no trabalho anterior.

Recomendações:

- Usar o utilitário mongoimport do MongoDB

**Questão 5 (2,0 pontos):** Montar consulta que mostre o valor total de transações (agregado da soma de “Valor\_Transação”) por favorecido (“Nome\_Favorecido”). Na mesma consulta, remover os valores cujo favorecido não aparece, casos de saques já identificados no trabalho anterior. Remover também os valores de “Informações protegidas por sigilo, nos termos da legislação, para garantia da segurança da sociedade e do Estado”, por não fazer sentido no resultado. Ainda na mesma consulta, ordenar o resultado em ordem decrescente e mostrar os 10 mais.

Recomendações:

- Utilizar a função *aggregate*, e colocar os filtros de consulta em *\$match*
- Reparem que o MongoDB não possui o conceito de nulos. Nos casos em que o favorecido não aparece, simplesmente não há esses campos no documento em questão. Há uma função para recuperar somente documentos cujo campo existe, pesquise!

**Questão 6 (2,0 pontos):** Montar consulta que mostre o valor total de transações (agregado da soma de “Valor\_Transação”) por portador (“Nome\_Portador”). Na mesma consulta, remover os valores de “Informações protegidas por sigilo, nos termos da legislação, para garantia da segurança da sociedade e do Estado”, por não fazer sentido no resultado. Filtrar somente para saques no exterior (Transação = "SAQUE - INT\$ - APRES"). Ainda na mesma consulta, ordenar o resultado em ordem decrescente e mostrar os 10 mais.

Recomendações:

- Utilizar a função *aggregate*, e colocar os filtros de consulta em *\$match*