

Sistemas de Computação em Cloud

MIEI - 2014/2015

2º teste, 6/12/2014, 1h30'

- 1- Considere um modelo de programação Map/Reduce como no Hadoop.
 - a) Apresente o pseudocódigo do mapper e do reducer que lhe permita inverter a ordem das linhas de um ficheiro de texto. Admita que os registos lidos do ficheiro de entrada têm como chave o número de linha e como valor o texto da linha.
 - b) Implementando o programa da alínea anterior em Hadoop, diga justificando, em que condições será possível paralelizar a fase de mapper, tirando partido de um cluster com vários nós (considere a situação onde processa um só ficheiro e a situação onde processa vários ficheiros).

- 2- Descreva as características específicas dos sistemas de virtualização do tipo “paravirtualization”. Apresente, justificando, as vantagens e desvantagens deste tipo face aos de “fullvirtualization”.

- 3-
 - a) Descreva o objectivo do serviço S3-Simple Storage System da Amazon. Resuma as principais operações suportadas.
 - b) Descreva uma situação em que seja usado (ou seja preferível usar) o serviço EBS-Elastic Block Store em vez do S3.

- 4-
 - a) Indique o objectivo de um serviço IaaS oferecido por uma Cloud. Descreva as principais operações suportadas.
 - b) Quais as vantagens e as desvantagens para um utilizador/empresa em recorrer a este tipo de serviço para a implementação de um determinada solução informática.

- 5- Para lidar com um grande volume de dados e a sua distribuição o Hadoop pode recorrer a um sistema de ficheiros distribuído chamado HDFS (semelhante ao Google FS).
 - a) Descreva como este sistema consegue guardar dados de dimensão superior à capacidade de um disco ou de todo um computador.
 - b) Descreva como o Hadoop, na execução das aplicações, procura otimizar a leitura dos dados pelos mappers minimizando os acessos a dados que estejam remotos.

- 6- Considere o serviço Chubby (ou Zookeeper) executando num cluster de vários servidores.
 - a) Será possível dois clientes criarem em concorrência um ficheiro com o mesmo nome, ficando cada um a usar a sua própria cópia? Justifique.
 - b) Explique como este serviço lida eficientemente com a carga de um grande número de clientes simultâneos?

- 7- Descreva, justificando, em que condições optaria por uma solução de base de dados dita NoSQL (como a BigTable) em vez uma tradicional BD relacional. (p.e indique os requisitos da aplicação que o levam a optar pela NoSQL)