

Sistemas de Computação em Cloud

MIEI - 2014/2015

Exame, 12/1/2015, 2h30'

- 1- Usando o modelo map/reduce como o do framework Hadoop pretende-se implementar um programa para contar as linhas de um conjunto ficheiros de texto.
 - a) Admita que na entrada cada registo corresponde a uma linha de um ficheiro, sendo a chave o número da linha e o valor essa mesma linha. Apresente as operações map e reduce do seu programa em pseudo-código que permita obter o respectivo número de linhas em cada ficheiro.
 - b) Suponha que se avaliou a execução deste programa usando um único processo sobre um conjunto de ficheiros e que a fase de map correspondeu a 70% do tempo de execução. Considerando que se configurou a aplicação para usar apenas um reducer, qual o speedup máximo possível ao usar vários processos em vários computadores? Justifique.

- 2- Pretende-se usar o middleware MPI num cluster para implementar uma solução baseada no modelo Map/Reduce que envolve o processamento de um grande número de ficheiros.
 - a) Descreva uma arquitectura para o programa MPI que permita executar de forma distribuída e em paralelo as operações de map e de reduce.
 - b) Explique a execução e as mensagens trocadas entre processos, que permitem a distribuição dos dados pelos processos, e a execução das operações de map e reduce deste programa. Assuma um sistema de ficheiros partilhado entre todos os computadores.

- 3- Apresente as características dos problemas que necessitam e que tiram melhor partido dos grandes sistemas de clusters, Grid ou Cloud, considerando os problemas que são limitados pelo tempo de computação e os que são limitados pelo processamento de dados. Justifique.

- 4- O sistema de Grid GT4 dispõem de serviços para gestão de dados (Data Management). Descreva o papel destes serviços e dê um exemplo de um destes serviços e da sua utilização.

- 5- Os clusters são tipicamente geridos por sistemas responsáveis pela gestão dos nós computacionais e pelo escalonamento dos trabalhos (exemplo: Condor).
 - a) Indique como é que um utilizador pede o lançamento de um trabalho (job) para execução, incluindo a informação que deve ser indicada nesse pedido.
 - b) O escalonador dos trabalhos em fila de espera pode incluir a técnica de *backfilling*. Diga em que consiste esta técnica e justifique como esta pode melhorar o uso dos recursos e qual o impacto nos tempos de atendimento dos trabalhos (jobs).

- 6- Descreva os principais problemas colocados à monitoração do estado dos grandes sistemas de Grid ou Cloud, para a obtenção da visão global do estado dos seus recursos.

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

- 8-
 - a) Indique o objectivo do um serviço IaaS oferecido por uma Cloud. Descreva as principais operações suportadas (pode basear-se no serviço EC2).
 - b) Descreva como este serviço pode ser usado na implementação de serviços do tipo PaaS, como é o caso do uso do serviço EC2 pelo serviço da Amazon EMR.

- 9- Considere uma base de dados dita NoSQL (como a BigTable).
 - a) Descreva como estes sistemas conseguem guardar dados de dimensão superior à capacidade de um disco ou de todo um computador.
 - b) Descreva como estas bases de dados conseguem responder ao aumento de carga colocado pelos pedidos dos clientes.
 - c) Como podem estes sistemas manter os tempos de acessos e latências baixas para os clientes, independentemente da localização geográfica desses clientes? Justifique.