

**** Protocolo de Autocorreção****Avaliação On-line – HADOOP E HDFS**

1) Considerando as afirmativas referentes aos subprojetos HDFS e MapReduce, assinale as alternativas verdadeiras.

Descrição	Resposta	Justificativa
Os elementos chave do Hadoop são o paradigma de programação HDFS e o sistema de arquivos distribuídos MapReduce	Falsa	O HDFS é responsável pelo Sistema de arquivos distribuídos, enquanto o MapReduce é um paradigma de programação.
MapReduce é um paradigma de programação com foco no processamento distribuído sobre uma grande massa de dados	Verdadeira	O paradigma de programação MapReduce permite o processamento sobre os dados armazenados no HDFS.
O Hadoop Distributed File System (HDFS) permite o processamento e transmissão de grandes massas de dados	Falsa	O HDFS é responsável apenas pelo armazenamento dos dados.
O modelo de programação MapReduce pode ser dividido em apenas duas funções a Map e Reduce	Verdadeira	A função Map é responsável pelo mapeamento dos dados, enquanto a função Reduce é responsável pela agregação dos dados.
Para o desenvolvimento de aplicações MapReduce deve-se utilizar o modelo de programação MapReduce, sendo que este pode ser implementado em diversas linguagens de programação	Verdadeira	O modelo de programação MapReduce é independente de linguagem de programação.
Para a utilização dos subprojetos HDFS e MapReduce torna-se necessário a utilização de máquinas de alto custo	Falso	As ferramentas disponíveis no Hadoop podem ser utilizadas em máquinas comuns.

2) Considerando seu conhecimento referente ao Hadoop, relacione a função a cada componente.

Descrição	Resposta	Justificativa
Responsável pelo gerenciamento do HDFS	NameNode	O NameNode atua como uma entidade que é responsável pelo gerenciamento dos blocos no HDFS, entre os DataNodes.
Responsável pelo armazenamento do HDFS	DataNode	O DataNode é responsável pelo armazenamento dos blocos no HDFS.
Responsável pelo gerenciamento das tarefas MapReduce	JobTracker	O JobTracker efetua o gerenciamento das tarefas MapReduce, gerenciando os estados das tarefas entre os TaskTrackers responsáveis.
Responsável pela execução das tarefas Map e Reduce	TaskTracker	O TaskTracker é responsável pela execução das tarefas atribuídas a ele pelo JobTracker..
Atua para a recuperação de falha do NameNode	SecondNameNode	Atua como um NameNode em caso de falha.

3) Considerando os componentes do Hadoop, analise cada afirmativa e assinale as verdadeiras.

Descrição	Resposta	Justificativa
O HDFS utiliza dois componentes em sua função, o NameNode e o DataNode. O NameNode é responsável pelo gerenciamento dos arquivos no HDFS, enquanto o DataNode é responsável pelo efetivo armazenamento.	Verdadeiro	O NameNode, efetua o gerenciamento, enquanto o DataNode efetua o armazenamento dos blocos.
O DataNode além do armazenamento dos dados, também é responsável pelo monitoramento de possíveis alterações realizadas nos mesmos.	Verdadeiro	Cada DataNode é responsável pelo monitoramento de seus blocos.
O DataNode é responsável pela execução das tarefas MapReduce.	Falso	A entidade responsável pela execução das tarefas MapReduce é o TaskTracker.
A função do JobTracker é gerenciar o HDFS, seu objetivo é determinar qual DataNode irá armazenar um bloco.	Falso	A entidade responsável pelo gerenciamento do HDFS é o NameNode.
O JobTracker é responsável pela execução das tarefas Map e Reduce do modelo de programação MapReduce	Falso	O JobTracker é responsável pelo gerenciamento das tarefas MapReduce.
O TaskTracker é responsável pela execução das tarefas Map e Reduce do modelo de programação MapReduce	Verdadeiro	O TaskTracker efetua o processamento das tarefas enviadas ao JobTracker.
O JobTracker é responsável pelo gerenciamento da execução das tarefas de MapReduce. Seu principal objetivo é determinar quais TaskTrackers irão executar a tarefa MapReduce, e posteriormente efetuar o monitoramento destas tarefas	Verdadeiro	O JobTracker efetua o gerenciamento dos TaskTrackers disponíveis para a execução das tarefas.

A função do SecondaryNameNode é auxiliar o DataNode a manter seu serviço, e ser uma alternativa de recuperação no caso de uma falha do DataNode.	Falso	O SecondaryNameNode atua apenas em caso de falha do NameNode.
Quando uma tarefa MapReduce é submetida, o JobTracker é responsável pelo seu gerenciamento, enquanto o TaskTracker pelo efetivo processamento.	Verdadeiro	O JobTracker efetua o gerenciamento das tarefas MapReduce, enquanto o TaskTracker efetua o processamento das tarefas atribuídas a ele.

4) Analise as assertivas a seguir, referentes às formas de execução do Hadoop e assinale as verdadeiras.

Descrição	Resposta	Justificativa
Os componentes do Hadoop, devem necessariamente executar de maneira distribuída (utilizando várias máquinas)	Falso	É possível executar todos os componentes em uma única máquina.
Em ambos os modos pseudo-distribuído e completamente distribuído todos os componentes do Hadoop são executados em máquinas distintas.	Falso	No modo pseudo-distribuído os componentes são executados de maneira similar a um ambiente distribuído, porém utiliza-se apenas uma máquina.
O modo local é recomendado para a fase de testes e depuração do código de aplicações.	Verdadeiro	A depuração de aplicações distribuídas é uma tarefa complexa, desta maneira a utilização de apenas uma máquina facilita o processo de análise do código.
No modo completamente distribuído é necessário indicar quais componentes serão executados em cada máquina.	Verdadeiro	Como várias máquinas são utilizadas, torna-se necessário definir quais máquinas irão executar quais componentes.

5) Analise as afirmativas referentes a sistemas de arquivos distribuídos e assinale as opções verdadeiras.

Descrição	Resposta	Justificativa
Sistemas de arquivos distribuídos surgiram com o advento do Big Data.	Falso	Sistemas de arquivos distribuídos já eram utilizados antes do HDFS.
O único objetivo de um sistema de arquivo distribuído é prover escalabilidade.	Falso	Um sistema de arquivos distribuídos provém diversas garantias para o armazenamento dos dados, como por exemplo a disponibilidade em caso de falha.
Um sistema de arquivos distribuídos deve atuar como um sistema de arquivos tradicional, porém, adicionalmente, deve permitir o uso de hardwares diferentes, sendo que estes são conectados através de uma rede.	Verdadeiro	Um Sistema de arquivos distribuídos deve permitir o uso de diversas máquinas, de maneira transparente ao usuário.
Atualmente, o HDFS é o único sistema de arquivos distribuído em utilização.	Falso	Existem diversos sistemas de arquivos distribuídos concorrentes ao HDFS.

6) Considerando as afirmativas referentes a HDFS, assinale as alternativas verdadeiras.

Descrição	Resposta	Justificativa
O HDFS permite a utilização de vários discos rígidos, sendo assim, a partir do momento em que o primeiro disco rígido esteja sem espaço disponível, o próximo disco rígido é utilizado.	Falso	O foco do HDFS é permitir o armazenamento distribuído através da utilização de várias máquinas distintas.
O HDFS atua sobre a arquitetura mestre/escravo, possuindo no lado mestre uma instancia do DataNode e em cada escravo uma instancia do NameNode.	Falso	O NameNode atua como o mestre, enquanto o DataNode atua recebendo as ordens do NameNode.
Para o armazenamento de arquivos com grandes volumes, o HDFS divide os arquivos em blocos, sendo que cada bloco possui o tamanho padrão de 64 MB.	Verdadeiro	O HDFS permite o armazenamento de grandes massas de dados através da quebra de um determinado arquivo em blocos, e a posterior distribuição destes blocos entre várias máquinas.
O HDFS é capaz de armazenar grandes volumes de dados uma vez que os arquivos são divididos em blocos menores, e diversos diversas máquinas são utilizadas para armazenar os blocos.	Verdadeiro	Uma vez que os blocos podem ser armazenados por várias máquinas.
O NameNode não é responsável pelo armazenamento dos dados, esta tarefa é efetuada pelo DataNode.	Verdadeiro	A função do NameNode é atuar no gerenciamento do HDFS.
Quando um arquivo é armazenado no HDFS, apenas os primeiros 64 MB, referentes ao primeiro bloco, são armazenados.	Falso	O arquivo inteiro é armazenado no HDFS, porém o mesmo é quebrado em blocos e os blocos são distribuídos em várias máquinas.
Independentemente da capacidade de armazenamento dos DataNodes, não existem limites de armazenamento no HDFS	Falso	O limite de armazenamento no HDFS está relacionado ao armazenamento disponível nas máquinas utilizadas como DataNodes.

A capacidade de armazenamento do HDFS é definida de acordo com a capacidade de armazenamento dos DataNodes disponíveis	Verdadeiro	Uma vez que os blocos são armazenados nos DataNodes.
Periodicamente, o DataNode envia uma requisição ao NameNode para receber mais blocos para armazenamento.	Falso	O NameNode é responsável por determinar se o DataNode irá armazenar um determinado bloco.
O HDFS é capaz de aumentar a capacidade de armazenamento quando comparado aos sistemas de arquivos tradicionais devido ao tamanho do bloco utilizado, sendo 64 MB do HDFS e 512 bytes dos sistemas tradicionais.	Falso	O HDFS utiliza um sistema de arquivos tradicional para armazenamento dos blocos, o ganho está relacionado a utilização de diversas máquinas.
Caso um arquivo inferior ao tamanho do bloco (64MB por padrão) seja adicionado ao HDFS, o espaço restante não é perdido, podendo ser assim utilizado pelos próximos arquivos.	Verdadeiro	Cada bloco utiliza apenas o tamanho necessário para o armazenamento dos dados relacionado ao conteúdo armazenado.
O HDFS efetua a replicação dos blocos na tentativa de melhorar a segurança, desta maneira, caso um escravo se torne indisponível o arquivo original não é perdido.	Verdadeiro	O HDFS é tolerante a falha caso utilize a replicação dos blocos.

O HDFS efetua a replicação dos blocos na tentativa de melhorar o desempenho, desta maneira, diversos escravos podem responder uma única requisição de acesso ao bloco	Falso	A divisão do arquivo em blocos está relacionada ao armazenamento dos dados.
Quando um usuário deseja armazenar um arquivo no HDFS, o mesmo necessita apenas dividir o arquivo em blocos de 64MB e enviar cada bloco aos DataNodes desejado.	Falso	O processo é transparente ao usuário, a entidade responsável pela divisão dos arquivos em blocos é o NameNode.
Para armazenar um arquivo no HDFS é necessário apenas efetuar uma requisição ao NameNode, o processo de divisão em blocos e efetivo armazenamento é transparente ao usuário.	Verdadeiro	O HDFS torna o Sistema de arquivos transparente ao usuário.
Em uma aplicação Hadoop, cada nó escravo contém um DataNode, que trabalha em conjunto com um TaskTracker, sendo que o primeiro efetua o processamento dos dados, e o segundo o seu gerenciamento.	Falso	O DataNode efetua o armazenamento no HDFS, enquanto o TaskTracker efetua o processamento das tarefas MapReduce.
É recomendável a utilização de um nó escravo como um DataNode ou um TaskTracker, uma vez que caso um nó seja utilizado tanto para armazenamento quanto para processamento, o mesmo pode ser sobrecarregado.	Falso	É recomendável que um nó atue tanto no armazenamento, quanto no processamento dos dados, diminuindo assim a utilização da rede para o processamento.
O DataNode tipicamente atua em conjunto com um TaskTracker, sendo que o TaskTracker processa os blocos armazenados pelo DataNode.	Verdadeiro	Tipicamente utiliza-se o MapReduce aliado ao HDFS.