

Gestão de Centros de Dados

Data: 13/Dezembro/2012

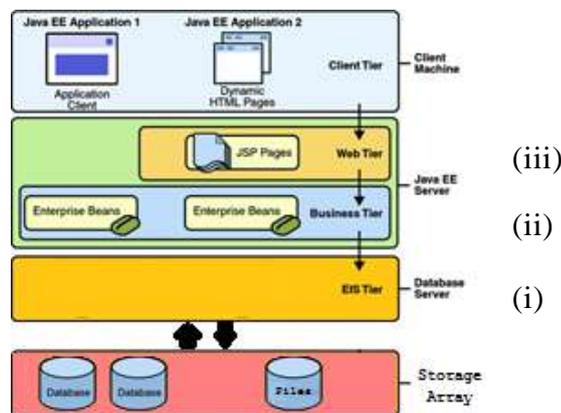
Duração: 1h45m

Nota: Algumas questões podem não estar completamente especificadas. Isto é intencional, e deverá assumir hipóteses razoáveis para depois formular as suas respostas.

No trabalho prático proposto para a disciplina foi-lhe sugerido que, para preparar um “protótipo bem sucedido” de uma aplicação que está concentrada numa única máquina, para funcionar no “mundo real”, em produção, fizesse o seu desacoplamento em camadas (níveis i a iii na Fig. 1), colocando os discos relevantes num “disk array” com acesso iSCSI.

Foi-lhe ainda pedido que, em cada camada, demonstrasse a fiabilidade e/ou escalabilidade da solução colocando vários (i.e., mais de um) servidores por camada, sendo-lhe sugerido que cada servidor (disk array incluído) fosse implementado como uma VM.

Figura 1



- i) uma base de dados, suportada por um SGBD relacional “open source” típico (e.g., MySQL);
- ii) um servidor aplicacional suportado por uma tecnologia distribuída (e.g., Java Beans);
- iii) uma camada de apresentação suportada numa tecnologia integrada com Web servers.

Q-1 [10 val.] Imagine que todo o ambiente desacoplado, fiável e escalável anteriormente descrito pôde ser implementado com VMs suportadas num único servidor físico que você tinha na sua empresa; tudo correu bem e é agora necessário implementar esse ambiente no CD do cliente, que já tem à sua espera num rack 2 switches Ethernet de 24 portas cada, um disk array com 2 portas Ethernet e 4 servidores com 4 portas Ethernet cada – tudo já ligado numa configuração com caminhos redundantes. Tem ainda disponíveis dois DVDs: um com uma distribuição Linux e outro com um hipervisor. Suponha que, se instalasse a solução num servidor, o desempenho “dava à justa” para as necessidades.

a) [4 val.] Descreva a melhor solução a propor ao seu cliente baseada em instalações do SO (Linux) directamente apoiadas no hardware. Descreva as tarefas que tem de realizar, os servidores vai usar (todos? Deixa algum “de fora”?). Tente, ainda que de uma forma muito simples, contabilizar os custos de instalação (em tempo) e de exploração (energia gasta, sabendo que se 1 servidor gasta 1 UFE - Unidade Fictícia de Energia, 4 servidores gastam 4 UFEs).

b) [4 val.] Descreva a melhor solução a propor ao seu cliente baseada em instalações de VMs apoiadas num hipervisor igual ao que usou na sua empresa. Descreva as tarefas que tem de realizar, e que servidores vai usar (todos? Deixa algum “de fora”?). Tente, ainda que de uma forma muito simples, contabilizar os custos de instalação (em tempo) e de exploração (energia, sabendo que se 1 servidor gasta 1 UFE - Unidade Fictícia de Energia, 4 servidores gastam 4 UFEs).

c) [2 val.] Suponha agora que o hipervisor usado na instalação “de produção” é “de outro fornecedor”, distinto portanto do que usou na sua empresa. Que tarefa(s) adicionais são de esperar para além das descritas em (b)? Essas tarefas têm um impacto significativo nos custos (tempo) de instalação? Justifique.

Q-2 [6 val.] Considere o “ambiente de produção” descrito na Questão 1; vamos agora tratar do sistema de backups. Considere-se que o volume de dados de um disco de sistema (de cada servidor) e software neste instalado ronda os 20 GB, mas que o volume de dados da BD ronda os 300 GB; considere-se ainda como negligenciável o espaço ocupado por páginas HTML, imagens, etc., servidas pelo(s) Web servers. A unidade de “tape backup” disponível tem uma interface FC e faz um máximo de 100 MB/s, e a janela de backup é de 30 minutos; note que dos 4 servidores, só um tem interface FC, e portanto é a esse que tem de ser ligada a unidade.

a) [3 val.] Descreva a solução a propor ao seu cliente para fazer backups na configuração descrita em Q-1-a), i.e., baseada em instalações do SO (Linux) directamente apoiadas no hardware: como são transferidos os dados dos servidores para a tape, com que periodicidade faz as diferentes cópias (discos de sistema/software, BD, etc.) que estratégias usa para minimizar o volume de dados copiado, etc.

b) [3 val.] Idem, mas para a configuração descrita em Q-1-b), i.e., baseada em instalações de VMs suportadas sobre um hipervisor.

Q-3 [4 val.] [Nota: nesta questão ignore tudo o que diz respeito a virtualização] Recorde que, na Questão 1, se disse: “Suponha que, se instalasse a solução num servidor, o desempenho “dava à justa” para as necessidades”; como havia 4 servidores idênticos, ao desacoplar as 3 camadas pode, por exemplo, colocar cada camada num servidor – correndo-se o risco de estar a subutilizar muito um servidor enquanto outro está bastante mais ocupado. O “ideal” seria usar para cada camada o servidor mais “apropriado” (i.e., o mais “barato” que cumpre os objectivos). Imagine que tem à sua disposição um leque variado de servidores (para os quais tem os resultados dos benchmarks típicos usados para caracterizar estes sistemas) e, claro, tem a sua aplicação. Se quisesse produzir uma tabela que lhe permitiria escolher, para cada camada, qual o “servidor ideal” para um determinado número de utilizadores da aplicação, como procederia? E, como daí a algum tempo, alguns servidores deixariam de ser fabricados e outros (para os quais teria de novo os resultados dos benchmarks típicos usados para caracterizar estes sistemas) tomariam o seu lugar, como faria para tentar manter essa “tabela” actualizada?