

Além do mascaramento e da subdivisão: compreendendo a importância do gerenciamento de dados

Huw Price
CA Technologies

Sumário

Introdução	3
Os obstáculos de uma abordagem puramente logística para TDM	3
A alternativa ideal: pessoas, processo e tecnologia	7
Referências	11
A superioridade da CA Technologies	11
Sobre o autor	12

Seção 1

Introdução

O conceito de TDM (Test Data Management – Gerenciamento de Dados de Teste) pode ser novo, mas costuma ser negligenciado, subestimado e mal compreendido, de modo que todo o valor que ele pode agregar à empresa acaba não sendo aproveitado. Para muitas organizações e vários fornecedores, ele é visto apenas como uma questão de ambiente, que se resume aos processos de cópia, mascaramento e, possivelmente, subdivisão dos dados de produção. Depois de migrados, esses dados são considerados uma "gold copy", que é utilizada nos ambientes de desenvolvimento e controle de qualidade, e o objetivo de um TDM aprimorado se torna o gerenciamento mais ágil dessa cópia, bem como o cumprimento da necessidade de novas cópias.

A geração de dados sintéticos, se utilizada, é realizada de maneira isolada em projetos específicos. Isso costuma ser sustentado pela premissa de que a geração de dados é uma boa prática para uma equipe ou um projeto individual, mas é inviável para uma empresa inteira. Em geral, as empresas modernas têm bancos de dados grandes e complexos, que armazenam dezenas de milhões de registros em diversos formatos e utilizam vários conjuntos de ferramentas. O argumento é que é mais fácil copiar os dados de produção existentes do que tentar definir o perfil de dados tão complexos como esses e modelá-los, como seria necessário para gerar dados sintéticos realistas.

No entanto, essa premissa é questionável, e as organizações que desejam obter os benefícios de um TDM aprimorado devem reavaliar a geração de dados sintéticos no nível corporativo e a forma como armazenam, gerenciam e provisionam dados. A geração de dados sintéticos não é só mais eficiente em termos de tempo, qualidade e custo, muitas vezes, ela também prova ser mais fácil e segura do que o mascaramento total dos dados de produção – desde que realizada com a tecnologia certa, os processos adequados e as alterações estruturais de equipe que são necessárias.

Este documento compara os problemas comuns e os possíveis benefícios do TDM do ponto de vista de ambiente, considerando a tecnologia, as estruturas de equipe e os processos normalmente encontrados nas organizações¹.

Seção 2

Os obstáculos de uma abordagem puramente logística para TDM

Para organizações que utilizam dados de produção para os ambientes de desenvolvimento e teste, o mascaramento e a subdivisão são práticas obrigatórias. O mascaramento é necessário para cumprir a lei atual de proteção de dados, já a subdivisão dos dados de produção pode ajudar a reduzir os altíssimos custos de infraestrutura.

No entanto, se o TDM se limitar ao mascaramento e à subdivisão, seu valor real não será realizado para a empresa. Isso pode ser demonstrado com base em duas premissas: 1 - os dados coletados por qualquer organização moderna conterão informações pessoais que, pela lei, não podem sair dos ambientes de produção em uma forma reconhecível; 2 - como os dados mascarados são para fins de desenvolvimento e teste, eles precisam ser suficientes para todos os casos de teste que serão executados e qualquer funcionalidade nova.

Seguindo essas premissas, primeiro, será mostrado por que o mascaramento pode, na verdade, não ser a maneira mais fácil nem mais a eficiente de provisionar dados seguros "adequados à finalidade" dos ambientes de desenvolvimento e teste. Serão considerados a tecnologia, as estruturas de equipe e os processos implementados por uma organização típica, argumentando que a proteção deficiente dos dados de produção, além da forma como esses dados são normalmente gerenciados, indica que os problemas comuns de atraso de projeto, orçamento excedido e defeitos na produção continuarão se a empresa depender apenas do mascaramento e da subdivisão.

O mascaramento não é a solução fácil

Antes de mais nada, é difícil acreditar que copiar e mascarar os dados de produção de maneira segura para os ambientes de desenvolvimento e teste seja a opção mais simples em um departamento complexo de TI. Na verdade, o esforço para mascarar os dados de produção e, ao mesmo tempo, manter a integridade referencial será, muitas vezes, maior do que o esforço para modelar os dados, já que de qualquer forma será preciso definir o perfil de dados para o mascaramento ser eficiente. Além disso, certos aspectos complexos dos dados costumam não ser mascarados, representando uma espécie de comprometimento. Portanto, em muitos casos, o mascaramento também pode não ser a opção mais segura.

Ao mascarar dados para uso em ambientes de teste, é preciso manter os relacionamentos de coluna dos dados originais, mesmo quando o conteúdo confidencial está mascarado. Em um nível mais simples, isso significa que a integridade referencial é mantida. Já em um nível mais complexo, significa que são mantidos os relacionamentos complexos de coluna. Por exemplo, se houver um campo "total", calculando o total de duas ou mais colunas, essas colunas terão que permanecer alinhadas nos dados mascarados. Os relacionamentos temporais e causais de dados (por exemplo, um total baseado em tempo) são mais complexos e extremamente difíceis de serem mascarados de maneira consistente.

E mais, para uma organização típica de grande porte com vários tipos de bancos de dados e aplicativos compostos complexos, esses relacionamentos não precisam só ser mascarados de maneira consistente em um banco de dados (intrassistema), eles precisam ser mantidos em todos os bancos de dados (intersistema). Quanto mais complexos os dados, mais difícil isso será, embora se torne mais fácil violar os dados, já que há mais informações para serem correlacionadas.

Geralmente, o mascaramento de dados tem como foco principal o conteúdo, e os relacionamentos inter e intrassistemas, para manter a integridade referencial. Devido à complexidade de manter o conteúdo, os relacionamentos de coluna e os relacionamentos temporais ao mesmo tempo, geralmente uma dessas frentes acaba sendo sacrificada – por exemplo, é comum os relacionamentos temporais entre colunas serem mantidos inalterados. Embora esse conteúdo não seja confidencial, esses dados inalterados podem ser utilizados para identificar informações confidenciais, por meio da correlação deles com informações externas aos dados.

Tome como exemplo uma situação na qual há um conjunto de logs mascarados de transação e o invasor sabe que certa pessoa fez uma transação de determinado valor em uma hora específica. Embora essas informações não sejam encontradas no conteúdo confidencial, que foi mascarado, informações temporais ainda podem estar presentes no banco de dados mascarado, já que o mascaramento consistente do conteúdo, dos totais numéricos e dos horários de transação é extremamente difícil. Uma vez identificadas na primeira instância, as informações confidenciais podem ser identificadas em todos os bancos de dados e sistemas, pois a integridade foi mantida tanto dentro dos sistemas quanto entre eles. Isso, na verdade, descaracterizará a anonimidade dos dados.

Mesmo se o esforço da definição de perfil e da modelagem dos dados tiver sido realizado em toda sua complexidade, os dados ainda não poderão ser considerados como protegidos da maneira adequada, porque as informações ainda estarão preservadas em seus aspectos mais complexos. Por exemplo, um invasor poderia identificar os efeitos causais em cascata de um log de transação, combinando informações causais inter e intrassistema com métodos temporais, como uma comparação de instantâneos. Assim, seria possível deduzir como os dados foram alterados durante uma operação e qualquer efeito causal em cascata existente (por exemplo, qualquer gatilho acionado no banco de dados). Como diz o ditado, "Uma corrente é tão forte quanto seu elo mais fraco". Depois que essas informações são identificadas, o invasor pode trabalhar com os relacionamentos de coluna, decifrando requisitos comerciais confidenciais ou, pior, informações pessoais.

Considerando a premissa de que os dados de praticamente qualquer organização conterão algumas informações pessoais, o mascaramento não apresenta uma maneira segura nem eficiente de provisionar dados para ambientes de teste. Isso se comprova com a pressão crescente sobre as organizações para que elas protejam totalmente os dados: a multa média para uma violação de dados aumentou 13%, chegando a US\$ 3,5 milhões em 2013 [5, Ponemon, 2014], além disso, espera-se que o futuro regulamento geral de proteção de dados da UE reforce a lei existente que proíbe o uso de informações pessoais para qualquer fim diferente daquele para o qual foram coletadas.

Levando em conta o resultado do relatório "State of Privacy Report 2015" da Symantec, que indica que a preocupação com a segurança de dados determina onde e como 88% dos consumidores da Europa fazem compras, e o risco do uso de dados mascarados em ambientes de teste, o mascaramento se torna uma opção inviável. Agora que já estabelecemos os riscos legais da migração de dados de produção para os ambientes de desenvolvimento e teste, podemos considerar a eficácia e eficiência dessa prática.

Pessoas

Dependências de dados

Não é comum uma organização ter uma equipe central responsável pelo TDM, em geral, equipes individuais são responsáveis por gerenciar, localizar e criar os próprios dados. Assim sendo, as equipes trabalham isoladas uma das outras, em seus próprios ambientes de desenvolvimento ou teste, mas costumam utilizar as mesmas fontes de dados. Essa falta de centralização e colaboração leva a restrições de dependência de dados – se uma equipe faz uma alteração em um banco de dados, todas as equipes são afetadas por essa alteração. Isso gera frustração, atrasos e retrabalho, já que os testes apresentam falha aparentemente sem motivo, e as equipes não conseguem identificar se isso foi causado por um defeito no código ou um erro de dados. Muitas vezes, as equipes também não sabem determinar a versão ou os parâmetros dos dados de teste e não têm como reverter os bancos de dados se outra equipe os canibaliza.

O SDLC (Software Development Life Cycle – Ciclo de Vida de Desenvolvimento de Software) é visto sequencialmente, como uma série de fases lineares, na qual uma equipe termina seu trabalho e o encaminha para a próxima. Como consequência, é comum as equipes de desenvolvimento e teste verem seus projetos parados devido a atrasos "ascendentes". Elas têm que esperar os dados serem disponibilizados ou especificações e feeds de dados serem concluídos por outras equipes. Assim, boa parte do tempo do SDLC pode ser gasta aguardando dados.

Essa falta de paralelismo vai contra a propensão atual de muitas organizações para a entrega contínua, na qual as equipes precisam dos dados em cada etapa do SDLC. Isso torna impossível, por exemplo, desativar as versões antigas de um sistema de maneira eficiente e, ao mesmo tempo, utilizar os dados existentes para novas versões. Em vez disso as equipes ficam com um novo ambiente de desenvolvimento, mas sem os dados necessários nele.

Tecnologia

Dependências de sistema

Além das dependências de dados, as restrições de hardware e sistema são um problema comum para as equipes modernas de desenvolvimento e teste. Nas últimas duas décadas, os aplicativos foram se tornando cada vez mais complexos, com um conjunto crescente de dependências de outros sistemas, tanto dentro quanto fora da organização. Geralmente, isso leva a empecilhos quando se deseja testar um sistema, já que os serviços podem ser instáveis ou não estar disponíveis. Outra preocupação relatada pelas equipes de teste e desenvolvimento é a falta de ambientes completos. As equipes podem descobrir que outra equipe tem prioridade sobre o ambiente que elas desejam utilizar ou podem não conseguir obter os dados corretos porque eles estão sendo usados por outra equipe. Embora a princípio essas restrições possam não parecer diretamente relevantes aos dados, posteriormente, será explicado como uma infraestrutura de TDM aprimorada é primordial para a resolução deles.

Armazenamento de dados

As organizações modernas armazenam grandes volumes de dados de produção. Ter cópias desses bancos de dados e executá-las nas máquinas de desenvolvimento é uma prática tanto cara quanto lenta. O armazenamento desses dados gera altos custos de infraestrutura, incluindo custos relacionados a programas de hardware, licenças e suporte. Além disso, os servidores são submetidos a cargas cada vez mais pesadas, já que precisam fornecer altos volumes de dados para as tarefas e, ao mesmo tempo, executar conexões abertas, manter arquivos abertos e manipular injeções de dados².

A menos que a forma como os dados são gerenciados seja reavaliada, é improvável que os custos diminuam. O volume de dados coletados e armazenados por uma empresa normal dobra a cada ano³ e, agora, com o advento do "Big Data", as organizações falam em termos de manipulação de petabytes, não mais de terabytes. Como consequência, o armazenamento de dados se tornou uma das partes dos orçamentos de TI que cresce mais rápido. Acredita-se que o setor de armazenamento de dados possa ter apresentado recentemente um crescimento anual de 20%⁴. Assim sendo, as organizações devem analisar se cada cópia dos dados de produção é realmente necessária – considerando-se que algumas organizações costumam ter várias cópias de um único banco de dados, é provável que não.

Mineração e definição de perfil de dados

Sem a tecnologia parcial ou totalmente automatizada, a detecção de dados é um dos maiores desafios enfrentados pelas equipes que desejam entregar o software totalmente testado dentro do prazo, especialmente em um contexto de desenvolvimento ágil ou em uma estrutura de entrega contínua. Os responsáveis pelos testes podem perder até metade do tempo de trabalho procurando dados, o que os força a decidir entre executar todos os testes necessários para evitar que defeitos onerosos passem para o ambiente de produção ou entregar o software dentro do prazo.

Isso é ainda piorado pelo armazenamento inconsistente dos dados em planilhas não controladas (com pouca ou nenhuma referência cruzada ou indexação, por exemplo), em vez de em um data warehouse ou repositório centralizado. Além disso, geralmente, os bancos de dados não são documentados corretamente, as organizações não costumam manter dicionários centrais dos atributos de dados de teste nem consultas SQL relacionadas. A mineração e alocação de dados são, portanto, prejudicadas, e as equipes não conseguem solicitar dados com base nos modelos ou formulários padronizados nem de acordo com os critérios específicos necessários. Como consequência, muitas vezes, as equipes precisam encontrar, manualmente, um pequeno conjunto de dados que atenda aos requisitos e seja adequado aos casos de teste – um processo longo e propenso a erros.

A falta de ferramentas automatizadas de mineração e definição de perfil de dados também aumenta o risco de não conformidade. Como os dados são armazenados em planilhas não controladas, é possível encontrar informações confidenciais em qualquer lugar – por exemplo, em uma coluna de observações, quando não há um campo de coluna adequado para os dados ou se todos os campos relevantes já tiverem sido preenchidos. Sem poder procurar campos específicos, torna-se mais difícil encontrar essas informações, que podem acabar sendo transferidas para ambientes de teste. Isso vai contra o regulamento de proteção de dados, o qual estabelece que os dados só podem ser utilizados para o fim que foram coletados. Isso pode gerar multas pesadas de, em média, US\$ 3,5 milhões⁵ e prejudicar gravemente o resultado de uma empresa.

Processo

Os dados de produção não são uma "gold copy"

A incapacidade de localizar dados "adequados ao fim" para casos de teste e requisitos nos leva ao problema maior do uso de dados de produção em ambientes de teste: eles simplesmente não conseguem atender à maioria das solicitações de dados que uma organização recebe, especialmente aquelas vindas de ambientes de desenvolvimento. Como descrito, as equipes de desenvolvimento e teste precisam dos dados em cada etapa do desenvolvimento.

Assim, um banco de dados "gold copy", propriamente dito, precisa conter um conjunto padrão de dados, com o qual são realizados testes várias vezes, e armazenar dados que satisfaçam a todos os testes possíveis. Além disso, ele deve ser semelhante ao da produção e mantido atualizado, incluindo os "dados inválidos" e todos os dados anteriores. Os dados de produção só atendem a duas dessas condições: são semelhantes aos da produção e podem conter todos os dados anteriores. Portanto, eles não são uma "gold copy".

Nas transações usuais, muitos dos dados de produção são bem similares, além de serem esterilizados por sua própria natureza de excluir os dados que podem causar o colapso de um sistema. Assim sendo, esse conjunto de dados não contém "dados inválidos", deixando os novos cenários e os caminhos negativos inalterados durante os testes. No entanto, são exatamente os resultados inesperados, as exceções e as condições de restrição que costumam fazer um sistema entrar em colapso. O objetivo das equipes de desenvolvimento e teste deve ser testar os casos extremos, os caminhos infelizes e os cenários inesperados. Quando a empresa conta apenas com métodos de amostragem, os defeitos passam para o ambiente de produção, onde custam até 1.000 vezes mais para serem corrigidos⁶ e podem demorar 50 vezes mais para serem resolvidos⁷.

Além disso, os dados de amostra da produção quase nunca vão estar atualizados, devido às alterações de ambiente e às constantes mudanças da demanda corporativa enfrentadas pelas equipes de TI. Como os dados não são armazenados de maneira independente dos ambientes, quando um ambiente é alterado, é necessária uma atualização de sistema ou versão. Isso pode canibalizar dados e cenários criados com base em várias fontes de dados de produção e causar a perda de conjuntos de dados úteis, que posteriormente precisarão ser recriados manualmente. Essas atualizações de sistema também costumam ser incrivelmente lentas. Por exemplo, já trabalhamos com organizações que levaram até nove meses para concluir esse processo.

A criação manual dos dados pode ser uma solução temporária, permitindo que os casos de teste iminentes sejam realizados. No entanto, como esses dados são criados tendo em mente casos de teste ou requisitos específicos, eles ficarão desatualizados quase que imediatamente. Um bom exemplo disso são as taxas de câmbio ou os padrões comerciais. Para eles, os dados ficam desatualizados de um dia para outro, ou seja, os dados criados manualmente não podem ser reutilizados, sendo, em geral, "queimados". Assim, é preciso realizar a cansativa tarefa de criar novos dados para cada teste, o que leva a atrasos de projeto no efeito "bola de neve", enquanto os testes avançam para o próximo sprint ou são deixados para a próxima fase do ciclo de vida de desenvolvimento.

Se o mascaramento eficiente nos bancos de dados exige, primeiro, a definição do perfil de dados e ainda assim não garante a segurança nem fornece dados de qualidade satisfatória, a pergunta que não quer calar é: por que as organizações não geram os dados necessários sintaticamente assim que o perfil dos dados de produção é definido com êxito?

Seção 3

A alternativa ideal: pessoas, processo e tecnologia

Uma diretiva de TDM aprimorado envolve a adoção de uma abordagem estruturada e centralizada (requisito crucial) para gerenciar os dados no nível corporativo. Isso não só resolve muitos dos pontos problemáticos detalhados acima, mas também costuma ser uma maneira mais econômica, eficiente e fácil de provisionar dados para ambientes de desenvolvimento e teste – com a tecnologia certa.

A prática de armazenar os dados úteis como ativos fora dos ambientes e transferi-los para os ambientes sob demanda elimina os problemas de ambiente do TDM, que passa a se preocupar com a forma como os dados são devolvidos. Isso, por sua vez, pode ser realizado de maneira eficiente com uma abordagem centralizada para o armazenamento de dados, na qual os dados são modelados como ativos reutilizáveis, que podem ser extraídos como subconjuntos precisos, sob demanda. Não se contente com o mascaramento e a subdivisão como operações

necessárias. A geração de dados sintéticos apresenta um objetivo estratégico, que, quando incorporado a uma diretiva mais ampla de TDM, permite a entrega de programas de software totalmente testados, dentro do prazo e do orçamento.

Tecnologia

Definição automatizada de perfil de dados

Já foi argumentado que, para mascarar os dados de produção de maneira eficiente, primeiro, é preciso definir o perfil deles, porém, mesmo depois disso, os dados mascarados não estarão totalmente seguros, e esse processo não fornecerá os dados necessários para desenvolvimento e teste. No entanto, há uma tecnologia automatizada para reduzir o esforço necessário para definir o perfil dos dados em um cenário complexo de TI. Para definir o perfil dos dados, primeiro é preciso "registrá-los", coletando o máximo possível de metadados. Esses metadados incluem, por exemplo, nomes de tabela, nomes de coluna e tipos de coluna, e existem mesmo para sistemas de bancos de dados não relacionais, sob a forma de copybooks para sistemas de mainframe e documentos de mapeamento para arquivos simples delimitados ou de largura fixa.

Assim que o registro é realizado, é possível aplicar algoritmos de detecção de dados com base em fórmulas matemáticas. Por exemplo, com o CA Test Data Manager (anteriormente chamado de Grid-Tools Data Maker), isso é feito primeiro para esquemas individuais, identificando informações pessoais e realizando a engenharia reversa dos relacionamentos de bancos de dados, se necessário. Depois disso, é possível unificar os sistemas. Assim, o CA Test Data Manager utiliza exibições em cubo para definir o perfil até dos relacionamentos mais complexos existentes nos dados, criando um conjunto de dados multidimensional no qual cada dimensão do "cubo" representa um atributo dos dados. Essa definição de perfil permite que uma organização compreenda exatamente quais dados existem e onde eles estão armazenados e identifique lacunas na cobertura funcional.

Geração de dados sintéticos

Com uma imagem precisa dos dados existentes e a identificação dos dados que ainda são necessários para os ambientes de teste, é possível gerar automaticamente os dados ausentes. Como cada universo real pode ser visto com outro ponto de dados, é possível modelar e criar dados para cobrir 100% das variações funcionais. Esses dados incluem cenários futuros que nunca ocorreram, além de "dados inválidos", exceções e resultados inesperados. Isso permite testes negativos eficientes e o desenvolvimento de um novo sistema ou subsistema. A geração de dados sintéticos oferece uma maneira sistemática de teste, para que resultados inesperados e cenários considerados impensáveis pelos testadores não causem um colapso de sistema, e detecta os defeitos antes de eles passarem para o ambiente de produção.

Caso não haja dados suficientes para realizar testes várias vezes, também é possível criar altos volumes de dados utilizando a tecnologia automatizada. O CA Test Data Manager oferece uma ferramenta automatizada que funciona diretamente com RDBMSs (Relational Database Management Systems – Sistemas de Gerenciamento de Banco de Dados Relacional) ou camadas de API do ERP para gerar dados na maior velocidade permitida pelo poder de processamento. Seus scripts volumosos podem dobrar a quantidade de dados de uma organização tão rápido quanto o permitido pela capacidade de processamento da infraestrutura. Em resumo, essa tecnologia automatizada permite a criação rápida de conjuntos de dados semelhantes aos de produção, com todos os dados necessários para executar os testes (inclusive testes negativos) em quantidade suficiente para a repetição do processo.

Gerenciamento de dados centralizado

Com os aprimoramentos tecnológicos mencionados anteriormente, as organizações já estão bem adiantadas no processo de estabelecimento de uma "gold copy", como definido neste documento (consulte a seção Os dados de produção não são uma "gold copy"). Além disso, armazenando dados, modelados como objetos reutilizáveis, em um data warehouse central de testes, elas também podem desenvolver a capacidade de identificar rapidamente subconjuntos de dados específicos para ambientes de desenvolvimento e teste, sob demanda.

Clonagem de dados

Assim que os dados estiverem modelados como objetos em um "TestMart" ou data warehouse de testes e os dicionários dos ativos de dados, bem como as consultas associadas, estiverem estabelecidos, será possível identificar, clonar e extrair subconjuntos de dados específicos para os ambientes de desenvolvimento e teste.

Por exemplo, o módulo de clonagem de dados do CA Test Data Manager obtém conjuntos de dados de teste pequenos e coerentes de vários sistemas de desenvolvimento e produção inter-relacionados, substituindo o processo lento e dispendioso de cópia e movimentação de bancos de dados grandes e complexos. Com a capacidade de extrair, copiar e fornecer apenas os dados necessários, as organizações não precisam mais manter inúmeras cópias completas dos bancos de dados de produção.

A presença de um data warehouse centralizado e a capacidade de clonar dados podem ainda acabar com as dependências de dados entre as equipes, separando o provisionamento e consumo dos dados. Isso significa que os dados podem ser clonados e fornecidos a várias equipes ao mesmo tempo, eliminando os atrasos relacionados à espera pela disponibilização dos dados "ascendentes" e evitando que as equipes atrapalhem o trabalho umas das outras quando fazem uma alteração nos dados.

A modelagem e o armazenamento central dos dados como objetos maleáveis e reutilizáveis também permitem a fácil reprodução de bugs e cenários do interesse da empresa. Com as informações sendo expressamente obtidas na geração de relatórios de bug, a tecnologia rápida e flexível de clonagem permite que testes raros e complexos sejam repetidos, sem os dados se esgotarem. Isso é particularmente útil ao realizar uma atualização de dados, porque significa que os dados não precisam ser mesclados, e os conjuntos de dados de seu interesse não são perdidos.

Restrições de hardware

Combinada com um kit de ferramentas de virtualização, a geração de dados sintéticos também pode ajudar a resolver restrições de hardware e sistema. As camadas de mensagem podem ser simuladas, utilizando os metadados de um sistema para definir o perfil de maneira precisa e gerar mensagens de serviço realísticas (incluindo arquivos de MQ, SOAP e REST, além de arquivos simples). Aqui, há um mecanismo automatizado de geração de dados subjacente a uma máquina virtual, para preencher as respostas às mensagens realísticas, por exemplo, pares de solicitação/resposta.

A virtualização de máquinas inteiras permite a criação de vários ambientes de desenvolvimento. Isso significa que as equipes podem trabalhar nos ambientes, mesmo quando os componentes interdependentes estão indisponíveis, evitando atrasos ascendentes. Além disso, os onerosos sistemas e programas de hardware legados também podem ser virtualizados para testes.

Processo

Desenvolvimento em paralelo e reusabilidade

Além das questões de ambiente, a capacidade de detectar e clonar dados com base em atributos de dados específicos também resolve outra preocupação central do TDM, que é a forma como os dados serão provisionados de maneira eficiente para os testadores individuais ou as equipes de teste. Isso permite que os dados sejam solicitados, compartilhados e reutilizados em paralelo, sob demanda.

O acesso a um portal centralizado de dados de teste sob demanda com base na web, como o oferecido pelo CA Test Data Manager, permite que os testadores e desenvolvedores solicitem exatamente os dados de que precisam para a tarefa em questão. Quando os critérios específicos (ou seja, os atributos de dados de teste) são enviados, o portal envia uma tarefa ao mecanismo de lote, que encontra os dados adequados nos sistemas de back-end ou clona os dados e os extrai de lá. Isso elimina a necessidade de procurar ou criar os dados à mão, reduzindo significativamente o tempo necessário para atender às solicitações de dados.

Quanto mais padronizadas as perguntas no formulário, melhor, porque isso permite que as equipes reutilizem os trabalhos de maneira mais eficiente. Por exemplo, se você tiver dados sintéticos que foram criados anteriormente, poderá determinar os parâmetros da entrada e expor o resultado por meio de menus suspensos no portal, permitindo que todos solicitem os dados, mesmo que o caso de teste seja diferente. Além dos dados de teste, estruturas de criação de dados, testes de unidade, ativos virtuais e scripts de automação podem ser armazenados e utilizados como módulos para futuros trabalhos.

Controle da versão

O poderoso controle da versão permite o paralelismo necessário para desenvolver continuamente novos sistemas, dentro do prazo e do orçamento. Por exemplo, o data warehouse de testes do CA Test Data Manager permite que uma equipe copie os dados de um repositório, que na verdade os "herda", junto com indicadores das versões anteriores. Isso acomoda a evolução dos dados entre múltiplas versões, porque bloqueia os dados para uma equipe específica, ao mesmo tempo em que permite que eles sejam facilmente revertidos ou atualizados e reconciliados com versões diferentes. Quando é feita uma alteração em um local, ela é refletida em todas as versões, embora o original permaneça intacto. Digamos que uma equipe precise adicionar uma nova coluna a um banco de dados inteiro: com o CA Test Data Manager, se essa equipe tiver pais codificados, ela poderá encontrar todos os filhos vinculados e definir valores padrão ou gerar dados utilizando sequências/funções padrão.

Pessoas

Por fim, geralmente, as alterações estruturais de equipe podem complementar esses aprimoramentos tecnológicos e processuais e ajudar a resolver os pontos problemáticos de ambiente discutidos neste documento. A centralização do TDM em uma equipe dedicada está correlacionada à presença de um recurso central de armazenamento, gerenciamento e provisionamento de dados que possa atender melhor à necessidade da empresa. Essa equipe pode deter o controle do provisionamento de dados, bem como do gerenciamento dos dados em si, e ser responsável ainda por criar novos dados e definir o perfil deles, se necessário.

Isso não só ajuda a evitar restrições de dependência de dados entre equipes, mas também permite que as solicitações de dados e a geração de relatórios de bug sejam reunidas. Assim, as passagens de qualidade podem ser aplicadas, enquanto a propriedade de dados pode ser centralizada em uma equipe de segurança de TI. A criação de formulário dinâmico oferecida pelo portal de dados de teste sob demanda do CA Test Data Manager dá suporte a isso e vai além dos direitos de acesso com base em função, provisionando dados confidenciais apenas para a equipe autorizada que os solicitou.

Seção 4

Referências

- 1 Os obstáculos da utilização real de dados de produção em ambientes de teste foram abordados em outros documentos. Consulte, por exemplo, as documentações técnicas "Reduce Time to Market with Test Data Management" e "How better Test Data Management is the only way to drive Continuous Delivery", também de Huw Price.
 - 2 Jacek Becla e Daniel L. Wang, Lessons Learned from managing a Petabyte, P. 4. Recuperado em 19/02/2015, de: <http://www.slac.stanford.edu/BFROOT/www/Public/Computing/Databases/proceedings/>
 - 3 Lessons Learned from managing a Petabyte
 - 4 <http://www.computerweekly.com/feature/Meeting-the-demand-for-data-storage>
 - 5 <http://www.ponemon.org/blog/ponemon-institute-releases-2014-cost-of-data-breach-global-analysis>
 - 6 <http://benderrbt.com/Bender-Requirements%20Based%20Testing%20Process%20Overview.pdf>
 - 7 <http://www.softwaretestingclass.com/why-testing-should-start-early-in-software-development-lifecycle/>
-

Seção 5

A superioridade da CA Technologies

A CA Technologies (NASDAQ: CA) fornece soluções de gerenciamento de TI que ajudam os clientes a gerenciar e proteger ambientes de TI complexos, a fim de oferecer suporte a serviços de negócios ágeis. As organizações utilizam o software e as soluções de SaaS da CA Technologies para acelerar a inovação, transformar a infraestrutura e proteger dados e identidades, do datacenter à nuvem. A CA Technologies tem o compromisso de garantir que os clientes alcancem os resultados desejados e o valor comercial esperado através do uso da nossa tecnologia. Para saber mais sobre os programas de sucesso de nossos clientes, visite ca.com/customer-success. Para obter mais informações sobre a CA Technologies, visite ca.com/br.

Seção 6



Sobre o autor

Com uma carreira de quase 30 anos, Huw Price foi o arquiteto técnico líder de várias empresas norte-americanas e europeias e forneceu suporte de arquitetura de alto nível a bancos multinacionais, grandes fornecedores de serviços públicos e provedores de serviços de saúde. Eleito o "Diretor de TI do ano de 2010" pela QA Guild, Huw passou anos especializando-se em ferramentas de automação de testes e lançou vários produtos inovadores que reformularam o modelo de teste utilizado no setor de software. Atualmente, Huw dá palestras em eventos conhecidos internacionalmente, e seu trabalho foi publicado em várias revistas, como a Professional Tester, a CIO Magazine e outras publicações técnicas.

A última empreitada de Huw, a Grid-Tools, foi adquirida pela CA Technologies em junho de 2015. Há quase uma década, ela já vinha redefinindo a maneira como as grandes organizações abordam sua estratégia de teste. Com a liderança e a abordagem visionária de Huw, a empresa introduziu uma abordagem forte centrada em dados para os testes, lançando os novos conceitos concebidos por Huw, como "Objetos de dados", "Herança de dados" e "Um data warehouse central de testes".



Conecte-se à CA Technologies em ca.com/br



A CA Technologies (NASDAQ: CA) cria software que acelera a transformação das empresas e permite que elas aproveitem as oportunidades da economia dos aplicativos. O software está no cerne de todas as empresas, em todos os setores. Do planejamento ao desenvolvimento e do gerenciamento à segurança, a CA está trabalhando com empresas de todo o mundo para mudar a maneira como vivemos, fazemos negócios e nos comunicamos – usando dispositivos móveis, as nuvens privada e pública e os ambientes distribuídos e de mainframe. Obtenha mais informações em ca.com/br.