

Escalabilidade do Tableau Server

Guia técnico para administradores do Tableau Server

Neelesh Kamkolkar
Gerente de produtos, dados e desempenho

Sumário

- Resumo executivo3
- O VizQL quebra um paradigma ultrapassado.....5
- Arquitetura do Tableau7
- Arquitetura unificada em tempo real e de processamento na memória..... 8
- Metodologia e abordagem dos testes9
- Metodologia9
- Caracterização das cargas de trabalho reais 11
- Etapas de modelagem dos testes11
- Metodologia do Processador em segundo plano.....12
- Ambiente isolado padronizado13
- Topologia de implantação14
- Medições e relatórios15
- Cenário15
- Tempo de resposta16
- Taxa de transferência do cenário16
- Usuários ativos17
- Resultados18
- A escalabilidade do Tableau Server 10 é linear18
- Resultados do Processador em segundo plano 22
- Isolamento do processo do Processador em segundo plano 26
- Considerações sobre o Processador em segundo plano 28
- Práticas recomendadas: faça você mesmo o teste de escalabilidade28
- Práticas recomendadas para otimização de ambientes reais29
- Resumo30

Resumo executivo

O Tableau é um recurso essencial para muitas organizações. Como a possibilidade de encontrar informações relevantes nos dados oferece grandes benefícios a vários departamentos dessas empresas, as equipes de TI estão colaborando com o restante da organização para distribuir o Tableau como uma plataforma de análise disponível para toda a empresa. À medida que as empresas começam a implantar o Tableau dessa forma, é fundamental que os arquitetos empresariais e os líderes de TI entendam como o Tableau Server permite a escalabilidade de dados, conteúdos e usuários e como ele pode ser implementado e integrado a plataformas de TI corporativas diversas e heterogêneas para suprir as necessidades analíticas da empresa hoje e no futuro.

Este whitepaper é destinado a arquitetos empresariais e líderes técnicos de TI. Ele proporciona uma visão detalhada da arquitetura do Tableau e de como ela pode ser escalonada à medida que as cargas de trabalho crescem.

Examinamos e testamos como o Tableau Server 10 é escalonado e como os resultados se comparam às versões anteriores do produto. Atendendo a pedidos dos clientes, ampliamos o escopo dos testes de escalabilidade para incluir cargas de trabalho em segundo plano além da escalabilidade de usuários.

Existem vários fatores relacionados aos sistemas que podem influenciar o desempenho e a escalabilidade do Tableau Server. Algumas importantes variáveis de sistema incluem o design das pastas de trabalho, a configuração do servidor, ajustes da infraestrutura, o ambiente de dados, a capacidade computacional e o sistema de rede. Esses fatores variam muito dependendo dos perfis de uso e dos tipos de implantação. Os resultados de qualquer teste de escalabilidade podem ser diferentes conforme essas variáveis são ajustadas ou alteradas. Na tentativa de isolar as variáveis e reduzir seu impacto, realizamos nossos testes em um laboratório fechado de máquinas físicas conectadas em rede. Nosso objetivo foi minimizar a variabilidade das medidas devido a influências de sistemas externos. Assim, pudemos medir métricas de escalabilidade modelando o uso real do Tableau Server.

Para isso, começamos analisando uma implantação de produção real do Tableau Server em seu pico de uso e, em seguida, modelamos esse uso em testes automatizados. Essa abordagem de duas etapas mimetiza uma carga de trabalho bastante realista. Ela também simula variações realistas em termos de como os usuários reais e as cargas de trabalho do Processador em segundo plano (sobretudo extrações de dados e notificações de usuário) podem usar o sistema. Nesse contexto, as variações incluem o tempo que os usuários esperam entre interações com a visualização ou o número de tarefas do Processador em segundo plano sendo executadas dentro de um cronograma.

Em nossos testes, modelamos as variações a fim de simular condições reais de produção. Observamos que o sistema escalonou de forma linear quando adicionamos mais nós de trabalho a um cluster de servidores do Tableau Server. Nossos experimentos colocaram o servidor em condições de carga máxima que excedem o que observamos no ambiente de produção. O uso do ambiente de produção geralmente apresenta picos curtos frequentes ao longo do dia de trabalho. Medimos a carga usando a taxa de transferência como critério. A taxa de transferência é o volume de trabalho que o servidor processa durante determinado período de tempo ou, em termos mais simples: transações por segundo. Como mostrado abaixo, observamos que a taxa de transferência escalonou de 4 para 18 transações por segundo em nossos testes. Cada barra indica um teste realizado na topologia descrita no cabeçalho de coluna (Servidor individual, Primário + 1 nó de trabalho etc.).

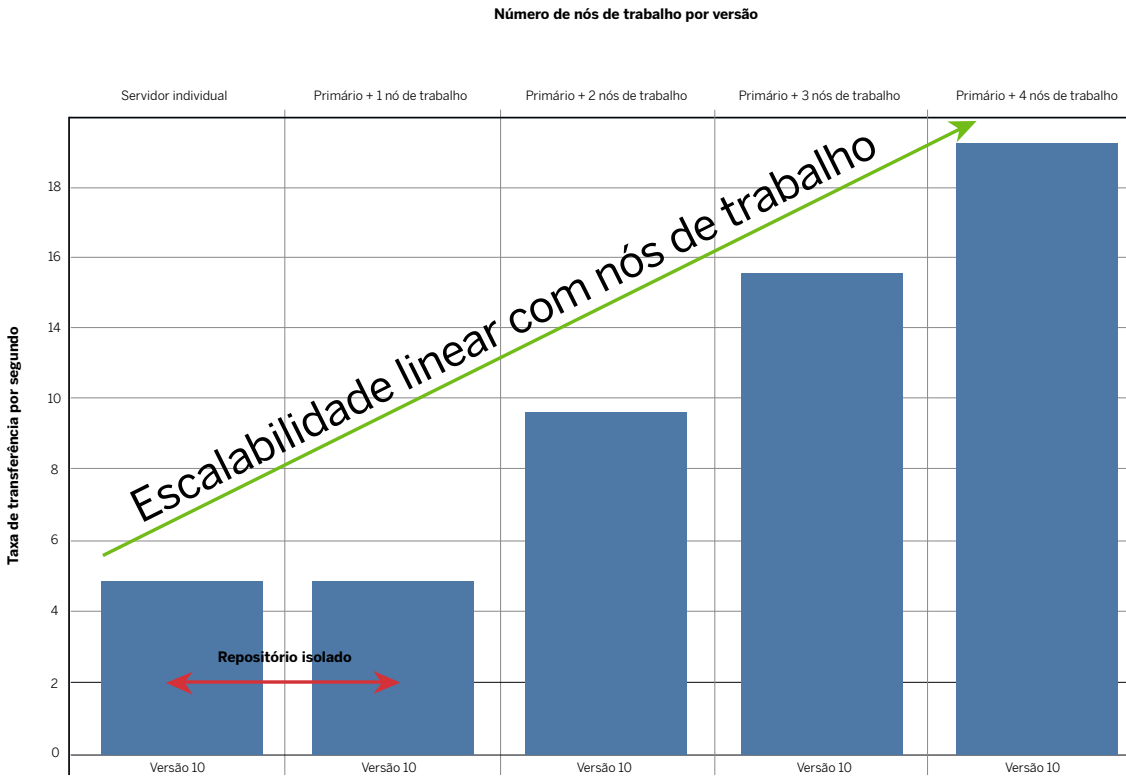


Figura 1: Taxa de transferência por segundo do cenário

Identificamos o número máximo de cargas de usuário ativo que puderam ser acomodadas em um único servidor com oito núcleos físicos. Uma carga de usuário ativo consiste em um conjunto automatizado de operações sendo processadas pelo Tableau Server. Mais adiante neste whitepaper, descrevemos em detalhes quais operações estão incluídas em uma carga de usuário ativo. Depois que medimos o número máximo de usuários ativos que um Tableau Server com oito núcleos físicos foi capaz de suportar, usamos esse número de usuários ativos como unidade de escalabilidade para testar se é possível aumentar linearmente a carga adicionando nós de trabalho uniformes de oito núcleos ao cluster.

O resultado: observamos que um único servidor de oito núcleos é capaz de dar suporte a até 112 usuários ativos operando nos picos de carga realistas modelados de forma prolongada. Quando escalonamos essa referência de forma linear, adicionando quatro nós de trabalho homogêneos com oito núcleos cada e um nó controlador primário somente com o repositório e a instalação de base, foi possível dar suporte a 448 usuários ativos. No caso, estamos falando de usuários simultâneos realizando tarefas no sistema, portanto, fazendo a extrapolação de que nem todos os usuários estarão usando o sistema sempre, pode-se esperar que seja possível dar suporte a entre 500 e 10.000 usuários com uma configuração de cluster de 8 a 32 núcleos. Como veremos, essa extrapolação depende de diversos parâmetros. Considerando que todas as outras variáveis são iguais, o dimensionamento dos servidores no contexto do número de usuários dependerá do uso da análise e das condições dos dados. O que nossos testes mostraram é que a arquitetura do Tableau permite que você continue escalonando linearmente sua base de usuários por meio da adição de mais nós de trabalho ao cluster de servidores do Tableau Server.

Em nossos testes, pressionamos o servidor ao limite para descobrir em que ponto ele começava a apresentar erros ou tempo de inatividade. Esse teste de pressão revelou um limite máximo que define os limites de escalabilidade e desempenho do servidor. Ao quantificar esses limites, continuamos testando o modelo com picos de carga prolongados que excedem a carga-limite de produção caracterizada. Portanto, na prática,

em que uma situação mais estável é o padrão, a plataforma deve comportar mais uso do que o sugerido neste documento. Você pode ter a certeza de que a arquitetura do Tableau Server é escalonável o suficiente para atender às suas necessidades atuais e futuras à medida que elas crescem linearmente.

Cada uma das topologias e dos casos de uso são descritos na tabela a seguir. A coluna “Perfil de risco” representa variáveis como, por exemplo, se a topologia pode estar exposta a maior ou menos riscos de acordo com vários fatores, com failover automático, exposição a falha de hardware e tolerância disponível para picos de carga.

Implantação Configuração	Caso de uso	Perfil de risco
Servidor individual	Implantação simples de um único servidor	Alto risco para disponibilidade, alta concorrência de recursos, sem tolerância a pico de carga problemas de desempenho
1 primário + 1 nó de trabalho	Implantação de dois nós	Alto risco para disponibilidade, menor concorrência de recursos com o repositório, sem tolerância para pico de uso
1 primário + 2 nós de trabalho	Implantação de três nós	Risco moderado para disponibilidade, menor concorrência de recursos, escalabilidade horizontal otimizada, tolerância mínima para pico de uso
1 primário + 3 nós de trabalho	Implantação de quatro nós	Risco moderado para disponibilidade, menor concorrência de recursos, escalabilidade otimizada, tolerância moderada para pico de uso
1 primário + 4 nós de trabalho	Implantação de cinco nós	Risco menor para disponibilidade, escalabilidade otimizada, tolerância disponível para pico de uso ¹

Tabela 1: Topologias de implantação e casos de uso

¹A tolerância disponível depende de vários fatores e não é garantida

É importante entender esses números no contexto da metodologia e dos testes. Para isso, o restante do whitepaper concentra-se na arquitetura e no que torna a arquitetura do Tableau especial em relação às tecnologias tradicionais de BI, bem como na nossa metodologia e na análise dos resultados de escalabilidade.

O VizQL quebra um paradigma ultrapassado

Se você está acostumado com as soluções tradicionais de BI ou se está apenas começando a usar Tableau, pode ser útil entender algumas diferenças importantes entre o funcionamento do Tableau e do BI

tradicional. Nós contrariamos o pressuposto fundamental da abordagem de “consulte primeiro, visualize em seguida” do BI tradicional. Em vez disso, acreditamos que os dados revelam informações quando explorados e questionados em tempo real. Há mais de uma década, o Tableau inclui uma tecnologia patenteada chamada VizQL™, que une consulta e visualização em uma única plataforma. Essa eficiente combinação das funcionalidades permite que os usuários façam perguntas ilimitadas aos dados: consultando, filtrando e analisando ao mesmo tempo em que visualizam os dados. O VizQL™ é uma linguagem fundamental no Tableau que expressa as perguntas e ações dos usuários, traduzindo-as em consultas que podem ser feitas a qualquer conjunto de dados na empresa ou na nuvem. Essa tecnologia se desenvolveu ao longo de mais de uma década de investimentos em engenharia e continua evoluindo para possibilitar a próxima geração de fontes de dados e requisitos de análise.

Diferentemente dos relatórios de BI tradicionais, que são criados e desenvolvidos para atender a um número estático e predefinido de necessidades, as visualizações do Tableau são feitas para proporcionar interatividade e colaboração. Os usuários podem fazer qualquer pergunta aos seus dados sem a necessidade de escrever uniões ou consultas SQL complexas. Eles não precisam esperar outro ciclo inteiro do processo tradicional de desenvolvimento de software para que suas perguntas sejam respondidas. Em vez disso, podem manipular as visualizações existentes de forma iterativa e continuar sua análise para responder a perguntas sobre seus negócios ou projetos.

Com o BI tradicional, você talvez esteja acostumado com relatórios estáticos de testes de carga que atendem a um SLA (contrato de nível de serviço) específico, em que as consultas são projetadas para serem feitas nos sistemas de destino com otimizações predefinidas. Os relatórios estáticos têm escopos fixos, números limitados de consultas e geralmente são otimizados um a um por um desenvolvedor, ao longo de várias semanas. Embora seja relativamente fácil definir e configurar um SLA para um relatório estático, modificar esse relatório redefine todo o ciclo de desenvolvimento, uma limitação que geralmente não é levada em consideração no SLA.

Já as visualizações do Tableau geram e enviam novas consultas em resposta às ações analíticas do usuário. As otimizações e o cache no navegador local do VizQL, que aceleram o acesso aos dados, podem ajudar o usuário a continuar no fluxo de análise, sem que ele precise esperar pelos resultados de uma consulta.

O VizQL como linguagem faz parte do processo do nosso servidor do VizQL e funciona em conjunto com muitos outros processos de servidor distribuídos pra oferecer uma plataforma de análise empresarial escalonável, disponível, segura e confiável.

Arquitetura do Tableau

A arquitetura flexível do Tableau foi projetada para oferecer escalabilidade tanto vertical como horizontal. Ela permite que o Tableau seja executado como a plataforma padrão da empresa ou como uma plataforma para potencializar a análise na nuvem. O Tableau Server é capaz de atender às necessidades das mais complexas infraestruturas de produção das empresas, bem como de implantações simples de grupos de trabalho individuais ou de departamentos.

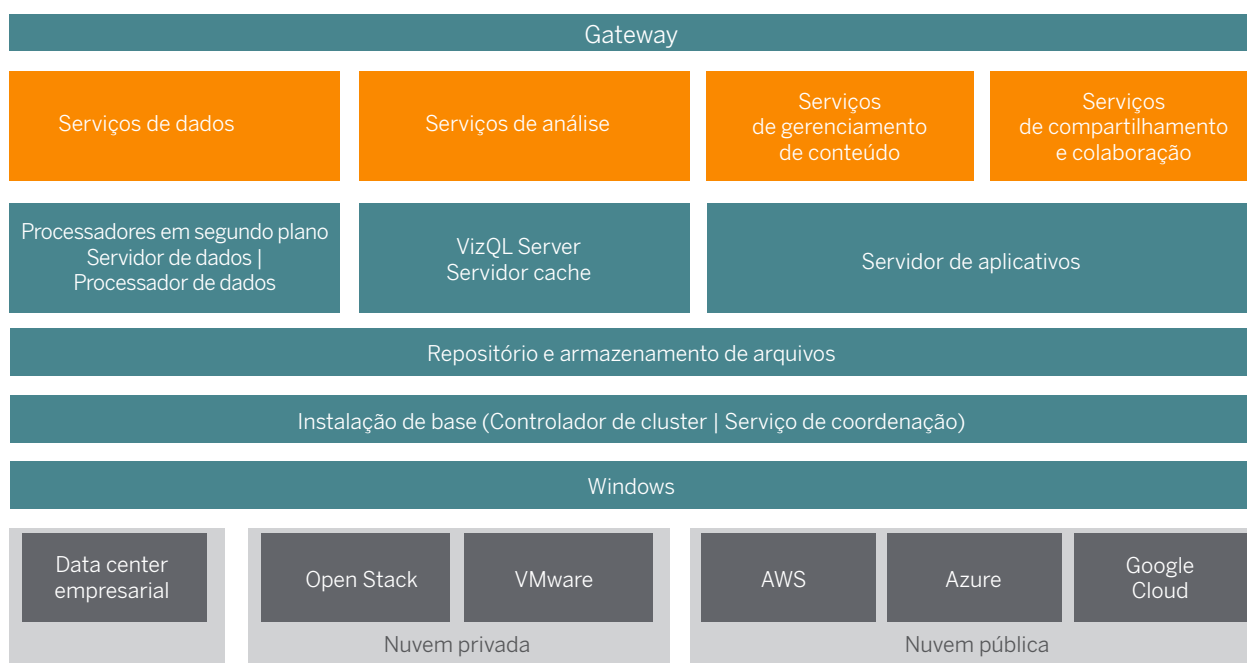


Figura 2: Arquitetura do Tableau Server

O Tableau Server segue um processo simples de instalação e configuração. Depois de instalado, vários processos de servidor (mostrados em azul na Figura 2) funcionam em conjunto para fornecer serviços em várias camadas.

O processo Gateway é o componente que redireciona o tráfego de todos os clientes do Tableau para os nós de servidor disponíveis no cluster.

Os Serviços de dados são um agrupamento lógico de serviços responsáveis pela atualização de dados, gerenciamento de metadados personalizados, fontes de dados governadas e dados em processamento na memória. Os processos subjacentes que alimentam os Serviços de dados são o Processador em segundo plano, o Servidor de dados e o Processador de dados.

Os Serviços analíticos, formados pelos processos do VizQL e do Servidor cache, oferecem serviços de visualização e análise para o usuário, bem como serviços de cache.

O Gerenciamento de conteúdo e o Serviço de compartilhamento e colaboração são alimentados pelo processo do Servidor de aplicativos. As funcionalidades principais do Tableau Server, como login do usuário, gerenciamento de conteúdo (projetos, sites, permissões de segurança etc.), e as atividades de administração são providenciadas pelo processo do Servidor de aplicativos.

Todos os serviços mencionados acima usam e dependem do processo do Repositório, que contém dados relacionais estruturados, como permissões, pastas de trabalho, extrações de dados, informações dos usuários e metadados. O processo de Armazenamento de arquivos permite a redundância de arquivos de extração de dados em todo o cluster e garante que as extrações de dados estejam disponíveis localmente em todos os nós do cluster. Em caso de cargas mais pesadas, os arquivos de extração ficam disponíveis localmente em todo o cluster para oferecer processamento e renderização mais rápidos.

A arquitetura do Tableau é flexível, permitindo executar a plataforma em qualquer lugar. Você pode instalar o Tableau Server localmente, em seu data center ou sua nuvem privada, no Amazon EC2™, no Google Cloud Platform™ ou no Microsoft Azure™. A plataforma de análise do Tableau também pode ser executada em plataformas de virtualização como o VMware ESXi™ ou o Microsoft Hyper-V™. Recomendamos que você siga as práticas recomendadas de cada plataforma de virtualização para garantir o melhor desempenho do Tableau Server.

Para obter detalhes sobre cada processo de servidor específico, consulte o [guia de administração](#) do Tableau Server.

Arquitetura unificada em tempo real e de processamento na memória

O Tableau oferece um sólido suporte a diversas plataformas, conectando-se a mais de 50 das mais populares fontes de dados de dezenas de fornecedores. A arquitetura oferece flexibilidade para você escolher a análise rápida processada na memória ou conduzir análises conectando-se diretamente aos repositórios de dados em tempo real. Independentemente da opção escolhida, também é fácil alternar entre a conectividade em tempo real e a análise na memória para atender a necessidades comerciais e fontes de dados variáveis.

O Tableau permite a análise processada na memória extraindo os dados em um armazenamento de colunas exclusivo chamado TDE (Extração de dados do Tableau), que é carregado em arquivos mapeados na memória para proporcionar acesso rápido. As extrações podem ser criadas pelos usuários no Tableau Desktop ou com processos comerciais externos por meio das APIs do Tableau. A arquitetura do Tableau Server oferece suporte integrado para garantir que as extrações criadas pelos seus usuários sejam atualizadas com uma frequência otimizada controlada por você.

Independentemente de como você se conecta aos dados e os integra ao Tableau Server, é necessário ter recursos de memória suficientes para a análise de dados, cache, atualizações de extrações e outras operações relacionadas.

Diferentemente das ferramentas que são processadas puramente na memória, o consumo de memória do Tableau engloba todo o cluster e não se acumula em um único servidor. Em vez disso, o uso de memória é distribuído entre todo o cluster por meio dos servidores de cache, oferecendo um cache de consultas compartilhado externamente, bem como outros processos que armazenam caches no nível da sessão. Assim, o impacto do aumento das cargas na memória é distribuído em todo o cluster de acordo com suas cargas de trabalho.

Metodologia e abordagem dos testes

Ao contrário da maioria dos projetos tradicionais de testes de carga, em que o aplicativo de destino é tratado como uma caixa-preta, o teste de carga do Tableau Server requer conhecimentos suficientes de sua arquitetura. Projetamos o Tableau Server para ser executado em qualquer lugar: localmente ou na nuvem. Ele foi desenvolvido para contemplar equipes e organizações de todos os tamanhos. Portanto, a instalação padrão do Tableau Server deve funcionar bem para a maioria das implantações padrão. Antes de começar a escalar e implantar o Tableau Server em toda a sua empresa, você deve entender como as diferentes cargas de trabalho são processadas e como alguns poucos ajustes de configuração podem melhorar os resultados para seu cenário de implantação.

Para as metas de escalabilidade da versão 10, tratamos de responder às seguintes perguntas:

1. O Tableau Server 10 é escalonado linearmente por meio da adição de mais hardware para dois cenários comuns:
 - Escalabilidade da carga de trabalho dos usuários finais
 - Escalabilidade da carga de trabalho do Processador em segundo plano
2. Qual é o momento certo de migrar os processos do Processador em segundo plano para um nó de trabalho dedicado?
3. Como o desempenho do Tableau Server 10 se compara às versões anteriores?

As versões anteriores deste whitepaper priorizavam especificamente a escalabilidade dos usuários finais, mas muitos clientes nos perguntaram como escalar a carga de trabalho do Processador em segundo plano. As cargas de trabalho do Processador em segundo plano controlam a atualidade dos dados (atualizações de extrações) e a amplitude do consumo da análise (notificações de assinaturas).

Metodologia

Melhorar o desempenho e a escalabilidade eram os objetivos principais da versão 10 do Tableau Server. Conseqüentemente, o desenvolvimento de uma metodologia de teste de categoria empresarial ideal para produção também era uma necessidade essencial. Começamos a implementar a metodologia descrita nesta seção em versões pré-lançamento da versão 10. Durante o processo iterativo de desenvolvimento ágil, descobrimos e corrigimos mais de 20 erros específicos de escalabilidade e desempenho revelados por nossa metodologia. Continuamos os testes após o lançamento da versão 10 e corrigimos erros com a versão de manutenção 10.0.1. Este whitepaper geralmente se refere à versão 10 do Tableau Server. No entanto, os resultados dos testes baseiam-se na versão de manutenção 10.0.1. Para obter os melhores benefícios de escalabilidade e desempenho da versão 10, execute a versão de manutenção 10.0.1 ou posterior na sua organização.

Otimizamos continuamente nossas práticas de escalabilidade para coletar e testar cargas de trabalho que representam cenários reais dos clientes. Embora haja muitas variáveis que determinam a escalabilidade de uma implantação, estes são os fatores importantes que você deve considerar ao planejar sua implantação:

- Impacto para o usuário – uso de autoatendimento e adoção do usuário: Quantos usuários utilizarão a análise? Com que frequência os usuários empregarão a análise para tomar decisões embasadas? Qual é o nível de complexidade das visualizações que os usuários criarão?
- Impacto dos dados – atualização, volume e local: Qual é o volume dos seus dados? Onde os dados estão localizados? Quão atualizados os dados precisam estar para fundamentarem as decisões comerciais com precisão?

Uso da análise para tomar decisões comerciais eficazes	Alta (a cada segundo)	7. Exemplos: Exploração mundial de dados Tableau Public (eleição presidencial dos EUA) 30 mil exibições/hora	8. Exemplos: Painel de meta de vendas, Tableau na TV	9. Exemplos: Controlador de tráfego aéreo, Monitoramento financeiro, Comercialização
	Moderada (uma vez por hora)	4. Exemplos: Estoque diário de loja Cliente de seguradora Marketing analítico (segmentação)	5. Exemplos: Capacidade de pacientes Gestão de concessionárias	6. Exemplos: Painel de encaminhamento de suporte, painel de portfólio financeiro Investigação de fraudes
	Baixa (uma vez por dia)	1. Exemplos: Engenharia - Quartos de navios Financiamento imobiliário Estoque BI tradicional	2. Exemplos: Personalidades em evidência Rastreamento de clientes em potencial	3. Exemplos: Painéis on-line de tráfego em rodovias
	Baixa (uma vez por dia)	Moderada (uma vez por hora)	Alta (a cada segundo)	
Frequência de atualização dos dados para tomar decisões comerciais eficazes				

Figura 3: Matriz de uso da análise e frequência de atualização dos dados

Para testar essas duas considerações, precisamos incorporar testes em que aumentamos gradualmente o número de usuários finais de um servidor e, ao mesmo tempo, aumentamos as cargas de trabalho do Processador em segundo plano. Essa metodologia nos permitiu avaliar o impacto das cargas de trabalho do Processador em segundo plano na qualidade de serviço para o usuário final. Para modelar o uso baixo, moderado e pesado da análise e o impacto do isolamento das cargas de trabalho do Processador em segundo plano, executamos mais de 400 iterações de testes em nossos laboratórios com várias cargas de trabalho. Examinamos a escalabilidade do sistema e também corrigimos erros que só se manifestaram em condições de carga pesada.

Para que os testes fossem aplicáveis a situações reais, primeiro tivemos que coletar um conjunto adequado de cargas de trabalho para executá-los.

Caracterização das cargas de trabalho reais

Para a versão 10, observamos e caracterizamos como um Tableau Server de produção real era usado durante períodos com pico de utilização.

Para determinar quais cargas de trabalho modelar e as características das cargas de trabalho usadas nos testes, analisamos os arquivos de registro do Tableau Server de um ambiente de produção com 3.000 usuários. Identificamos visualizações e cargas de trabalho usadas de forma intensiva. Calculamos a distribuição de uso entre essas pastas de trabalho e depois analisamos as características de uso. Por exemplo, investigamos o intervalo de tempo entre as solicitações (também conhecido como tempo de pensamento). Abaixo estão as etapas específicas que executamos para modelar a carga de trabalho do servidor de produção.

Etapas de modelagem dos testes

1. Obter os registros do servidor de produção correspondentes ao período com pico de uso
2. Identificar as N principais exibições de pastas de trabalho de acordo com a média ponderada do tempo ocupado no servidor.
3. Média ponderada = Tempo de resposta médio * Número de solicitações
4. Calcular os pesos relativos entre N principais exibições de pastas de trabalho
5. Para cada exibição de pasta de trabalho selecionada
 - a. Encontrar a porcentagem das cargas de visualização com: atualização=y (essa foi uma forma de descobrir quantos usuários estavam atualizando ativamente seus dados para obter os dados mais recentes)
 - b. Encontrar as N principais interações por peso
 - c. Calcular o tempo de pensamento médio entre interações
6. Para a confirmação do modelo:
 - a. Determinar o tempo médio entre inicializações (tempo entre testes) para a exibição de pasta de trabalho
 - b. Determinar as N principais tarefas do Processador em segundo plano pela média ponderada de tempo ocupado no servidor
7. Categorizar o tráfego do Processador em segundo plano por diferentes tarefas (por exemplo, assinaturas e atualizações de extrações)
 - a. Determinar o tempo médio entre as principais tarefas do Processador em segundo plano e inclui-lo no modelo
 - b. Determinar o número de assinaturas em vários agendamentos
 - c. Determinar o tamanho e o tipo das atualizações de extrações (extração publicada no servidor de dados, extração de pasta de trabalho)

Usamos todos os dados acima para modelar uma combinação realista de cargas de trabalho que representasse como os usuários reais estavam utilizando o servidor de produção durante o pico de uso. Por fim, usamos nossas análises dos registros de produção para criar um modelo com base na carga de trabalho para assinaturas e atualizações de extrações do Processador em segundo plano.

As combinações de cargas de trabalho estão resumidas na tabela a seguir.

Nome da carga de trabalho	Descrição	Como comparar com as versões anteriores
Carga de trabalho de servidor de produção real	Esta carga de trabalho baseou-se na análise e na caracterização do uso de um Tableau Server de produção que atende a 3.000 usuários de uma organização como um aplicativo essencial gerenciado pela TI.	Os dados dos resultados somente deste whitepaper são comparáveis entre as versões 9.3 e 10.0 como especificado aqui. Os resultados não são comparáveis com os whitepapers anteriores que publicamos (incluindo o whitepaper da versão 9.0) pois a metodologia de teste era consideravelmente diferente.
Carga de trabalho de servidor de produção real + novos recursos	Esta carga de trabalho combina a carga de trabalho acima com pastas de trabalho que usaram os novos recursos da versão 10.	Não é possível comparar as versões 9.3 e 10.0 porque a versão 9.3 não tem os recursos da versão 10 para serem testados. A melhor aplicação dos resultados deste whitepaper devem ser para fundamentar as características de escalabilidade da versão 10.
Combinação de Processadores em segundo plano	Esta carga de trabalho baseia-se na análise de cargas de trabalho de produção e modela cargas de trabalho e programações reais que refletem o ambiente de produção.	A nova combinação introduzida no whitepaper de testes de escalabilidade da versão 10.0 não é comparável aos resultados de nenhum whitepaper anterior.

Tabela 2: Descrições das combinações de cargas de trabalho

Em seguida, executamos cada uma dessas combinações em um laboratório de escalabilidade isolado em máquinas físicas com cargas crescentes de usuários finais e do Processador em segundo plano. Quando o cluster atingiu a capacidade máxima, continuamos aumentando a carga após adicionar um nó de trabalho por vez. Observamos como o sistema se comportou durante cada um desses testes. Em cada execução, registramos os indicadores-chave de desempenho, como tempo de resposta, taxa de transferência e taxas de erros. Também registramos métricas de sistema e métricas do servidor de aplicativos usando JMX. Correlacionamos os dados e analisamos o comportamento do sistema com cargas de trabalho cada vez maiores. Durante esse processo, também encontramos e corrigimos problemas de escalabilidade como parte do nosso processo de desenvolvimento ágil.

Metodologia do Processador em segundo plano

O servidor do Processador em segundo plano processa tarefas em segundo plano nos níveis do sistema e do usuário. As tarefas no nível do sistema, como tarefas rotineiras de manutenção de repositório, são executadas pelo Processador em segundo plano. As tarefas no nível do usuário são aquelas que um usuário pode ter enviado para o sistema executar. Por exemplo, os usuários podem publicar extrações no servidor e, em seguida, configurar uma atualização de dados regular para a extração de acordo com uma agenda. Esse conjunto de operações cria uma tarefa de atualização. O Processador em segundo plano é o processo que analisará a lista de tarefas e as executará. Este cenário é fundamental para a eficácia do autoatendimento, pois o usuário não precisa esperar o departamento administrativo atualizar os dados. No entanto, se a equipe de administração que gerencia o Tableau Server não planejar a capacidade para esse tipo de carga, poderá haver problemas de qualidade de serviço para o usuário final. A otimização

adequada aos serviços do Processador em segundo plano é um componente fundamental do processo de dimensionamento e planejamento dos servidores. Você deve considerar se é melhor separar os serviços em segundo plano em outro computador para isolar as cargas de trabalho.

Há várias práticas recomendadas simples abordadas no fim deste documento que permitem separar o processamento das cargas de trabalho com base na hora do dia. No entanto, se você estiver executando o Processador em segundo plano no mesmo computador que os Serviços de análise, poderá observar um impacto na qualidade do serviço para o usuário final quando em condições de carga pesada devido às restrições de compartilhamento de recursos do servidor.

Por esse motivo, quisemos avaliar o impacto do Processador em segundo plano na escalabilidade de usuários finais quando ele é colocado no mesmo computador que os Serviços analíticos. Além disso, quisemos quantificar a escalabilidade do Processador em segundo plano com o aumento das cargas de trabalho quando ele foi isolado em um hardware dedicado.

Para avaliar isso, usamos um computador com quatro núcleos físicos para executar o Processador em segundo plano em isolamento. Não executamos nenhum outro processo do Tableau Server no mesmo computador. Executamos a mesma carga de trabalho de produção modelada nas versões 9.3 e 10.0 do Tableau Server. A carga de trabalho incluía atualizações de extrações e assinaturas para que pudéssemos nos concentrar nas tarefas no nível do usuário. A carga de trabalho incluía 400 assinaturas com oito agendamentos. Avaliamos o sucesso das notificações de assinaturas e também o tempo que o Tableau levou para concluir todas as assinaturas.

Ambiente isolado padronizado

Realizamos os testes de escalabilidade em nosso laboratório de desempenho utilizando máquinas físicas idênticas com as seguintes especificações.

Tipo de servidor	Dell PowerEdge R620
Sistema operacional	Microsoft Windows Server 2012 R2 Standard de 64 bits
CPU	2,6 GHz, 1x8 núcleos físicos, hyper-threading ativado (16 núcleos lógicos)
Memória	64 GB

Tabela 3: Especificação de hardware de cada nó no ambiente de teste

Embora esta tabela liste os núcleos físicos, recomendamos que você não desative o hyper-threading (hiperprocessamento). Para fins de consistência, com exceção desta referência específica, neste whitepaper nos referimos aos números de núcleos físicos e sempre pressupomos que o hyper-threading está ativado nas máquinas físicas.

Topologia de implantação

Os clusters são compostos por um ou mais nós primários (controlador) e um ou mais nós de trabalho. Em nossos testes, os nós de trabalho compartilhavam o mesmo perfil de configuração de processos:

Processo	Primário tsperf-212.perf.dev.tsi.lan	Nó de trabalho 1 tsperf-213.perf.dev.tsi.lan	Nó de trabalho 2 tsperf-215.perf.dev.tsi.lan	Nó de trabalho 3 tsperf-216.perf.dev.tsi.lan	Nó de trabalho 4 tsperf-217.perf.dev.tsi.lan
Controlador de cluster	✓	✓	✓	✓	✓
Gateway	✓	✓	✓	✓	✓
Servidor de aplicativos		✓	✓	✓	✓
VizQL Server		✓ ✓	✓ ✓	✓ ✓	✓ ✓
Servidor cache		✓ ✓	✓ ✓	✓ ✓	✓ ✓
Pesquisar e procurar		✓	✓	✓	✓
Processador em segundo plano		✓	✓	✓	✓
Servidor de dados		✓ ✓	✓ ✓	✓ ✓	✓ ✓
Processador de dados		✓	✓	✓	✓
Armazenamento de arquivos		🔄 Sincronizando	🔄 Sincronizando	🔄 Sincronizando	🔄 Sincronizando
Repositório	✓				
<input type="button" value="Atualizar status"/> ✓ Ativo 🔄 Ocupado ✓ Passivo ⚠ Não licenciado ✖ Inativo □ Status indisponível					

Figura 4: Configuração de processos

A configuração dos processos de trabalho mostrada aqui é a configuração padrão. A escalabilidade pode ser maior ou menor dependendo do número e dos tipos de processos que você configura para seu ambiente e cenário de uso.

Os nós primários são configurados com uma instalação de base que inclui os processos de Controlador de cluster, Gateway e Repositório. Vale a pena observar que, quando implantado com um modelo de licenciamento por núcleos, esse tipo de configuração de nó primário não está incluído na contagem de núcleos.

Aumentamos a carga de trabalho usando geradores de carga (TabJolt) para simular a carga de trabalho de usuários descrita anteriormente. **TabJolt** é uma ferramenta automatizada de teste de carga e desempenho projetada especificamente para funcionar com o Tableau Server 9.0 e versões posteriores. A figura abaixo mostra uma exibição lógica da execução do teste.

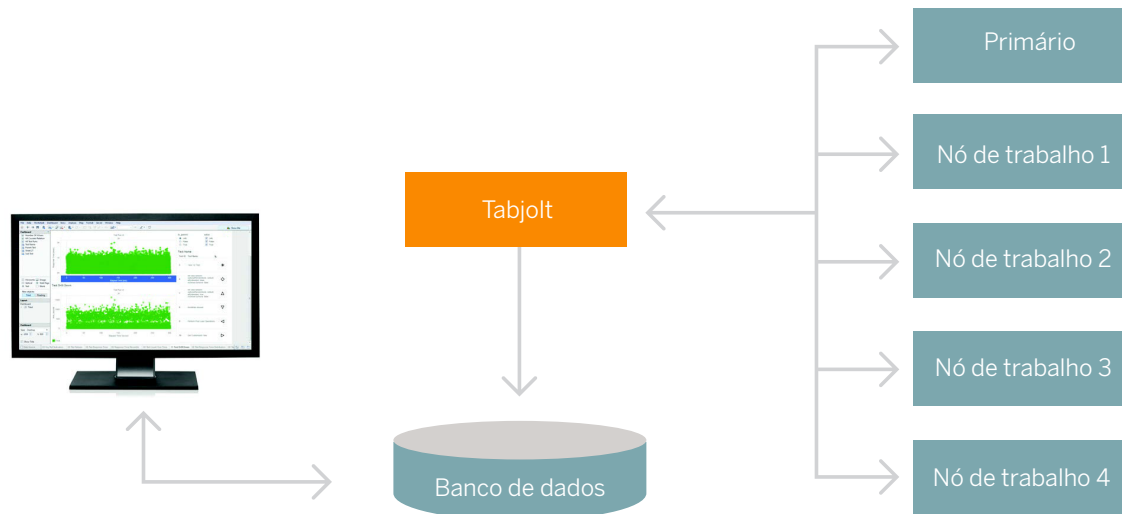


Figura 5: Exibição lógica do ambiente de teste

Coletamos dados de cada uma das iterações de teste para análise, mas, antes de examinar os resultados, vamos entender algumas das métricas e definições.

Medições e relatórios

Medimos várias métricas para entender o desempenho e a escalabilidade do hardware, incluindo métricas de sistema referentes a CPU, memória e disco. Também medimos métricas de desempenho e escalabilidade, como tempos de resposta, taxa de transferência, taxas de erros, duração da execução, entre outras.

Para entender os dados abordados neste whitepaper, vamos examinar rapidamente algumas definições.

Cenário

O cenário é a atividade geral do usuário no servidor. Nas versões anteriores deste documento, nosso foco eram os tempos de interação e carga de visualização quando o servidor tinha o acesso para convidados ativado. Nesta versão do documento, expandimos as cargas de trabalho para incluir os serviços do Servidor de aplicativos (como login) e outros. Os usuários finais executam várias etapas com base na modelagem de carga de trabalho de produção que simulamos usando uma versão personalizada do TabJolt.

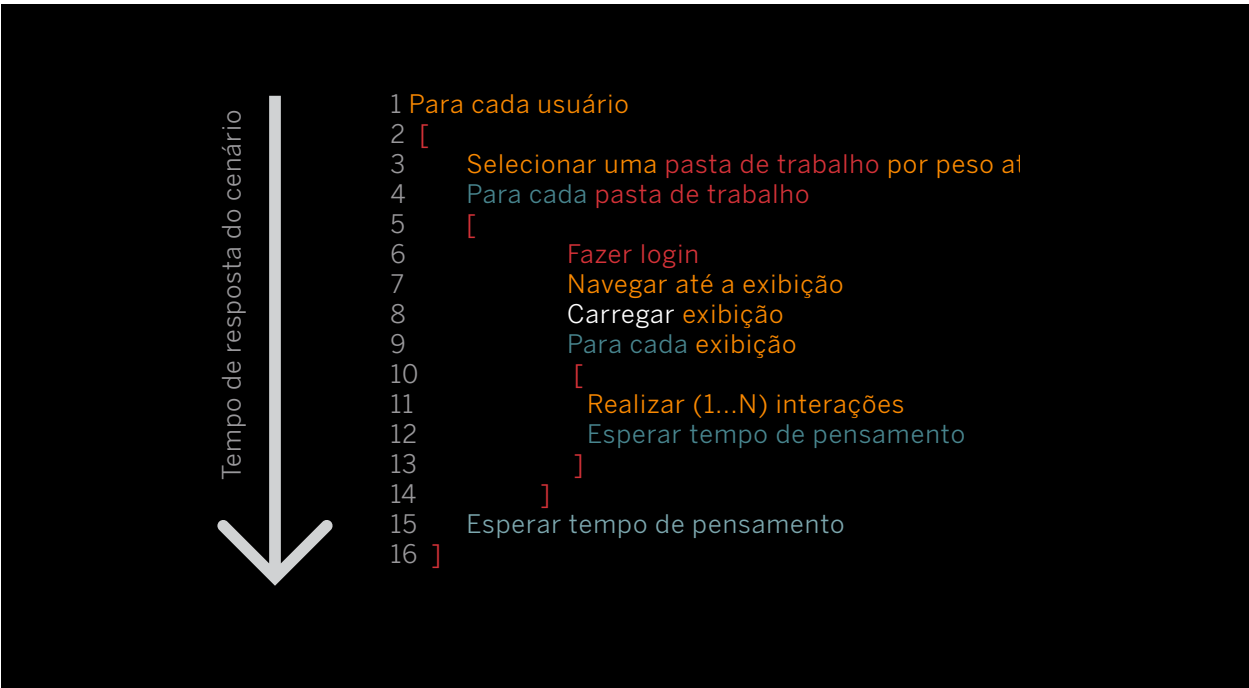


Figura 6: Algoritmo de teste do cenário

Os tempos de resposta informados neste documento são maiores do que modelos de carga e interação mais simples porque nossos modelos de teste incluem o tempo necessário para concluir o cenário. É medido e informado o tempo de resposta de uma transação da perspectiva do cliente. Isso significa que o cenário descreve a perspectiva real do usuário final, incluindo variáveis de rede do ambiente, como a latência.

Tempo de resposta

O tempo de resposta é medido como o número de segundos que o servidor leva para responder à solicitação do usuário final. Considere este exemplo: um usuário faz login no servidor, navega até uma visualização, altera um filtro nela, espera que a visualização seja atualizada e renderizada e a analisa (tempo de pensamento). O tempo total real desde o momento em que o usuário faz login no servidor até o fim do tempo de pensamento do usuário é informado como um tempo de resposta para essa iteração do cenário.

Taxa de transferência do cenário

A taxa de transferência do cenário é o número de cenários bem-sucedidos concluídos por segundo, representado com a sigla TPS (taxa de transferência por segundo). Calculamos a TPS do cenário executando cenários por uma hora em determinada topologia no limite computacional do sistema. O número total de cenários dividido pelo número de segundos em uma hora (3.600) é a TPS.

Por exemplo, a execução de nossos cenários em um Tableau Server 10 com oito núcleos durante o teste completo concluiu 16.372 cenários. Isso se traduz em uma TPS de aproximadamente 4,5 (16.372/3.600). Esses testes colocou o servidor em cargas muito mais altas do que observamos na produção, em que a TPS era inferior a 1.

O aumento significativo nos dados de teste ocorre porque estamos levando o servidor ao limite, na tentativa de entender a máxima capacidade de desempenho para o hardware e as unidades de escalabilidade escolhidos.

No entanto, as implantações de produção apresentam aumentos significativos de carga durante picos de uso, e os usuários geralmente são muito mais lentos do que um sistema de execução de teste automatizado. Atentando a essa diferença, o Tableau Public, por exemplo, processa escalas imensas de aproximadamente 12 cargas de visualização (chamadas de impressões) por segundo, o que se corresponde a mais de 7 milhões de impressões de visualizações do Tableau por semana.

Usuários ativos

A métrica de usuários ativos mede o número de usuários usando o Tableau Server simultaneamente em um período de pico de uma hora. Os cenários agora incluem login, carga de visualização, interações do usuário, pesquisa e outras ações. Definimos usuários ativos como usuários que estão executando qualquer um desses tipos de ação durante um período de pico de uma hora.

Para saber quantos usuários ativos nossa instalação de servidor suportaria, começamos determinando quantos usuários um único servidor poderia comportar sem queda no tempo de resposta, mais de 2% de taxa de erro (erros HTTP do Tableau Server) ou mais de 80% de utilização de CPU.

Com base na carga de trabalho de produção que caracterizamos, atribuímos pesos a pastas de trabalho específicas. Em seguida, geramos um segmento (um usuário virtual) que selecionaria uma pasta de trabalho aleatória com base nos pesos atribuídos a ela e, em seguida, completaria o cenário inteiro descrito anteriormente. No fim do cenário, o segmento aguardaria por um determinado tempo de pensamento. No fim desse período, o segmento concluiria sua iteração e seria encerrado. Durante o teste, medimos a taxa de transferência, o tempo de resposta, as taxas de erro, a CPU e o uso de memória do cenário, entre outras métricas. Continuamos aumentando o número de segmentos ativos enquanto a CPU se mantinha abaixo de 80%, a taxa de erros era inferior a 2% e o tempo de resposta não havia piorado. Quando um desses três valores era atingido, considerávamos esse o limite de usuários ativos que essa topologia poderia comportar razoavelmente.

Para confirmar que o servidor escala de forma linear, identificamos o limite para uma única máquina virtual, depois aumentamos linearmente o número de usuários virtuais para o próximo incremento e confirmamos que nossas condições predefinidas ainda eram atendidas com a nova carga.

Resultados

Agora que já sabemos como os testes foram realizados, o tipo de implantação adotado e as métricas utilizadas, vamos ver os resultados.

A escalabilidade do Tableau Server 10 é linear

Nossa primeira pergunta foi: como é a escalabilidade do Tableau Server 10 na prática? Com o aumento das cargas de usuário, observamos que o Tableau Server 10 escala de forma linear por meio da adição de mais nós de trabalho ao cluster. A figura abaixo mostra o número de cenários de usuário que foram concluídos por segundo com o aumento dos nós de trabalho.

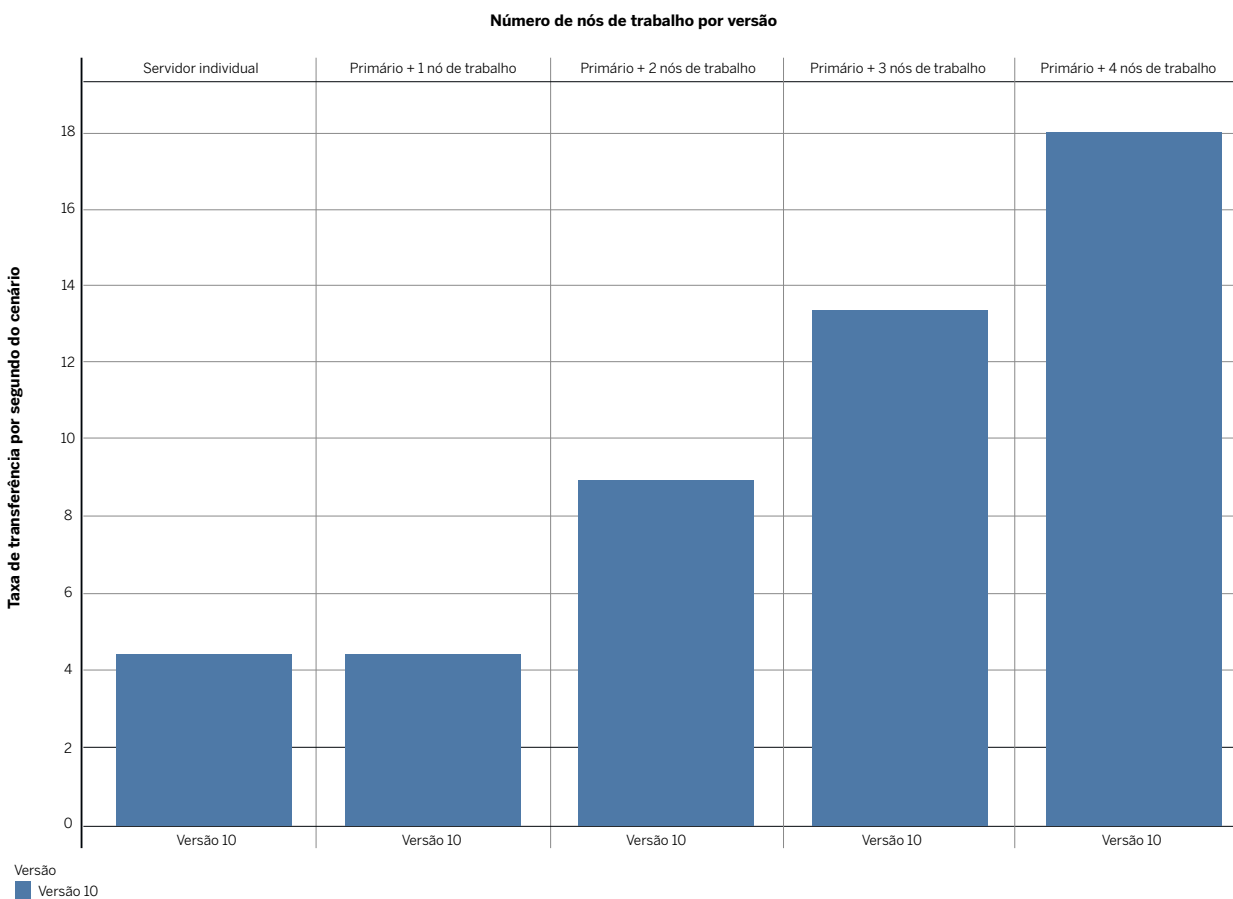


Figura 7: Taxa de transferência por segundo do cenário

Cada coluna mostra a topologia do cluster. A primeira coluna mostra a configuração de um único servidor. A segunda coluna mostra a configuração de cluster de nó primário/de trabalho com um único nó de trabalho. Para cada coluna restante, adicionamos um nó de trabalho, como descrito anteriormente. A altura da barra mostra a taxa de transferência média por segundo do cenário. A TPS representa o volume de trabalho que o servidor está processando. Como mostrado, a TPS aumentou linearmente à medida que adicionamos mais nós de trabalho. Observamos que, à medida que aumentamos as cargas no cluster, a média da utilização de CPU em todo o cluster passou a ser cerca de 80%. A figura abaixo mostra a utilização de CPU em todo o cluster com o aumento das cargas de trabalho, com uma faixa de confiança de 95%.

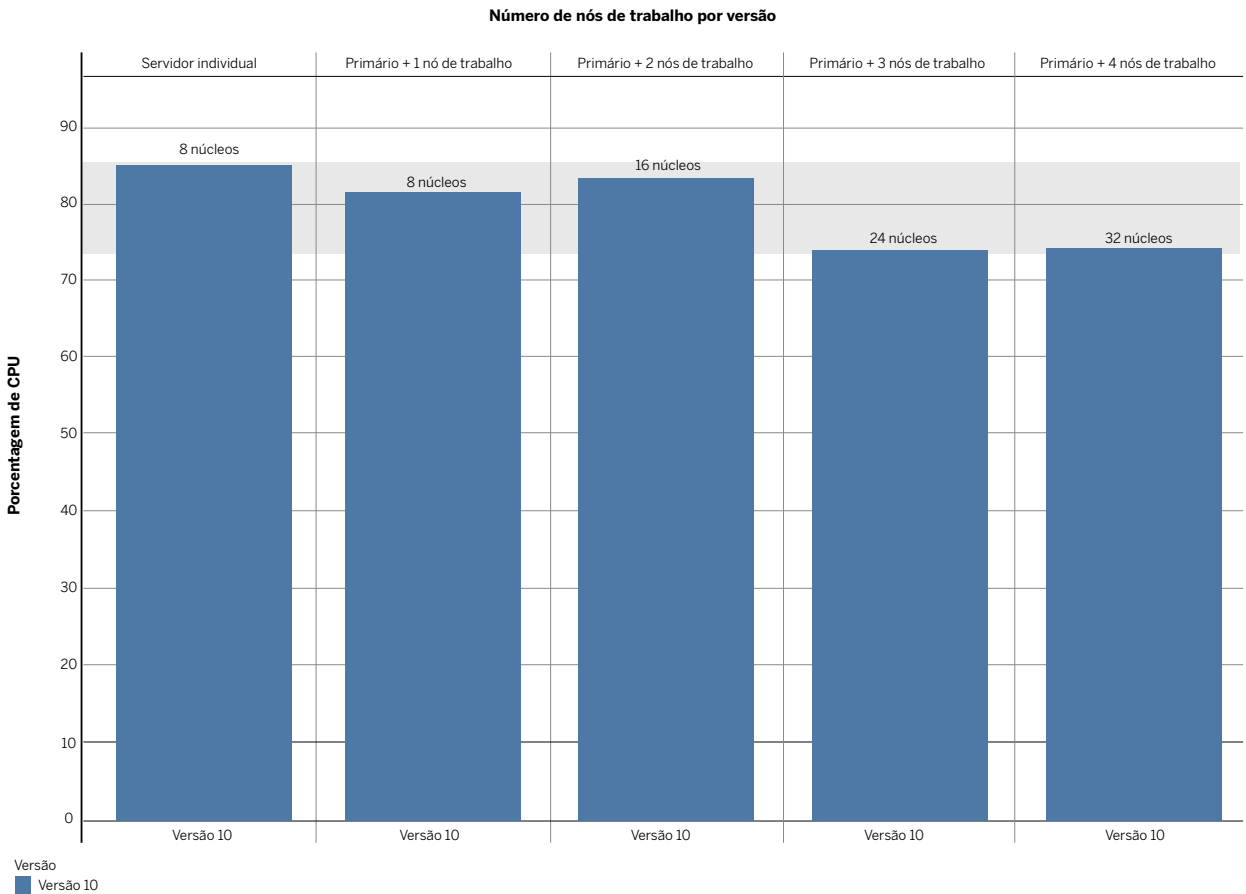


Figura 8: Utilização de CPU com o aumento das cargas em todo o cluster

Como mostra a Figura 8, a adição de mais nós de trabalho para distribuir a carga de CPU por todo o cluster otimiza o sistema criando uma tolerância para picos de carga. Em um cluster em que menos nós de trabalho foram configurados, a utilização de CPU foi comparativamente maior do que em clusters com mais nós de trabalho. Nesses casos de menos nós de trabalho, os processos vinculados à computação competem por recursos limitados. Durante esses testes, observamos as taxas de erro do servidor e elas estavam satisfatoriamente dentro da meta de 2% que havíamos definido como parte da metodologia. Dependendo das suas cargas de trabalho, você pode observar taxas maiores de erros (menor qualidade de serviço) quando os clusters são limitados a menos máquinas e/ou tem menos capacidade.

Nossa próxima pergunta foi como o Tableau Server 10 se compara ao Tableau Server 9.3 com a mesma metodologia de teste?

Os resultados: com a mesma metodologia de teste aplicada a cada versão, a taxa de transferência do Tableau Server 10 foi melhor em relação ao Tableau Server 9.3.

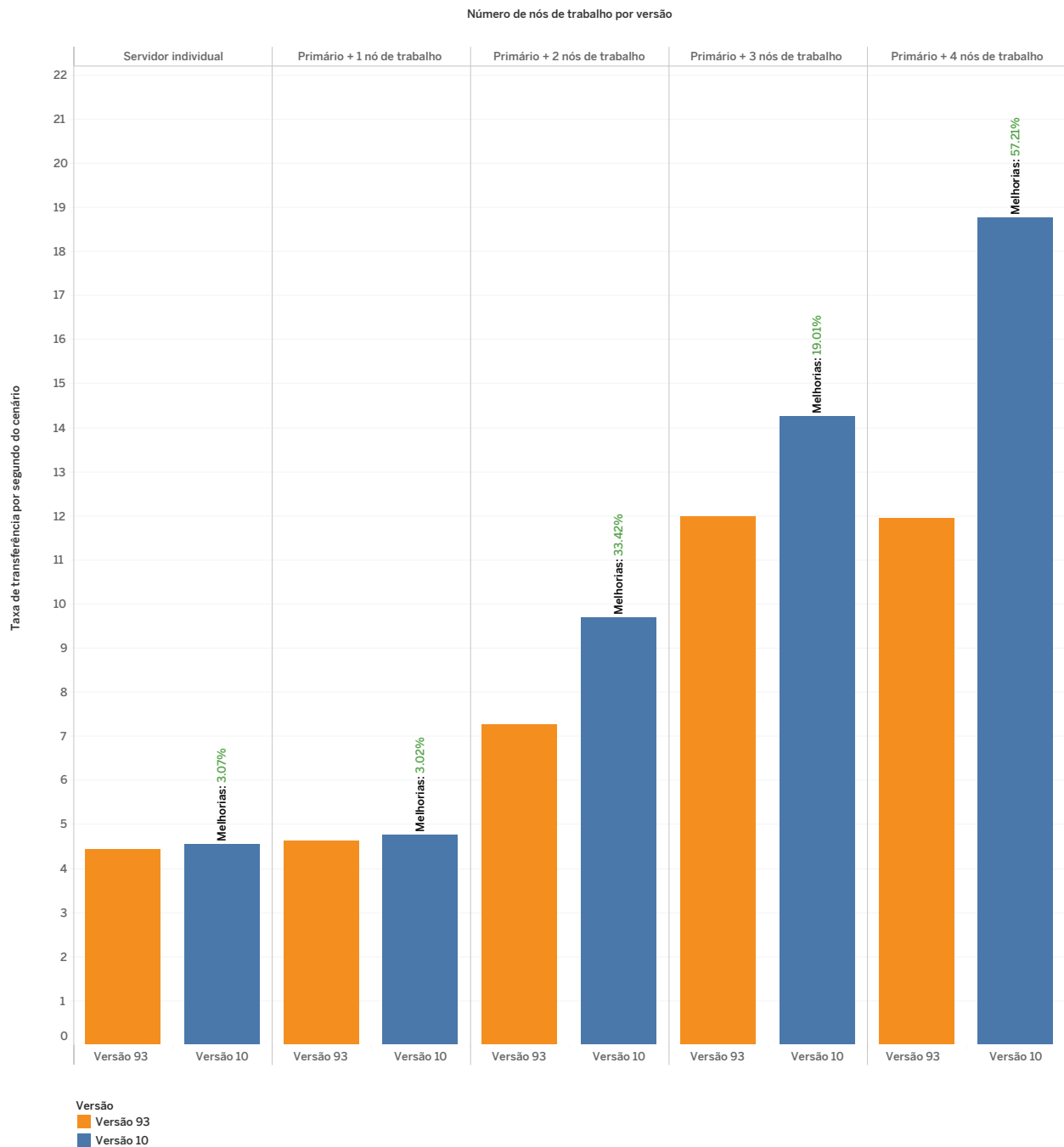


Figura 9: Comparação da taxa de transferência por segundo do cenário entre as versões 9.3 e 10.0.1

Na Figura 9, a taxa de transferência do cenário do Tableau Server 9.3 (barras laranja) é comparada à taxa de transferência do cenário do Tableau Server 10 (barras azuis). Cada painel mostra a configuração da topologia do cluster conforme descrito pelo cabeçalho de coluna. Como observado anteriormente, quando adicionamos mais nós ao cluster, o Tableau Server 10 não só escalonou linearmente, mas também escalonou a taxa de transferência do cenário de melhor forma quando comparado ao Tableau Server 9.3.

Ao comparar as versões, entenda que a única comparação adequada que podemos fazer é comparar as duas versões usando a metodologia de teste otimizada que captura os cenários de usuário reais. Por esse motivo, comparar os resultados do Tableau Server 10 com os resultados do whitepaper anterior não consiste em uma comparação precisa, pois a metodologia de teste mudou significativamente.

Embora seja importante observar a taxa de transferência, queríamos ter certeza de que o tempo de resposta para o usuário final fosse satisfatório durante todo o cenário de teste, sem falhas quando as cargas aumentassem.

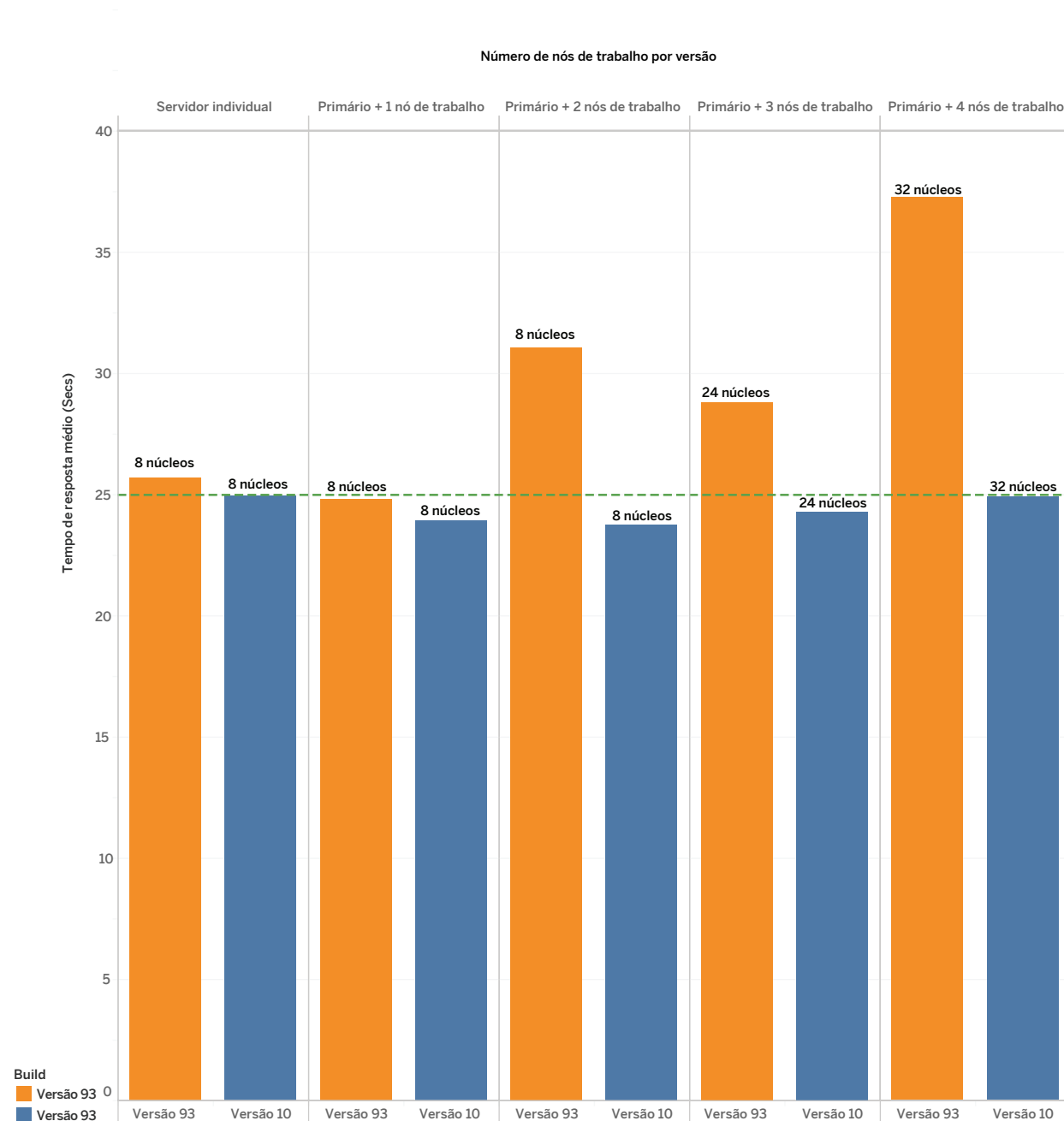


Figura 10: Comparando os tempos de resposta médios do cenário (segundos): 9.3 x 10.0.1

Na Figura 10, a altura da barra laranja mostra o tempo de resposta médio do cenário para a versão 9.3. A altura da barra azul mostra a mesma métrica para a versão 10.0.1. À medida que aumentamos as cargas de usuários finais no cluster que hospedava a versão 9.3, observamos aumentos incrementais nos tempos de resposta. Já o Tableau Server 10 mostrou um tempo de resposta mais consistente à medida que continuamos adicionando usuários virtuais ao sistema. Isso mostra que o Tableau Server 10 oferece desempenho e capacidade de resposta melhores aos usuários finais em relação ao Tableau Server 9.3 para as mesmas cargas de trabalho.

Algumas das melhorias que observamos no Tableau Server 10, com tempos de resposta relativamente consistentes e estáveis com o aumento das cargas, foram o resultado de melhorias que fizemos nos componentes de cache. Especificamente, trabalhamos na versão 10 para otimizar o cache no cenário de resposta de inicialização. O cenário de inicialização é a primeira chamada feita para inicializar e armazenar em cache os dados para uma sessão de usuário. Sem cache, a versão anterior do Tableau Server computava e salvava os dados de inicialização para cada solicitação subsequente, mesmo que eles fossem muito semelhantes ou idênticos. Na versão 10, realizamos operações de cache inteligentes para esse cenário. Essas melhorias aumentaram a eficiência do Tableau Server 10, permitindo que ele processe uma carga maior de usuários sem prejudicar os tempos de resposta. No entanto, o servidor agora está fazendo mais trabalho, como observamos antes com a taxa de transferência maior.

Todos os resultados acima foram observações sobre como o Tableau Server escalonou com o uso crescente da análise pelos usuários finais. À medida que a carga de trabalho de usuários finais aumenta no servidor, você deve garantir que haja capacidade suficiente adicionando mais nós de trabalho ao cluster para oferecer uma boa qualidade de serviço aos usuários finais.

O próximo conjunto de perguntas concentrou-se no impacto para o usuário em virtude de onde e como o Processador em segundo plano é configurado. Especificamente, qual é o impacto para o usuário de hospedar os processos do Processador em segundo plano no mesmo computador que os Serviços de análise (servidor do VizQL) em relação a isolar os processos do Processador em segundo plano em um computador diferente. Abordamos esses resultados na próxima seção.

Resultados do Processador em segundo plano

Primeiro, queríamos quantificar o impacto da execução do Processador em segundo plano em nossa instalação padrão de um único servidor. Qual é o impacto desse cenário para os usuários finais e a escalabilidade em geral? Fizemos nossos testes de carga de trabalho e registramos a TPS em uma implantação de um único Tableau Server. Em seguida, testamos as cargas de trabalho em clusters, expandindo a topologia do cluster ao adicionar nós de trabalho.

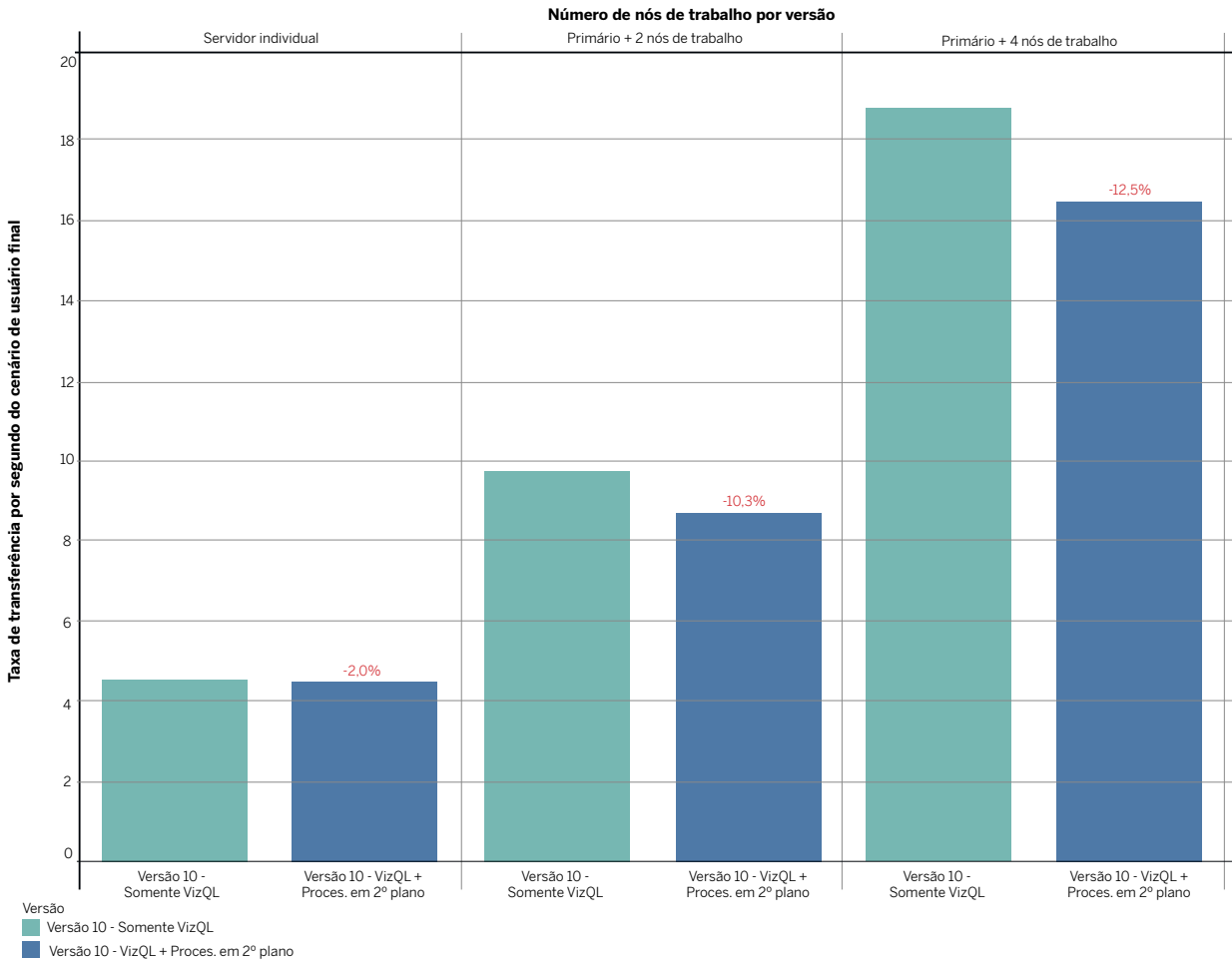


Figura 11: Impacto na escalabilidade dos usuários finais de ter o Processador em segundo plano e os serviços do VizQL no mesmo computador

A Figura 11 mostra os resultados de dois cenários de teste. As barras verdes representam o cenário de teste em que executamos somente uma carga de trabalho do VizQL. A carga de trabalho somente do VizQL simula as cargas de trabalho do usuário final. As barras azuis representam o cenário de teste em que adicionamos um volume fixo de carga de trabalho do Processador em segundo plano à carga de trabalho do VizQL no mesmo cluster. Depois registramos as mudanças no TPS para a carga de trabalho dos usuários finais nos dois cenários. Sabíamos que isso causaria um impacto porque o Processador em segundo plano e o servidor do VizQL são cargas de trabalho de computação intensiva, mas nossa intenção era medir esse impacto para as cargas de trabalho específicas.

Observamos entre 2 e 12% de redução na taxa de transferência geral do cenário de usuário final. Essa redução mostra que a capacidade do cluster de atender aos usuários finais é afetada pelo custo de processamento da carga de trabalho do Processador em segundo plano. Extrapolar estes dados mostra que, se o cluster estava processando 100 cenários de usuário final por unidade de tempo, então a adição da carga constante do Processador em segundo plano reduz a capacidade para 88 cenários de usuário na mesma unidade de tempo. Essa redução na taxa de transferência poderia traduzir-se em um impacto significativo na escalabilidade de usuário final e/ou na qualidade do serviço, dependendo de vários fatores, tais como cargas de trabalho, picos de carga, limitações de hardware e variáveis de infraestrutura.

Este teste ilustra a importância de providenciar os recursos de hardware adequados de acordo com o planejamento de cargas de trabalho do Tableau Server. A execução do Tableau Server em hardware com baixa capacidade ou limitado pode resultar em menores taxas de transferência, falha de tarefas do Processador em segundo plano, atrasos nas notificações de assinaturas e/ou erros para o usuário final manifestados como problemas de desempenho. Quando essas condições ocorrem, considere expandir o cluster para isolar as cargas de trabalho do Processador em segundo plano em hardware dedicado. Ao isolar o serviço do Processador em segundo plano, você liberará processos concorrentes vinculados à computação que, caso contrário, estariam disputando pelos mesmos recursos quando localizados no mesmo computador.

Vamos analisar esse impacto em uma carga de trabalho específica no caso das assinaturas quando o Processador em segundo plano está localizado no mesmo computador que o servidor do VizQL e quando é isolado em sua própria máquina.

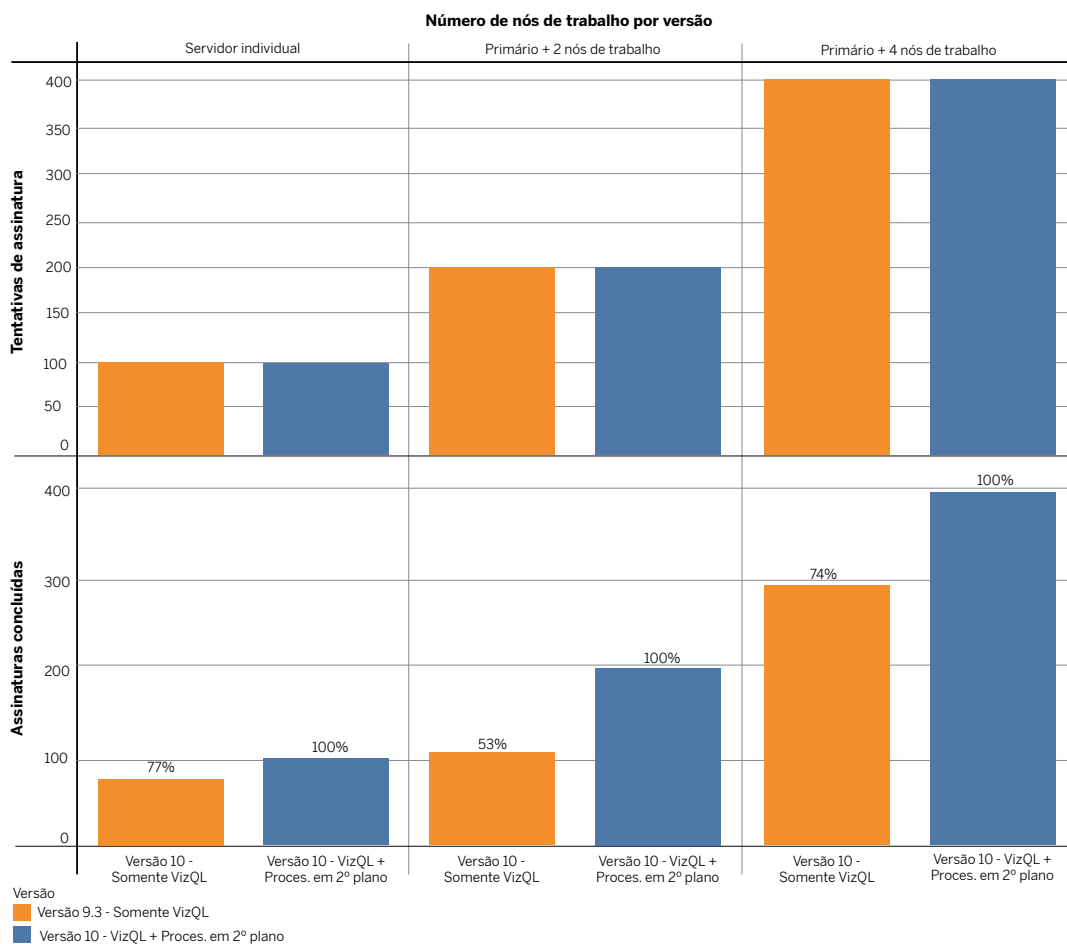


Figura 12: Melhorias na notificação de assinaturas

Na Figura 12, cada painel mostra a topologia de cluster do Tableau Server na versão 9.3 (laranja) e na versão 10 (azul), enquanto cada barra representa o número de tentativas de notificações de assinaturas e o número de notificações concluídas com êxito. Em cada iteração, aumentamos o número de nós de trabalho em passos de 0, 2, 4 e cada nó de trabalho tinha apenas um processo do Processador em segundo plano configurado. (Omitir o teste de 1 e 3 nós de trabalho não comprometeu a fidelidade dos resultados que almejávamos.) Limitamos o processo do Processador em segundo plano a um por nó de trabalho para estabelecer uma medição consistente para comparações.

À medida que aumentamos a carga de assinatura do Processador em segundo plano, o Tableau Server 10 foi capaz de concluir todo o trabalho enviado. Nos mesmos testes, o Tableau Server 9.3 começou a saturar-se e algumas assinaturas falharam. Esse comportamento pode ser agravado em uma topologia com poucos recursos e excessivamente carregada. Além disso, observamos que o Tableau Server 10 pôde assumir um número cada vez maior de assinaturas à medida que mais nós de trabalho eram adicionados ao sistema. Uma melhoria no Tableau Server 10 que explica a observação acima tem a ver com a introdução do cache de imagens para as notificações. Esse recurso permite que o Processador em segundo plano faça muito menos trabalho para as mesmas tarefas vinculadas a um agendamento. Em cenários em que o mesmo trabalho está sendo executado no mesmo agendamento, o Tableau Server 10 agora armazena em cache os resultados da primeira execução e fornece os resultados das solicitações subsequentes para a mesma carga de trabalho. Isso significa que o mesmo trabalho é concluído com mais eficiência.

Para as assinaturas que usam as mesmas pastas de trabalho no mesmo agendamento, observamos que o tempo necessário para concluir a assinatura foi reduzido de 60 a 90% em comparação com a versão 9.3. A Figura 13 ilustra essa melhoria.

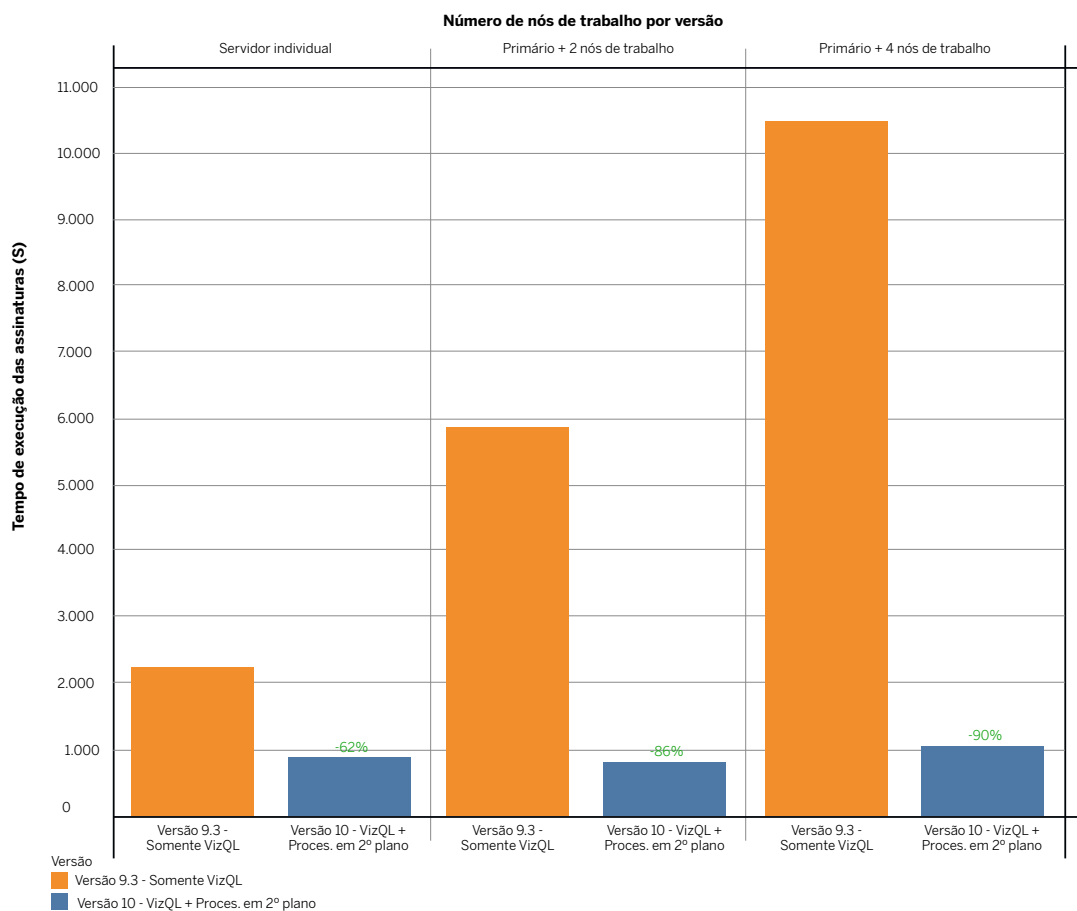


Figura 13: Tempo de execução da assinatura (tempo de conclusão): 9.3 x 10

O que isso significa é que as notificações para a mesma pasta de trabalho no mesmo agendamento levarão muito menos tempo para serem concluídas. No entanto, se você tiver notificações para pastas de trabalho diferentes ou filtros de usuário estiverem definidos, as assinaturas consumirão tempo de processamento e computação antes que a notificação possa ser entregue para o usuário. Pastas de trabalho diferentes e filtros de usuário exigem que o Tableau Server execute o funil inteiro de visualização do usuário final, que inclui a consulta de dados e processos visuais de dados para cada pasta de trabalho ou exibição filtrada.

O benefício das assinaturas é que elas oferecem aos usuários corporativos os dados importantes para eles de forma pontual. Atualizações de extrações eficazes ajudam sua organização a tomar boas decisões com base em dados devidamente atualizados. Como o processo do Processador em segundo plano gerencia ambas essas funções essenciais, planejar seus agendamentos de assinaturas de tal forma que o trabalho duplicado seja vinculado ao mesmo agendamento pode garantir os benefícios do cache.

Isolamento do processo do Processador em segundo plano

Com o processo do Processador em segundo plano isolado em seu próprio nó de trabalho, executamos os mesmos testes com assinaturas. Observamos que o nó de trabalho do Processador em segundo plano foi capaz de concluir 400 assinaturas em uma máquina única de quatro núcleos. Esse é o mesmo número de assinaturas que registramos quando um único processo do Processador em segundo plano também foi colocado em cada nó de trabalho do VizQL em quatro máquinas de trabalho.

A lição importante aqui é que, embora o próprio processo do Processador em segundo plano seja escalonado, isolar o processo proporciona escalabilidade similar, mas não afeta a qualidade do serviço para o usuário final. O processo do Processador em segundo plano é de thread único e é projetado para concluir trabalhos o mais rápido possível. Considerando esse design, um processo do Processador em segundo plano consumirá um núcleo inteiro quando tiver trabalho a fazer. Em uma máquina isolada, o uso intensivo de computação do processo do Processador em segundo plano não interferirá em outros serviços do Tableau voltados para o usuário.

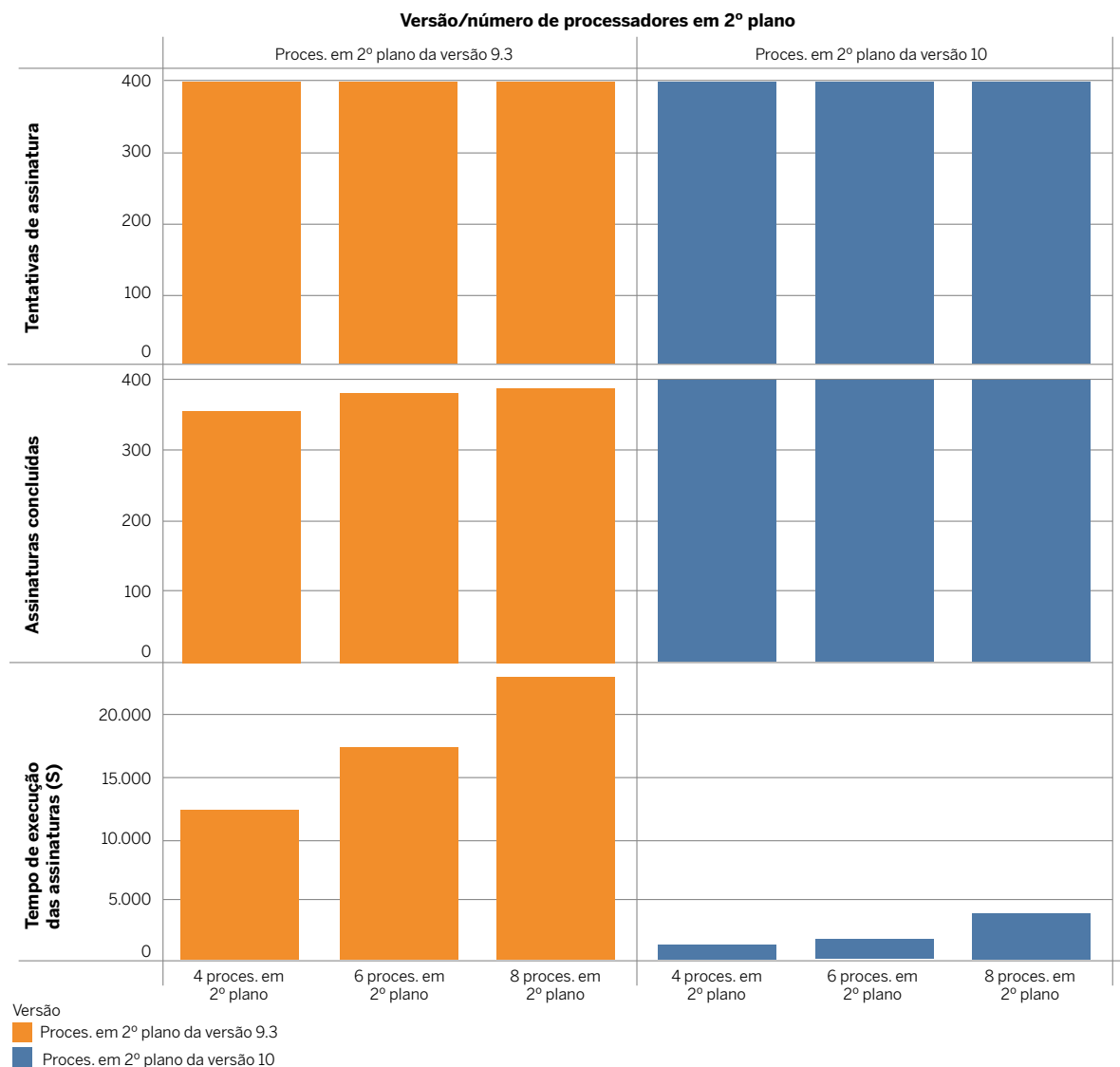


Figura 14: Adicionando processos do Processador em segundo plano a um computador de quatro núcleos: 9.3 x 10

Como mostrado na Figura 14, adicionar processos do Processador em segundo plano a um único computador de quatro núcleos permitiu que o Tableau Server 10 (azul) concluísse as assinaturas em muito menos tempo quando comparado à versão 9.3 (laranja). No entanto, continuar adicionando processos do Processador em segundo plano em um computador fisicamente limitado pelos núcleos tem um impacto negativo. Como mostrado, adicionar oito processos do Processador em segundo plano a um computador limitado a quatro núcleos retarda o tempo de conclusão. Isso ocorre devido ao design de thread único do Processador em segundo plano.

Por último, em nossos testes das atualizações de extração em nós de trabalho executando processos isolados do Processador em segundo plano, observamos que o Tableau Server 9.3 e o Tableau Server 10 foram equivalentes. Ambas as versões concluíram o mesmo número de atualizações de extração aproximadamente no mesmo tempo. Um detalhe importante a considerar ao escalonar para as atualizações de extração é que as atualizações de extração são altamente dependentes de bancos de dados externos para ter um desempenho adequado. (Em nossos testes, atualizamos os dados usando pastas de trabalho que usavam extrações publicadas do Microsoft SQL Server). O desempenho e a escalabilidade da atualização da extração dependem

muito das especificações de hardware do banco de dados. Além disso, as características dos dados, tais como os tipos de uniões e a complexidade das consultas que são executadas, afetarão a escalabilidade. Por esse motivo, você deve garantir que as notificações e atualizações de extrações tenham capacidade suficiente disponível para concluir seu trabalho antes que picos de carga do usuário final ocorram no sistema.

Considerações sobre o Processador em segundo plano

O processo do Processador em segundo plano faz grande parte do trabalho relacionado a atualizações de extrações, assinaturas e outras tarefas agendadas em segundo plano. Essas tarefas não afetarão a capacidade se suas execuções forem agendadas em horários fora do pico. Quando isso não for possível, organize-se para adicionar a capacidade necessária para que seus processadores em segundo plano e cargas de trabalho não relacionadas a usuários sejam executados simultaneamente com os processos de usuários.

Os processadores em segundo plano foram desenvolvidos para consumir a capacidade por processo de um núcleo inteiro, porque foram criados para concluir o trabalho o mais rápido possível. Ao executar vários processos do Processador em segundo plano, considere o fato de que um processo do Processador em segundo plano pode competir pelos recursos computacionais e de rede com outros serviços em execução na mesma máquina.

Práticas recomendadas: faça você mesmo o teste de escalabilidade

Se você pretende fazer seus próprios testes de carga para determinar a escalabilidade do Tableau Server em seu ambiente com suas cargas de trabalho, veja as práticas recomendadas abaixo.

- 1. Não trate o Tableau Server como uma caixa preta.** Muitas vezes, os testes de carga tradicionais tratam um aplicativo testado como uma caixa preta. Ou seja, como se não fossem necessários ajustes/configurações na implantação para atender às condições de carga. O Tableau foi projetado para oferecer escalabilidade horizontal e vertical e, para ajudar a fundamentar os testes de escalabilidade, é necessário compreender a arquitetura do Tableau a fim de gerar os resultados que funcionam com sua situação de teste
- 2. Escolha a ferramenta certa para os testes.** O Tableau Server trabalha pesado e executa atividades complexas que exigem muitos recursos. Existem diversas ferramentas disponíveis para testar cargas no Tableau Server. Embora o Tableau não ofereça suporte diretamente a essas ferramentas, você deve escolher uma que seja fácil de usar e que melhor represente seu ambiente de produção. Também é importante saber usar bem a ferramenta e o Tableau Server ao realizar os testes de carga. Usamos o [TabJolt](#) em nossos testes. O TabJolt é uma ferramenta automatizada de testes de carga baseada no JMeter e foi criado para eliminar a manutenção de scripts, além de funcionar com soluções de análise ad hoc, como o Tableau.
- 3. Escolha pastas de trabalho representativas.** Muitas vezes, quando recebemos reclamações sobre problemas de desempenho e escalabilidade, o problema é que as pastas de trabalho usadas não foram criadas de acordo com as práticas recomendadas. Se um teste com um único usuário em sua pasta de trabalho revelar um tempo de resposta muito lento, a pasta de trabalho deve ser otimizada antes de você iniciar um projeto de teste de carga. Assim como você não manteria um painel de baixo desempenho em um ambiente de produção, não deveria usá-lo para testes.

4. **Comece pelos padrões.** Ao testar pastas de trabalho com conexões em tempo real, lembre-se de que com a introdução da paralelização no Tableau Server 9.0 você talvez não precise de todos os VizQL Server que implantou para sua versão anterior do Tableau Server. Comece com a nova configuração padrão de dois processos e expanda incrementalmente conforme necessário.

Práticas recomendadas para otimização de ambientes reais

Além de ter um sistema otimizado, há práticas recomendadas que podem ser usadas para melhorar muito o desempenho do software e reduzir o tempo de resposta médio.

- Crie pastas de trabalho visualmente atraentes e de alto desempenho. Geralmente quando ouvimos o cliente sugerir que sua pasta de trabalho é lenta é porque ela foi criada sem pensar no desempenho. Se o tempo de carga de um único usuário é lento, os tempos de resposta da pasta de trabalho serão lentos sob cargas pesadas também. Ao adotar uma cultura de análise, oferecer opções e equipes com administradores que podem ajudar os usuários a criar pastas de trabalho elegantes e inteligentes com alto desempenho garantirá também a criação e o oferecimento de visualizações escalonáveis. [Como criar pastas de trabalho eficientes](#) é um whitepaper que aborda em mais detalhes a criação de painéis com ótimo desempenho.
- O tempo de resposta total para o usuário final é uma combinação de vários elementos, mas é sobretudo tempo consumido pelo Tableau somado à recuperação dos dados. Se os bancos de dados de back-end ou os tempos de consulta forem lentos, as visualizações serão lentas. É importante levar isso em consideração na sua estratégia de dados. Geralmente, as fontes de dados de uma organização são administradas e compartilhadas. Você deve garantir o fornecimento de dados relevantes que contribuam para a produtividade dos usuários corporativos. Isso significa otimizar os dados. Por exemplo, você deve garantir uniões ideais e níveis relevantes de agregação para proporcionar consultas rápidas em tabelas indexadas. Ter um processo adequado de higiene dos dados é importante para manter a integridade das suas visualizações e um bom desempenho.
- Use extrações de dados do Tableau. Se as consultas em seu banco de dados estiverem lentas, considere usar extrações para aumentar o desempenho das consultas. As extrações são armazenadas localmente no servidor e executadas na memória para que os usuários possam acessar os dados rapidamente sem fazer solicitações ao banco de dados. As extrações podem ser filtradas e agregadas facilmente e são ideais quando os usuários não precisam de detalhamento no nível da linha. As extrações melhoram significativamente o tempo de resposta e permitem que seus usuários entrem no fluxo analítico.
- Agende atualizações para horários fora do pico. Normalmente, as fontes de dados são atualizadas em tempo real, mas os usuários só precisam acessar os dados uma vez por dia ou por semana. Agendar extrações para horários fora do pico pode reduzir a carga no banco de dados e no Tableau Server durante horários de pico. Além disso, você pode adicionar mais processadores em segundo plano nas máquinas existentes ou usar um hardware dedicado se tiver capacidade de núcleo suficiente. Cogite essa opção para agilizar a criação das extrações.

- Evite operações “caras” durante horários de pico. A publicação, principalmente de arquivos grandes, é uma tarefa que consome muitos recursos. Geralmente é fácil influenciar o comportamento de publicação: peça aos usuários para publicar fora dos horários de pico, evitando momentos de sobrecarga, como segundas-feiras. Para saber o período de maior utilização dos seus servidores, use as [Exibições administrativas](#) e crie uma política com base no uso real. Dependendo de como você configurou o Tableau Server 10.0, fazer uma publicação também pode significar que uma cópia das extrações é feita em cada um dos nós do cluster para possibilitar a alta disponibilidade. Fazer isso em horários fora do pico também permite que você maximize a largura de banda da rede.
- Armazene as exibições em cache. À medida que mais usuários começam a acessar o Tableau Server, o tempo de resposta inicialmente aumentará devido à disputa por recursos compartilhados. Com o armazenamento em cache ativado, as exibições de cada solicitação que chega ao sistema serão armazenadas em cache e processadas com mais rapidez para o próximo usuário que visualizar esse mesmo painel.
- Você pode fazer um aquecimento com o processo do Servidor cache, um recurso introduzido no Tableau Server 9.0, agendando um e-mail de exibições mais visualizadas depois da atualização de suas respectivas extrações. Dessa forma, os usuários futuros utilizarão os dados armazenados em cache da sua solicitação anterior. Existem também outras maneiras de fazer um aquecimento com o cache, como usar uma ferramenta automatizada para carregar visualizações populares que geralmente atraem muito tráfego. O usuário pode manualmente invalidar o cache de consulta externo a qualquer momento para atualizar seus dados da fonte de dados. Essa ação também força uma recriação do cache. Assim, os usuários poderão ter sempre uma cópia atualizada dos dados mesmo se já houver uma versão no cache.

Resumo

O Tableau Server 10 é uma plataforma escalonável de nível empresarial que pode dar suporte a organizações de qualquer tamanho. Ele pode ser executado localmente em nuvens privadas ou públicas e pode ser escalonado linearmente com a adição da capacidade de nós de trabalho. Embora cada ambiente tenha suas próprias características e configuração únicas, a arquitetura do Tableau Server permitirá escalonar suas implantações para atender às necessidades dos seus usuários.

Embora sua escalabilidade e seu desempenho possam variar e estas recomendações não sejam adequadas para todas as situações, o Tableau Server 10 pode dar suporte a equipes ou departamentos com 25 a 100 usuários com capacidade de 8 a 16 núcleos. Para dar suporte a equipes de 100 a 1.000 usuários, dependendo de suas necessidades de uso e atualização dos dados, 16 a 24 núcleos poderão ser um bom ponto de partida. Se você precisar dar suporte a mais usuários ou cargas adicionais em segundo plano, poderá expandir sua implantação para mais de 32 a 64 núcleos, dando suporte a cargas de trabalho cada vez maiores agregando nós de trabalho adicionais para expandir até mesmo para escalas de nuvem.

Sobre a Tableau

O Tableau ajuda as pessoas a transformar dados em informações acionáveis que causam impacto. Conecte-se facilmente a dados armazenados em qualquer lugar e em qualquer formato. Faça análises rápidas sob demanda que revelam oportunidades ocultas. Arraste e solte para criar painéis interativos com análises visuais avançadas. Em seguida, compartilhe com toda a organização e permita que seus companheiros de equipe explorem seus pontos de vista sobre os dados. De multinacionais a startups recém-fundadas e pequenas empresas, pessoas em todo o mundo usam a plataforma de análise do Tableau para ver e entender seus dados.

Recursos

[Tableau para a empresa: uma visão geral de TI](#)

[Guia de administração do Tableau Server](#)

[Alta disponibilidade do Tableau Server 10.0: possibilitando análises essenciais escalonáveis](#)

[Tableau no Amazon Web Services](#)

