

# Análise de Desempenho de Servidores de Arquivos em Nuvem Privada

Alison da Silva<sup>1</sup>, Gustavo Callou<sup>1</sup>, Thiago Valentim<sup>2</sup>, Jamilson Dantas<sup>2</sup>

<sup>1</sup>Departamento de Computação  
Universidade Federal Rural de Pernambuco (UFRPE) – Recife, PE – Brasil

<sup>2</sup>Centro de Informática  
Universidade Federal de Pernambuco (UFPE) – Recife, PE – Brasil

{alison.gsilva, gustavo.callou}@ufrpe.br, {tvb, jrd}@cin.ufpe.br

**Abstract.** *The need for cloud services has increased in recent years. Internet users are increasingly using services hosted in the cloud. Therefore, there is a growing need for computing resources to be offered. The introduction of cloud computing with a file system to dynamically provision resources is a possible solution to this problem. However, the large increase in the use of this type of service can lead to negative experiences for users if the required resources do not perform satisfactorily. Therefore, in this work, a performance analysis of a file system, Nextcloud, hosted in a private cloud is performed. The model results show that service time increases significantly as users arrive in the system. The average wait time ranges from just over two minutes for a scenario with 100 users to over 12 minutes for a scenario with 500 users.*

**Resumo.** *A necessidade de serviços em nuvem vem aumentando nos últimos anos, onde usuários da Internet passaram a consumir cada vez mais serviços hospedados na nuvem. Portanto, observa-se a necessidade crescente por recursos computacionais que precisam ser ofertados. Adotar a computação em nuvem com um sistema de arquivos para fornecer dinamicamente recursos é uma das possíveis soluções para esse problema. No entanto, o grande aumento na utilização desse tipo de serviço pode gerar experiências negativas para os usuários caso os recursos necessários não tenham um desempenho satisfatório. Portanto, esse trabalho realiza uma análise de desempenho de um sistema de arquivos, o Nextcloud, hospedado em uma nuvem privada. Os resultados do modelo demonstram que a medida que usuários entram no sistema, o tempo de atendimento aumenta consideravelmente. O tempo médio de espera parte de um pouco mais de 2 minutos para um cenário de 100 usuários pra mais de 12 minutos considerando um cenário de 500 usuários.*

## 1. Introdução

A pandemia do COVID-19 afetou toda a população mundial, e os seus impactos podem ser observados tanto nas universidades como em empresas que passaram a ter que incorporar, rapidamente, o conceito do trabalho *home office*. Devido a esse cenário, houve um aumento da demanda por recursos computacionais (ex., armazenamento e processamento) para dar suporte a tal modalidade de trabalho. E, nesse sentido, a computação em nuvem

se apresenta como uma grande aliada para a mitigação desses problemas por oferecer infraestrutura, plataforma e software como serviço. Além de suas principais características como elasticidade e escalabilidade.

No cenário em que empresas e órgãos governamentais tiveram que se adaptar a essa nova realidade, acelerada em virtude da COVID-19, investimentos foram conduzidos na área da tecnologia da informação a fim de suprir essa demanda não esperada. Então, soluções *open source* ganharam uma maior visibilidade. Ao se adotar tais soluções, se reduz o custo com licença de *softwares* e, ainda assim, é possível realizar as personalizações conforme as necessidades de cada negócio. Exemplos dessas soluções de *software open source* para a computação em nuvem são o Apache Cloudstack, e o Openstack.

Com o aumento da demanda da utilização de recursos computacionais para dar suporte tanto aos professores e aos alunos, passou a ser preciso adotar um serviço como o Nextcloud para o armazenamento dos dados. O Nextcloud é uma plataforma de colaboração de conteúdo acessível por meio de interfaces móveis, *desktop* e *web*. Nesse ambiente, tanto alunos, como professores e pesquisadores podem colaborar em tempo real em documentos, conversar e participar de videochamadas, acessar seus e-mails e agendar reuniões em calendários.

Esse trabalho tem como foco a avaliação de desempenho do Nextcloud hospedado em um ambiente de nuvem privada com o Apache Cloudstack. Os cenários avaliados têm como objetivo o de atestar a qualidade e a quantidade de acessos simultâneos do serviço oferecido. Dessa forma, esse trabalho propõe uma estratégia, baseada em modelos, para a avaliação de desempenho do Nextcloud em nuvem privada. Exemplos de métricas que são avaliadas são o tempo de resposta e a vazão considerando cenários com diferentes cargas de trabalho. Vale destacar que como cenário de estudo foi adotada uma instituição brasileira, a Universidade Federal Rural de Pernambuco (UFRPE). Além disso, é importante destacar que tanto governos como empresas estão se afastando dos provedores de nuvens públicas em meio a crescentes preocupações sobre a soberania e segurança digital dos dados.

O artigo está organizado da seguinte forma. A Seção 2 detalha os trabalhos relacionados a essa pesquisa. A Seção 3 apresenta os conceitos básicos. A Seção 4 apresenta a metodologia adotada. A Seção 5 apresenta o cenário de teste configurado em laboratório. A Seção 6 apresenta o modelo proposto em rede de Petri estocástica (SPN) para avaliar o desempenho. A Seção 7 apresenta os estudos de casos. Por fim, a Seção 8 conclui e apresenta os futuros direcionamentos desta pesquisa.

## **2. Trabalhos Relacionados**

Uma série de trabalhos vem sendo propostos para avaliação de desempenho de infraestruturas de nuvens com o Openstack e o Cloudstack. Em [Moreira et al. 2021], os autores avaliaram o Nextcloud em uma nuvem privada com o Openstack, e usando o GlusterFS como ambiente de armazenamento. Foram realizados testes de desempenho na leitura e escrita de arquivos, utilizando o *benchmarks* IOzone. Santos et al. [Santos et al. 2018] realizaram um estudo comparativo do desempenho da CPU, memória e disco do Nextcloud, Owncloud e Pydio. Para coletar as métricas de interesse foi utilizado a ferramenta Nmon.

Shahzadi et al. [Shahzadi et al. 2017] realizaram testes comparativos de desempe-

no e escalabilidade entre as nuvens Openstack e Devstack utilizando o *benchmark Rally*. Em todos os cenários foram coletados a métrica de tempo de resposta. Os autores em [Vogel et al. 2017] avaliaram o desempenho da infraestrutura do Cloudstack utilizando as métricas de tempo de resposta e latência da rede na criação de instâncias. Para a realização desse estudo e quantificação das métricas de interesse, os autores utilizaram os *benchmarks* NetPIPE e Uperf.

Lima et al. [Lima et al. 2021] avaliaram o desempenho e o consumo de energia de um ambiente virtual de aprendizagem hospedado no Apache Cloudstack e no Openstack. O ambiente foi estressado para quantificação de métricas de tempo de resposta, vazão, utilização de Cpu e memória e consumo de energia. Também foi construído um modelo em SPN para representar o cenário. Matos et al. [Matos et al. 2020] propôs modelos hierárquicos para detecção de gargalos no ambiente de nuvem privada. As métricas utilizadas nessa pesquisa foi disponibilidade, utilização da vm, tempo de resposta e vazão.

Diferente dos trabalhos anteriores, esse trabalho faz uma análise de desempenho do servidor de arquivos (NextCloud) hospedado na nuvem privada CloudStack. Para isso, medições das métricas de tempo de resposta e vazão foram realizadas e um modelo de desempenho em SPN foi proposto.

### 3. Fundamentação Teórica

Esta seção apresenta os conceitos necessários para um melhor entendimento deste trabalho.

#### 3.1. Rede de Petri Estocástica

As redes de Petri representam uma dos muitos formalismos matemáticos existentes utilizada para modelar sistemas discretos e distribuídos, como sistemas de produção, de controle e sistemas complexos. Essa rede é representada por um grafo bipartido, composto por lugares, transições e com arcos direcionados. As redes de Petri estocásticas (SPNs) são uma extensão das redes de Petri por adicionarem, principalmente, o conceito de tempo. Então, os tempos para a realização de determinadas atividades, e que fazem com que o sistema mude de estado, podem ser associados às transições temporizadas e que são exponencialmente distribuídas. Além disso, as SPNs também adicionam o conceito de transições imediatas, onde o tempo demandado para se realizar essa atividade é zero (0) [Marsan et al. 1998]. As transições imediatas possuem uma maior prioridade sob as transições exponenciais, ou seja, estando as duas habilitadas a dispararem, a transição imediata sempre irá ser disparada primeiro. A seguir, a Figura 1 mostra os elementos que compõe as SPNs.

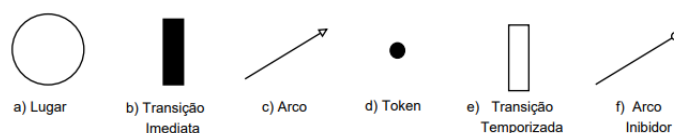
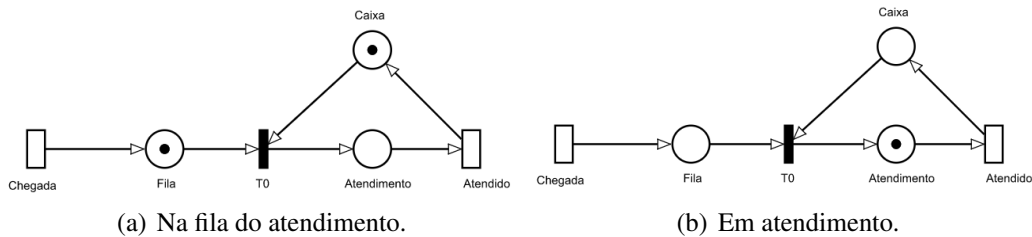


Figura 1. Elementos de uma SPN.

A Figura 2 apresenta um exemplo de um modelo SPN que representa o funcionamento de um sistema com uma única fila. Esse modelo contempla o percurso da chegada de clientes à fila até o momento de ser atendido. A transição *Chegada* representa

a frequência de chegada de clientes ao sistema. O disparo dessa transição faz com que um token seja gerado no lugar *Fila*, conforme pode ser visto na Figura 2(a). Um token no lugar *Caixa* representa a quantidade de caixas disponíveis para atender aos clientes no sistema. Sendo assim, tendo um token no lugar *Caixa* e um token no lugar *Fila*, pode-se disparar a transição *T0*. O disparo dessa transição tem o tempo zero (0), e consome os tokens dos lugares *Caixa* e *Fila* e gera um token no lugar *Atendimento*, representando que o caixa foi alocado para realizar o seu atendimento (ver Figura 2(b)). O disparo da transição *Atendido* representa o término do atendimento. Esse disparo consome um token do lugar *Atendimento* e gera um token para o lugar *Caixa*, representando que o caixa se encontra livre para atender outro cliente.



**Figura 2. Exemplo da SPN**

A seguir, iremos explicar como definir uma regressão linear. É importante destacar que a equação definida pela regressão pode ser adotada nas transições temporizadas nos modelos em SPN.

### 3.2. Regressão Linear

A regressão linear é uma técnica estatística para relacionar uma variável  $x$  independente com uma variável  $y$  que é dependente [Matos 1995]. A regressão linear simples é dada pela seguinte Equação 1.

$$y = a + bx \quad (1)$$

onde  $y$  é a resposta de saída prevista,  $x$  é a variável de entrada,  $a$  e  $b$  são os parâmetros da regressão linear.

O cálculo do parâmetro  $b$  da regressão linear é dado pela seguinte Equação 2.

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (2)$$

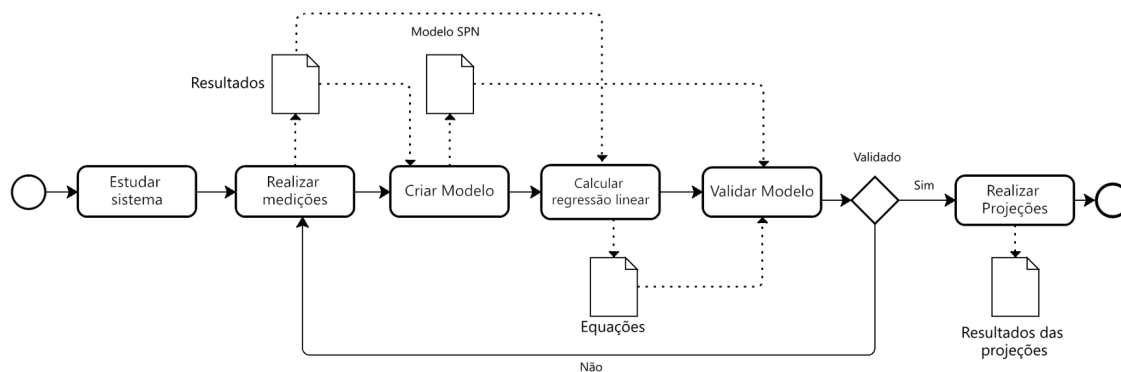
onde  $n$  é quantidade de pontos  $(x, y)$ .

Existe um grau de erro associado a regressão linear chamado de coeficiente de determinação, sendo representado por  $R^2$  [Lilja 2000]. Esse grau de erro varia entre zero e um. Quanto mais próximo o valor do coeficiente de determinação estiver de um, maior será a qualidade da expressão definida pela regressão. O cálculo do coeficiente de determinação é dado pela seguinte Equação 3.

$$R^2 = \frac{b \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

## 4. Metodologia

Esta seção apresenta a metodologia usada para a avaliação de desempenho do Nextcloud em nuvem privada. Também são contextualizados as atividades realizadas e os objetos adquiridos em cada uma das atividades. A Figura 3 ilustra a metodologia adotada para este trabalho, destacando as suas atividades que são: estudar o sistema, realizar medições, criar modelo, calcular regressão linear e validar o modelo e os objetos adquiridos ao final de cada atividade. A seguir, cada fase da metodologia adotada será detalhada.



**Figura 3. Metodologia proposta.**

A primeira etapa da metodologia corresponde ao estudo do sistema que será analisado. Sendo assim, o objetivo dessa atividade é realizar um estudo sobre a infraestrutura do Apache Cloudstack, entendimento da plataforma Nextcloud e definições de métricas que serão utilizadas na análise de desempenho. Em seguida, medições são realizadas, com o auxílio da ferramenta Jmeter, para gerar a carga de trabalho ao Nextcloud e coletar as métricas de tempo de resposta e vazão. Nesta etapa também deve-se definir as atividades de carga de trabalho que serão realizadas. Por exemplo, foram definidas seis atividades para serem realizadas no Nextcloud que são: acessar o ambiente do Nextcloud, realizar *login*, acessar a base de arquivos, realizar o *upload* de um arquivo de 1 MB e sair do ambiente do Nextcloud ao realizar o *logout*.

Na atividade de medição também deve-se buscar identificar a quantidade máxima de usuários simultâneos que o sistema irá ter capacidade de gerenciar simultaneamente. Por exemplo, em nossos testes foram utilizados no máximo 90 usuários simultâneos, pois ao se utilizar mais do que essa quantidade perdas de pacotes eram recorrentes.

A etapa seguinte corresponde a criação do modelo de desempenho. É importante destacar que o modelo SPN criado nessa atividade será refinado na atividade seguinte com a utilização da regressão linear. A expressão obtida através da regressão será parâmetro de entrada para as transições do modelo SPN proposto para representar os tempos que cada atividade demanda nos mais diferentes cenários sob análise. Por exemplo, ao se variar a quantidade de usuários no sistema. Uma vez criado o modelo, deve-se partir para a próxima etapa, a da regressão linear. Por exemplo, foi preciso fazer uma regressão para a definição da relação entre a quantidade de clientes no sistema e o tempo de atendimento. Ao final foram geradas equações que serão utilizados no modelo SPN de desempenho proposto.

A etapa seguinte da metodologia corresponde a validação do modelo SPN.

Esta fase tem como objetivo validar o modelo para garantir que ele é confiável e realiza o que foi proposto. Para a realizar a validação foi utilizado as técnicas do teste t pareado [Shapiro and Wilk 1965] e seus pressupostos que são o teste de normalidade de Shapiro-Wilk [Shapiro and Wilk 1965] e o teste de variâncias homogêneas [Snedecor and Cochran 1967]. Uma vez de posse do modelo validado, deve-se ir para a etapa de realização das projeções. Esta fase será responsável para realizar projeções do comportamento do sistema em cenários que não tivemos como analisar por limitações, sejam de usuários para testar o ambiente ou até limitações físicas dos *hardwares* disponíveis em laboratório.

## 5. Ambiente de Teste

Esta seção tem como objetivo descrever o ambiente implantado em laboratório para a realização dos cenários a serem avaliados no estudo de caso. A Figura 4 ilustra a arquitetura do ambiente de teste. Esse ambiente é composto por 4 máquinas físicas, sendo três computadores HP e um servidor Dell. As máquinas HP possuem as mesmas configurações (processador AMD A8-550B 3.2 GHz com 4 cores, 8 GB de memória RAM, 500 GB de armazenamento e sistema operacional CentOS 7). Já o servidor Dell tem um processador Intel Xeon e3-1220 3.0 GHz com 4 cores, 16 GB de memória RAM, 1 TB de armazenamento e sistema operacional CentOS 7.

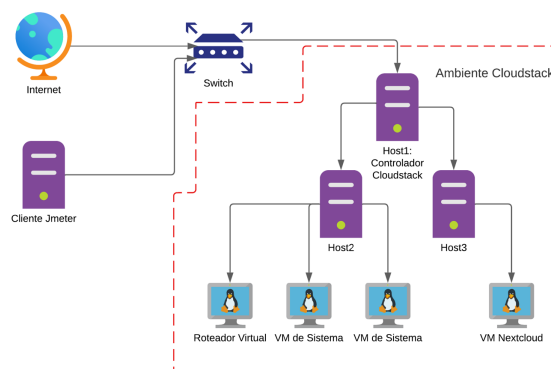


Figura 4. Arquitetura do ambiente de teste.

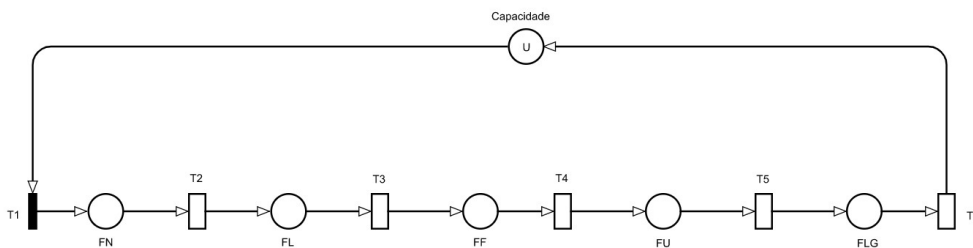
Uma máquina física HP foi utilizada como cliente, onde foi instalada a ferramenta de geração de carga utilizada que foi o JMeter. Um plano de teste foi definido nessa ferramenta para a realização dos experimentos no ambiente do Cloudstack implementado em laboratório com o NextCloud configurado. Por exemplo, foi construída uma base de dados de usuários com a função de executar uma carga de trabalho para os testes no ambiente gerenciado pelo Cloudstack. Foram definidas cinco atividades para realizar no ambiente: (i) acessar o Nextcloud, (ii) realizar o login, (iii) entrar nos arquivos, (iv) realizar um *upload* de arquivo de 1 MB e (v) finalizar realizando o logout.

O ambiente do Cloudstack foi configurado com três hosts (máquinas físicas), sendo o servidor Dell destinado ao ambiente que foi diagnosticado por demandar mais carga, no caso, o host3 com a VM do Nextcloud. O gerenciador da nuvem do Cloudstack foi configurado em uma máquina HP, sendo ela o responsável por fazer o gerenciamento dos recursos, por exemplo, descobrir novos host e os recursos por eles disponibilizados.

Com o objetivo de isolar mais o ambiente que estamos analisando, no caso o host3 com o Nextcloud, optamos por colocar todas as VMs que são nativas do ambiente do Cloudstack (máquinas virtuais do sistema e o roteador virtual) em um host em separado (host2). O host2 é responsável pela alocação de todos os recursos de rede (ex., IPs das VMs), alocação de hardware (ex., espaço em disco e memória RAM para as VMs), segurança, entre outros. A instância do Nextcloud (VM) foi configurada com 2 cores, 4 GB de memória RAM e 50 GB de armazenamento.

## 6. Modelo

Esta seção apresenta o modelo de desempenho proposto, em SPN, para representar as operações realizadas no Nextcloud contemplando o cenário previamente descrito. Além disso, as métricas utilizadas, os parâmetros adotados e as regressões lineares colocadas no modelo também são apresentados. A Figura 5 ilustra o modelo SPN proposto que representa a chegada de clientes e a realização das seguintes atividades: acesso ao Nextcloud, realizar o *login*, acesso aos arquivos, realizar o *upload* de um arquivo e realizar o *logout*.



**Figura 5. Modelo SPN de desempenho proposto.**

O lugar *Capacidade* representa a quantidade de clientes ( $U$ ) que estarão no sistema sendo atendidos pelo serviço simultaneamente. A transição imediata  $T1$  será disparada sempre que existir clientes para serem atendidos. Ao realizar o disparo dessa transição, um token é consumido do lugar *capacidade* e outro token é gerado no lugar  $FN$ , representando que um cliente chegou ao sistema. Os lugares  $FN$ ,  $FL$ ,  $FF$ ,  $FU$ ,  $FLG$  representam as filas de requisições para acessar o Nextcloud, realizar o *login*, acessar os arquivos, realizar o *upload* do arquivo e realizar o *logout*, respectivamente. As transições  $T2$ ,  $T3$ ,  $T4$ ,  $T5$  e  $T6$  representam o tempo demandado para realizar cada uma dessas atividades. O disparo destas transições faz com que  $U$  tokens sejam consumidos do lugar que antecede a transição e gere *tokens* para o lugar subsequente. Isso ocorre devido à semântica adotada em cada transição temporizada do modelo ( $T2$  a  $T6$ ) que é a de *Infinite Server*, conforme pode ser visto na Tabela 1. Além disso, a Tabela 2 apresenta uma descrição dos lugares do modelo e mostra o número inicial de tokens para cada lugar. A seguir, mais detalhes sobre os parâmetros de entrada e definições das métricas adotadas serão apresentados.

### 6.1. Métricas

No modelo SPN foram definidas duas métricas que são a de tempo de resposta e vazão. A métrica de tempo de resposta é representada no modelo por  $TR$ , sendo o período entre o envio da requisição para o sistema até a chegada da resposta ao cliente. A Equação 4 é a responsável por esse cálculo no modelo proposto.

**Tabela 1. Descrição das transições do modelo.**

Transição	Tipo	Tempo	Tipo Serv.
T1	Imediata	-	-
T2	Exponencial	$t_{Nextcloud}$	Infinite Server
T3	Exponencial	$t_{Login}$	Infinite Server
T4	Exponencial	$t_{File}$	Infinite Server
T5	Exponencial	$t_{Upload}$	Infinite Server
T6	Exponencial	$t_{Logout}$	Infinite Server

**Tabela 2. Descrição dos lugares do modelo.**

Lugar	N.º Token Inicial	Descrição
Capacidade	U	Quantidade de clientes que serão atendidos pelo serviço.
FN	0	Fila de clientes para realizar o acesso ao Nextcloud.
FL	0	Fila de clientes para realizar o login.
FF	0	Fila de clientes para realizar o acesso aos arquivos.
FU	0	Fila de clientes para realizar o upload do arquivo.
FLG	0	Fila de clientes para realizar o logout.

$$TR = \frac{E\{\#FN\} + E\{\#FL\} + E\{\#FA\} + E\{\#FU\} + E\{\#FLT\}}{\frac{E\{\#FLG\}}{T_6}} \quad (4)$$

onde  $E\{\#FN\}$ ,  $E\{\#FL\}$ ,  $E\{\#FF\}$ ,  $E\{\#FU\}$  e  $E\{\#FLG\}$  representa a esperança de tokens nos lugares de nomes correspondentes, ou seja, indicam o número médio de tokens nos seus respectivos lugares. O  $T_6$  indica o tempo para ser realizado o logout dos usuários.

A métrica do modelo  $V$  é responsável pelo cálculo da vazão do sistema. A vazão representa a quantidade de requisições que o sistema atende durante um período. A fórmula da vazão é definida na Equação 5.

$$V = \frac{E\{\#FN\} + E\{\#FL\} + E\{\#FA\} + E\{\#FU\} + E\{\#FLT\}}{T_2 + T_3 + T_4 + T_5 + T_6} \quad (5)$$

onde  $E\{\#FN\}$ ,  $E\{\#FL\}$ ,  $E\{\#FF\}$ ,  $E\{\#FU\}$  e  $E\{\#FLG\}$  indicam o número médio de tokens nos seus respectivos lugares. Já as transições  $T_2$ ,  $T_3$ ,  $T_4$ ,  $T_5$  e  $T_6$  indicam o tempo para realizar cada uma das atividades correspondentes.

## 6.2. Parâmetros de entrada

O modelo SPN proposto possui seis parâmetros que devem ser especificados para que se consiga fazer a avaliação do modelo. O primeiro parâmetro é  $U$  que representa a quantidade de clientes que serão atendidos pelo serviço. Os outros parâmetros são o  $t_{Nextcloud}$ ,



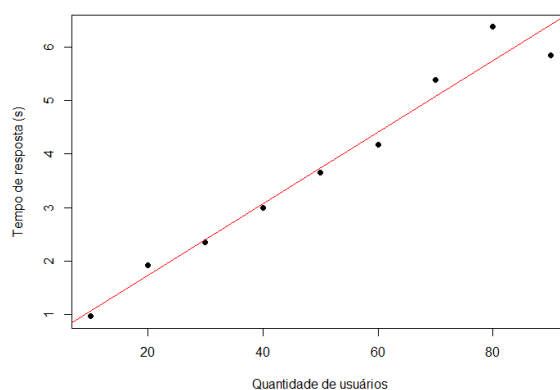
que é o tempo demandado para acessar e ter a resposta do Nextcloud;  $tLogin$ , tempo demandado para a realização do login no sistema do Nextcloud; o  $tFile$ , tempo decorrido para se acessar o ambiente de arquivos do Nextcloud;  $tUpload$ , tempo para se subir um arquivo local ao sistema; e o  $tLogout$ , tempo necessário para se sair do ambiente do sistema. É importante destacar que estamos lidando sempre com quantidades distintas de usuários acessando o sistema. Sendo assim, adotamos a estratégia de medir o comportamento do sistema com diferentes quantidades de usuários realizando operações simultaneamente. Dessa forma, foi possível definir equações matemáticas para cada uma das atividades a partir de regressões lineares conforme será explicado a seguir.

### 6.3. Regressão Linear

A Figura 6 apresenta a regressão linear entre a quantidade de clientes ( $x$ , independente) e o tempo decorrido para obter a resposta da requisição ( $y$ , dependente) da atividade de acesso ao Nextcloud. É possível notar que os pontos estão bem próximos da linha de tendência, e o coeficiente de determinação  $R^2$  foi de 0,9665. Quanto mais próximo o  $R^2$  for de um, significa um melhor ajustamento da regressão [Gujarati and Porter 2011]. A regressão linear entre a quantidade de clientes e o tempo de atendimento para acessar o Nextcloud é dada pela seguinte Equação 6.

$$y = (66,81 \times U) + 399,55 \quad (6)$$

Onde o  $y$  é a previsão do tempo de resposta para acessar o Nextcloud e  $U$  é a quantidade de clientes no sistema. A Equação 6 será utilizada na transição  $T2$  do modelo de desempenho.



**Figura 6. Regressão linear acessar o Nextcloud.**

As outras quatro regressões lineares seguem o mesmo rumo da apresentada anteriormente. A Tabela 3 apresenta os coeficientes de determinação  $R^2$  e as equações da regressão linear para cada atividade realizada. As equações das regressões lineares apresentadas nessa tabela são utilizadas nas transições  $T2$ ,  $T3$ ,  $T4$ ,  $T5$  e  $T6$  do modelo SPN de desempenho (ver Figura 5).

## 7. Estudo de Caso

Esta seção apresenta os estudos de caso que contemplam o cenário previamente apresentado. O primeiro estudo tem como objetivo o de apresentar a validação do modelo SPN

**Tabela 3. Coeficiente de determinação e equação.  $y$  corresponde a previsão do tempo de atendimento de um cliente e  $U$  é a quantidade de clientes no sistema.**

Atividade	Transição	$R^2$	Equação
Acessar Nextcloud	T2	0,9665	$y = (66,81 \times U) + 399,55$
Login	T3	0,9671	$y = (889,8 \times U) - 729,2$
Acessar Arquivo	T4	0,9542	$y = (84,17 \times U) - 357,35$
Upload	T5	0,9626	$y = (66,12 \times U) - 408,99$
Logout	T6	0,975	$y = (125,8 \times U) - 706,5$

de desempenho proposto através da métrica de tempo de resposta. Em seguida, adotamos o modelo para fazer algumas projeções que não temos como quantificar por medição.

### 7.1. Estudo de Caso I - Validação do Modelo

Esse estudo tem como objetivo a validação do modelo de desempenho proposto (ver Figura 5). A validação é um procedimento utilizado para certificar que os valores gerados pelo modelo apresentam coerência com os gerados pelo sistema real. No nosso caso, a validação foi feita a partir da comparação do tempo de resposta medido no sistema real com o tempo obtido via análise estacionária no modelo.

A Tabela 4 apresenta um resumo dos resultados obtidos para o tempo de resposta tanto para a medição como para a modelagem. Como parâmetros de entrada da análise do modelo, adotamos o nível de confiança de 95% e o erro relativo em 10%. É possível perceber que os intervalos de confiança apresentados para a medição e modelagem se sobrepõem.

**Tabela 4. Resultado para validação do modelo com tempo de resposta.**

Qtd. Clientes	Medição		Modelagem	
	T.(s)	Int. Conf.	T.(s)	Int. Conf.
10	8,53	[7,89; 8,75]	10,53	[8,41; 10,83]
20	18,74	[17,12; 19,43]	22,83	[19,29; 24,12]
30	37,97	[36,73; 39,20]	35,21	[29,15; 41,91]
40	51,24	[50,71; 52,47]	47,70	[39,85; 57,48]
50	71,18	[70,35; 72,89]	59,17	[53,10; 71,01]
60	77,56	[75,99; 78,72]	71,97	[62,24; 78,50]
70	86,73	[85,44; 87,21]	84,72	[75,68; 92,73]
80	101,74	[100,32; 103,04]	96,86	[82,14; 103,24]
90	103,33	[102,46; 104,98]	109,20	[99,35; 116,89]

A fim de realizar a validação do modelo, aplicamos o teste  $t$  pareado com os tempos de medição e os obtidos via modelagem. O teste  $t$  é um teste de hipótese e tem como objetivo comparar duas médias extraídas do mesmo grupo. Para a realização do teste  $t$  pareado é necessário validar alguns pressupostos.

O primeiro pressuposto é para verificar se as amostras seguem uma distribuição normal, então a hipótese nula seria que as médias são de uma distribuição normal e a

hipótese alternativa que não são de uma distribuição normal. Para a realização de todos os testes será adotado um nível de significância de 5%. Utilizando o teste de Shapiro-Wilk, o p-valor encontrado para medição e modelagem são de 0,5114 e 0,9133, respectivamente. Como os p-valores foram maior que o nível de significância, então aceitamos a hipótese nula.

O segundo pressuposto é para avaliar se as variâncias das amostras são homogêneas (hipótese nula) ou não são homogêneas (hipótese alternativa). Realizando o teste de variâncias homogêneas, o resultado do p-valor foi de 0,9394, como o resultado foi maior que o nível de significância aceita-se a hipótese nula. Após a realização dos pressupostos é realizado o teste-t, onde a hipótese nula é se as médias são iguais e a hipótese alternativa se são diferentes. O p-valor para este teste foi de 0,2843, como o p-valor foi maior que o nível de significância, aceitamos a hipótese nula que não há diferença entre as médias das amostras.

## 7.2. Estudo de Caso II - Projeções

Esse estudo tem como objetivo utilizar o modelo proposto para a realizar projeções com uma maior quantidades de usuários realizando as atividades propostas. As simulações foram feitas de 100 a 500 clientes. A Tabela 5 apresenta os resultados obtidos das projeções.

**Tabela 5. Resultado das projeções.**

Qtd. Clientes	T.(s)	Int. Conf.
100	132,38	[120,39; 140,01]
200	253,21	[241,05; 259,37]
300	403,94	[392,86; 415,34]
400	557,06	[460,78; 618,68]
500	751,70	[699,85; 803,49]

Como pode ser visto na Tabela 5 a medida que aumenta a quantidade de clientes simultâneos no mesmo sistema, o tempo de atendimento aumenta consideravelmente. Por exemplo, o tempo para realizar as atividades sequencialmente de acesso ao Nextcloud, realizar o *login*, acesso aos arquivos, realizar o *upload* de um arquivo e realizar o *logout* partiu de pouco mais de 2 minutos, cenário com 100 usuários, para acima de 12 minutos considerando 500 clientes simultâneos. Logo, nesse caso, pode ser inviável dado o tempo demandado de espera.

## 8. Conclusão

Este trabalho propôs um modelo para avaliação de desempenho do Nextcloud em nuvem privada. A abordagem proposta usa modelos em redes de Petri estocásticas (SPN) para avaliar o desempenho da arquitetura proposta. Foram realizadas medições do tempo de resposta de atividades feitas no Nextcloud e a validação do modelo proposto para representar o cenário de teste. Além disso, o modelo contempla regressões lineares realizadas em cada uma das atividades que representam os cenários de testes realizados. Sendo assim, foi possível também analisar outros cenários que foram inviáveis de medir, mostrando a validade do modelo na análise de cenários com maior quantidade de usuários no sistema.

Como trabalhos futuros, pretende-se analisar também o consumo de energia decorrente das atividades realizadas, além de, avaliar cenários que contemplam atividades de auto escalonamento de máquinas virtuais para garantir que trabalhem em níveis de desempenho aceitáveis.

## Agradecimentos

Os autores agradecem ao CAPES, CNPq e a FACEPE pelos suportes financeiros para este trabalho.

## Referências

- Gujarati, D. N. and Porter, D. C. (2011). *Econometria básica /*. AMGH,, Porto Alegre :, 5. ed. edition. Tradução de: Basic econometrics.
- Lilja, D. J. (2000). *Measuring Computer Performance: A Practitioner's Guide*. Cambridge University Press.
- Lima, C. J. d., Silva, A. V. G. d., Andrade, E. C. d., and Callou, G. R. d. A. (2021). Performance and power consumption evaluation of the moodle environment. *Research, Society and Development*, 10(5):e57010515191.
- Marsan, M. A., Balbo, G., Conte, G., Donatelli, S., and Franceschinis, G. (1998). Modeling with generalized stochastic petri nets. *SIGMETRICS Perform. Eval. Rev.*, 26(2):2.
- Matos, M. A. (1995). Manual operacional para a regressão linear. *Faculdade de Engenharia da Universidade do Porto*, page 63.
- Matos, R., Dantas, J., Araujo, E., and Maciel, P. (2020). Bottleneck detection in cloud computing performance and dependability: Sensitivity rankings for hierarchical models. *Journal of Network and Systems Management*, 28.
- Moreira, C., Monteiro, R., Martins, L., Albuquerque, R., Filho, F., and de Sousa Junior, R. (2021). Compartilhamento de arquivos em home office: uma solução de armazenamento de arquivos segura e altamente disponível em um ambiente de nuvem privada. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 01:409–424.
- Santos, D. L., Camboim, K., and Alencar, F. (2018). Avaliação de desempenho de nuvens privadas: Um comparativo entre owncloud, nextcloud e pydio. In *Anais do XVI Workshop em Clouds e Aplicações*, Porto Alegre, RS, Brasil. SBC.
- Shahzadi, S., Iqbal, M., Qayyum, Z. U., and Dagiuklas, T. (2017). Infrastructure as a service (iaas): A comparative performance analysis of open-source cloud platforms. In *2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–6. IEEE.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Snedecor, G. W. and Cochran, W. G. (1967). *Statistical methods / George W. Snedecor, William G. Cochran*. Iowa State University Press Ames, 6th ed. edition.
- Vogel, A., Griebler, D., Schepke, C., and Fernandes, L. G. (2017). An intra-cloud networking performance evaluation on cloudstack environment. In *2017 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 468–472.