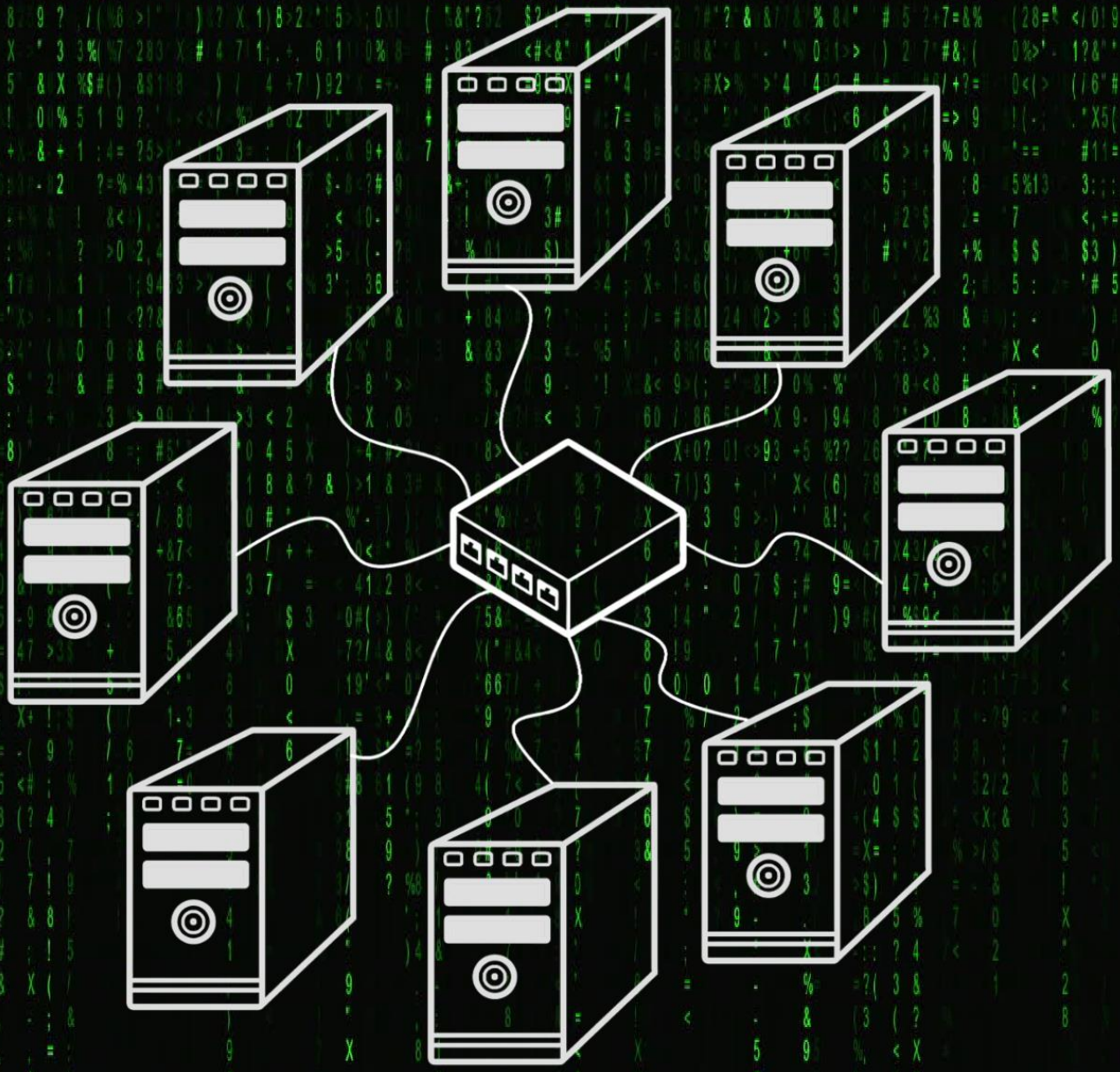
Montagem de um Cluster Beowulf

Primeiros Passos



Autores

João Fernandes Santos Filho

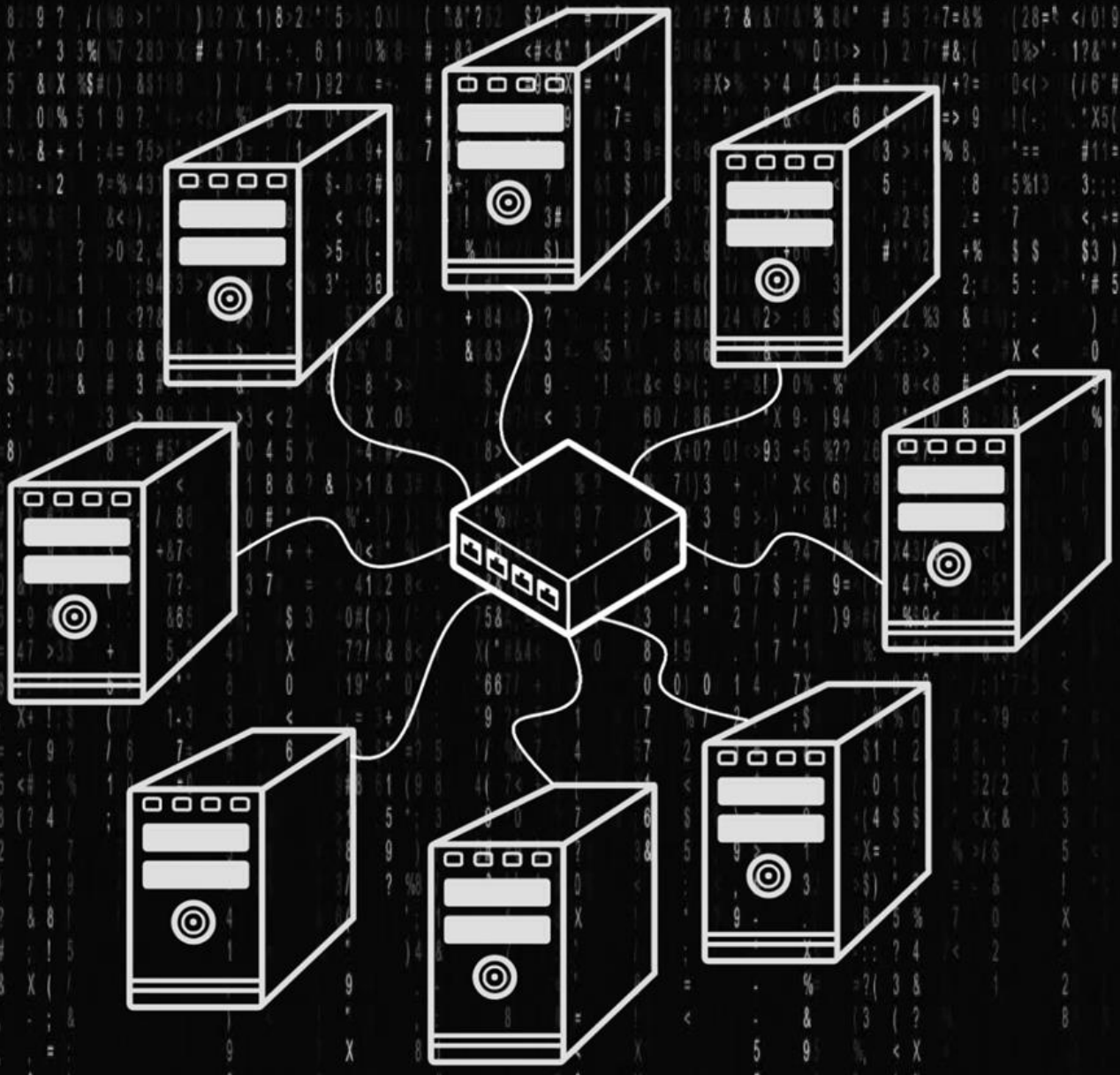
José Aprígio Carneiro Neto

FORMA EDUCACIONAL

2025

Montagem de um Cluster Beowulf

Primeiros Passos



Autores

João Fernandes Santos Filho

José Aprígio Carneiro Neto

FORMA EDUCACIONAL

2025

www.formaeducacional.com.br
formaeducacional@gmail.com

Autores
João Fernandes Santos Filho
José Aprígio Carneiro Neto

Editor Chefe: Jader Luís da Silveira
Editoração: Resiane Paula da Silveira
Capa: Os autores
Revisão: Os autores

Dados Internacionais de Catalogação na Publicação (CIP)

Montagem de um Cluster Beowulf: Primeiros Passos

S237m / João Fernandes Santos Filho; José Aprígio Carneiro Neto. – Formiga (MG): Forma Educacional Editora, 2025. 95 p. : il.

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui bibliografia
ISBN 978-65-85175-47-0
DOI: 10.5281/zenodo.17095562

1. Cluster Beowulf. 2. Guia para Montagem. I. Santos Filho, João Fernandes. II. Carneiro Neto, José Aprígio. III. Título.

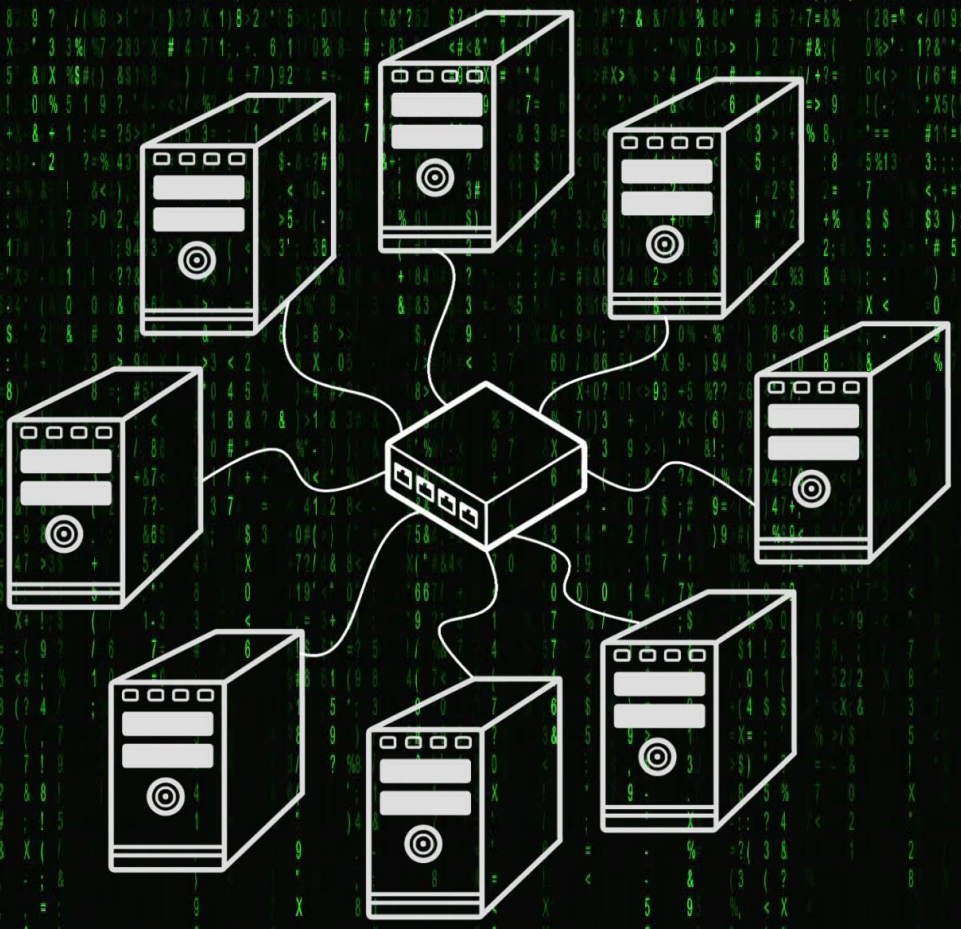
CDD: 004.07
CDU: 681.3

Forma Educacional Editora
CNPJ: 35.335.163/0001-00
Telefone: +55 (37) 99855-6001
www.formaeducacional.com.br
formaeducacional@gmail.com
Formiga - MG
Catálogo Geral: https://editoras.grupomultiatual.com.br/

Acesse a obra originalmente publicada em:
https://www.formaeducacional.com.br/2025/09/montagem-de-um-cluster-beowulf.html

Montagem de um Cluster Beowulf

Primeiros Passos



Autores

João Fernandes Santos Filho

José Aprígio Carneiro Neto

2025

PREFÁCIO

Estamos em uma era em que os dados crescem em uma velocidade assustadora, e com eles, crescem os desafios de processá-los, com inteligência, agilidade e propósito. Foi nesse contexto que iniciei minha caminhada pelo mundo dos *Clusters* e, posteriormente, mergulhei na compreensão das bibliotecas de paralelismo, como o *Apache Hadoop*. O que encontrei foi muito mais do que tecnologia, mas, uma nova maneira de pensar, de dividir e de cooperar.

Este livro nasceu dessa experiência, da vontade de compreender, de montar algo do zero, de ver máquinas que antes operavam sozinhas começarem a cooperarem entre si, de noites em claro tentando entender por que um nó (*node*) não respondia, de testes incansáveis com arquivos gigantescos, de euforia ao ver tarefas serem divididas e resolvidas em paralelo, como uma orquestra de máquinas em perfeita harmonia.

A primeira vez que vi um *cluster* funcionando, senti como se estivesse presenciando uma conversa silenciosa entre computadores. E mais do que isso, a uma orquestra de nós (*nodes*) trocando informações, dividindo tarefas, resolvendo problemas em conjunto. Foi aí que percebi que havia algo quase poético na computação paralela e distribuída.

A Arquitetura *de Clusters Beowulf* me chamou atenção não pela complexidade, mas pela proposta de usarmos o que temos ao nosso alcance, de reaproveitar e conectar máquinas novas e até mesmo máquinas que estão obsoletas, muitas vezes esquecidas, na maioria dos laboratórios de computação das instituições, em outras palavras, fazer mais com menos. E quando isso se junta com a filosofia da biblioteca *Apache Hadoop* — uma plataforma aberta, robusta, confiável, escalável e tolerante a falhas, feita para trabalhar com o caos (problemas complexos) e com um grande volume de dados (*Big Data*) — tudo ganha uma escala maior e mais sentido.

Portanto, este livro não é apenas um manual técnico. É um convite a uma reflexão, a uma narrativa construída a partir da prática, das tentativas frustradas, das pequenas vitórias, das madrugadas tentando entender porque um processo travou ou porque o *namenode* se recusava a responder. Além disso, trata-se de uma carta aberta a quem, como eu, já se perguntou: “Será que consigo montar um sistema distribuído com poucos recursos?” — a resposta, como você verá neste livro, é sim. E acredite, bem mais simples do que você imagina.

Mas, eu não quero aqui te convencer apenas com argumentos técnicos. Quero te lembrar que, por trás da computação paralela e distribuída, há um princípio profundamente humano: a da colaboração. O *Apache Hadoop*, com o seu *MapReduce* e seu sistema de arquivos distribuídos (*HDFS*), não resolve tudo sozinho — assim como nós, precisa de ajuda para resolver problemas complexos. Mas quando dividimos uma tarefa, onde cada nó (node) do *cluster* faz a sua parte, o impossível começa a acontecer.

Por isso, ao ler este livro, convido você a ir além das linhas de comando. Preste atenção nas relações entre os componentes do sistema (*cluster*). Nos detalhes que unem os nós (*nodes*) de um *cluster* e, principalmente, da colaboração entre eles. Nos pontos de falha que nos forçam a pensar melhor e refletir sobre uma solução. Dessa forma, a montagem de um *Cluster Beowulf*, com a utilização da biblioteca *Apache Hadoop*, não é só sobre desempenho. É sobre como pensamos soluções em rede — na tecnologia e na vida.

Que este livro possa te ajudar a montar, testar, errar e aprender. E que, ao final, você não só tenha um *cluster* rodando, mas também um novo olhar sobre o que significa construir algo em conjunto e, principalmente, sobre o que é o verdadeiro aprendizado das coisas.

Seja você também um pesquisador, um curioso apaixonado pela computação paralela e distribuída, por arquitetura de computadores. Vamos juntos construir, conectar, inspirar e transformar.

José Aprígio Carneiro Neto

SUMÁRIO

Capítulo 1

ARQUITETURAS EM *CLUSTER*

Uma das alternativas economicamente viáveis para explorar o paralelismo de *hardware* e suprir a demanda computacional pelo elevado poder de processamento, se dá por meio da construção de sistemas fracamente acoplados, que realizam o processamento paralelo por meio da agregação de dois ou mais computadores, conectados entre si através de uma rede local de computadores, denominados de *clusters*.

Os *clusters* são sistemas formados pela junção de computadores, homogêneos e/ou heterogêneos, onde cada computador possui seu próprio processador e memória, comportando-se como um sistema único e coerente, cujo objetivo é trabalhar de forma colaborativa e distribuída na execução de tarefas paralelas.

Para montar um *cluster* é preciso ter um *hardware* com vários núcleos no seu processador, um sistema operacional robusto e uma biblioteca de comunicação que faça a comunicação entre os nós (*nodes*) do sistema, para efetuar o paralelismo das tarefas.

Durante a montagem de um *cluster* é possível escolher várias combinações de *hardware*, mas, é essencial utilizar componentes que possuam o desempenho adequado para as tarefas que serão realizadas. Dentre os componentes de *hardware* utilizados na implementação de um *cluster*, alguns possuem um destaque maior, pois influenciam diretamente no seu desempenho, são eles: a memória de armazenamento (memória secundária), que precisa ter altas velocidades de leitura e escrita, além de uma boa capacidade de armazenamento, visto que, será responsável pelo armazenamento de grandes volumes de dados (*Big Data*); a memória *RAM* (*Random-Access Memory*), que precisa ter frequências elevadas para não prejudicar o desempenho do processador e, consequentemente, do *cluster*, tendo em vista que serão responsáveis pelo armazenamento temporário dos processos durante a execução dos algoritmos de *benchmark*; a placa-mãe (*motherboard*), que precisa

suportar a frequência máxima que as memórias *RAM* trabalham; e por fim, o processador, que precisa possuir, obrigatoriamente, vários núcleos (*multicore*), além de uma boa frequência de processamento.

Em um *cluster*, existe um computador principal, o nó mestre (*master*), responsável por coordenar a execução das tarefas e controlar os outros computadores do sistema, além dos nós escravos (*slaves*), que trabalham entre si para a realização das tarefas paralelas.

No entanto, para tirar o máximo de eficiência do poder de processamento das máquinas que compõem um *cluster*, torna-se indispensável o uso de uma biblioteca de programação paralela (*software*), que trabalhe de forma simultânea e distribuída na execução de tarefas complexas, envolvendo grandes volumes de dados, atuando de forma eficiente no gerenciamento dos processos, sendo confiável, escalável e tolerante a falhas.

Porém, a implementação de um *cluster* deverá levar em consideração o tipo de aplicação que será executada, tendo em vista que o seu desempenho não é tão eficiente no caso da resolução de problemas menos complexos, que envolvem uma granulosidade fina (divisão de um sistema em partes pequenas). Isso ocorre devido a troca elevada de informações entre os nós (*nodes*) do *cluster*, que provoca um aumento significativo no tráfego da rede, interferindo de forma direta no tempo de execução das tarefas, comprometendo o objetivo proposto pelo paralelismo, que é o ganho de tempo na execução das atividades.

Na Computação de Alto Desempenho (*High-Performance Computing – HPC*), os *clusters* são amplamente utilizados por possuírem acesso restrito e dedicado, além de ser uma arquitetura confiável, escalável e flexível, que possui uma série de benefícios, tais como: desempenho, expansibilidade, baixo custo, alta disponibilidade, balanceamento de carga e tolerância a falhas.

Atualmente, são identificados diferentes tipos de *clusters* na literatura, onde cada modelo possui suas próprias características e situações em que podem ser utilizados, são eles: de alto desempenho (com elevado poder de processamento, permitindo a execução de algoritmos complexos, que envolvem grandes volumes de dados); de balanceamento de carga (com o objetivo de dividir todas as tarefas de forma igualitária entre os recursos de *hardware* disponíveis no sistema, diminuindo o

tempo de resposta das tarefas); e de alta disponibilidade (com o objetivo de garantir o maior tempo de disponibilidade do sistema, onde as *nodes* podem ser substituídas por outras, caso apresentem falhas).

Neste livro, será abordado o desenvolvimento de um *Cluster Beowulf* Virtualizado, formado por computadores pessoais (*PCs*), não especializados, com o objetivo de auxiliar os estudantes da área de Computação a compreenderem, na prática, as arquiteturas paralelas, bem como o funcionamento das bibliotecas de paralelismo, sem a necessidade de utilização de equipamentos de alto custo de aquisição e alto consumo de energia, como supercomputadores.

Capítulo 2

MONTANDO UM *CLUSTER BEOWULF VIRTUALIZADO*

Neste capítulo, será descrito o processo de configuração de um *Cluster Beowulf*, utilizando Máquinas Virtuais (*VMs*) e a biblioteca de paralelismo *Apache Hadoop*.

Para o desenvolvimento do *cluster* foram utilizados três computadores pessoais (*PCs*), sendo um deles configurado como nó (*node*) mestre (*master*) e outros dois como nós (*nodes*) escravos (*slaves*).

Ao longo deste capítulo, serão abordadas as etapas de criação das máquinas virtuais (*VMs*), a instalação dos sistemas operacionais utilizados, a configuração do *cluster*, além da configuração da biblioteca de paralelismo *Apache Hadoop*.

2.1 CONFIGURAÇÃO DO *CLUSTER*

2.1.1 Criação das Máquinas Virtuais (*Virtual Machines - VMs*)

Nesta seção, será descrito o processo de criação das máquinas virtuais (*VMs*) utilizadas no desenvolvimento do *cluster*. Para isso, será necessário utilizar um *software* de virtualização, responsável por simular os recursos de *hardware* de um computador físico, possibilitando a criação e a execução de sistemas operacionais independentes, em um mesmo ambiente.

Atualmente, existem diversos tipos de *softwares* de virtualização disponíveis no mercado, como *Oracle VirtualBox*, *Microsoft Hyper-V*, *VMware Workstation*, entre outros. No desenvolvimento deste *cluster*, será utilizado o *software Oracle VirtualBox*, na versão 7.1.8, por ser uma ferramenta gratuita, de fácil acesso, com configurações simples e intuitivas.

1º Passo: Baixar a *ISO* do *Debian* Versão *11.11*

Primeiramente, será preciso fazer o *download* da imagem *ISO* do sistema operacional *Debian*, versão 11.11, que poderá ser obtida, por exemplo, através da seguinte *URL*: https://cdimage.debian.org/mirror/cdimage/archive/11.11.0/amd64/iso-cd/debian-11.11.0-amd64-netinst.iso.

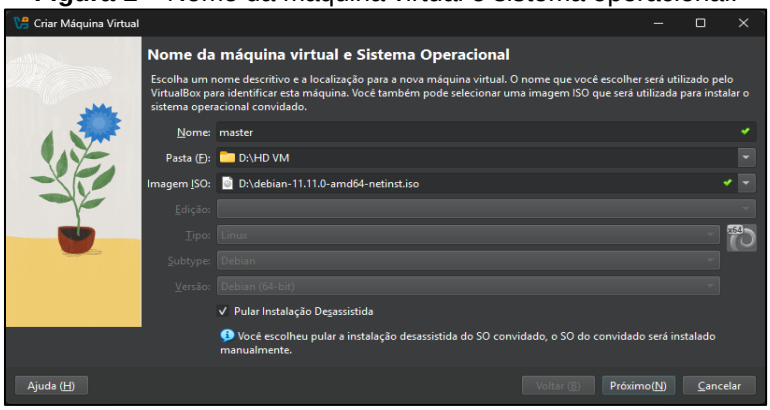
Apesar de existirem versões mais recentes, este livro utiliza a distribuição do *Linux*, *Debian* 11, devido à sua compatibilidade com a versão 1.11.0 do *Java*, que é um requisito obrigatório para a instalação e funcionamento adequado da biblioteca de paralelismo *Apache Hadoop*, na versão 3.4.1, utilizada no desenvolvimento deste *cluster*.

2º Passo: Criar as Máquinas Virtuais (*VMs*)

Para criar as máquinas virtuais (*VMs*), é preciso baixar e instalar o *software* ***VirtualBox*** e, em seguida, após a sua instalação e abertura, será preciso clicar na opção **Novo**, para iniciar o processo de criação de uma *VM*. Na janela **Criar Máquina Virtual**, defina o **Nome** da máquina, lembrando que neste livro, as máquinas virtuais (*VMs*) serão identificadas da seguinte forma: **master**, **slave1 e slave2**.

Na sequência, na opção **Imagem ISO**, selecione a imagem do sistema operacional que foi baixada, e marque a opção **Pular Instalação Desassistida**, como mostra a Figura 1 e, em seguida, clique em **Próximo**, para continuar com o processo de instalação da *VM*.

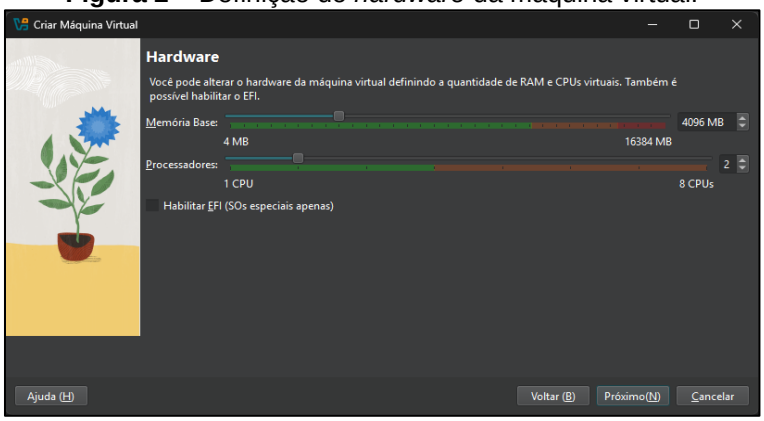
Figura 1 – Nome da máquina virtual e sistema operacional.



Fonte: Autores, 2025.

Na tela seguinte, defina a quantidade de **Memória *RAM*** (*4096 MB*) e de **Processadores** (2 - dois) de cada máquina virtual do *cluster* e, em seguida, clique em **Próximo**, para continuar com o processo, como mostra a Figura 2.

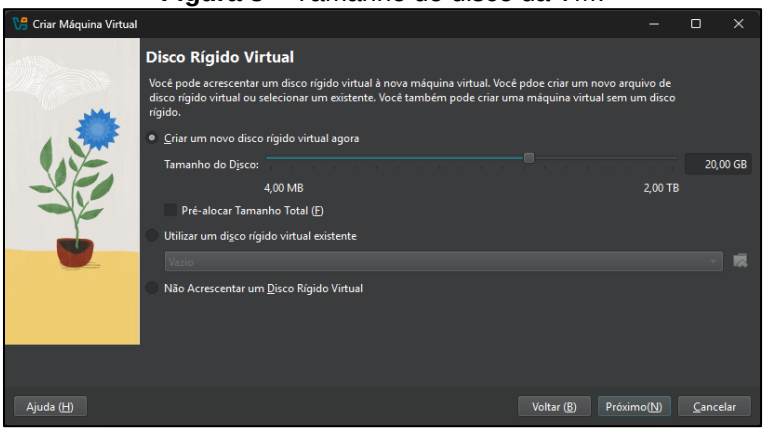
Figura 2 – Definição do *hardware* da máquina virtual.



Fonte: Autores, 2025.

Na próxima tela, defina o **Tamanho do Disco** (*20 GB*) e, na sequência, clique em **Próximo** e, em seguida, em **Finalizar**, para concluir com o processo de criação do disco da *VM*, como mostra a Figura 3.

Figura 3 – Tamanho do disco da *VM*.



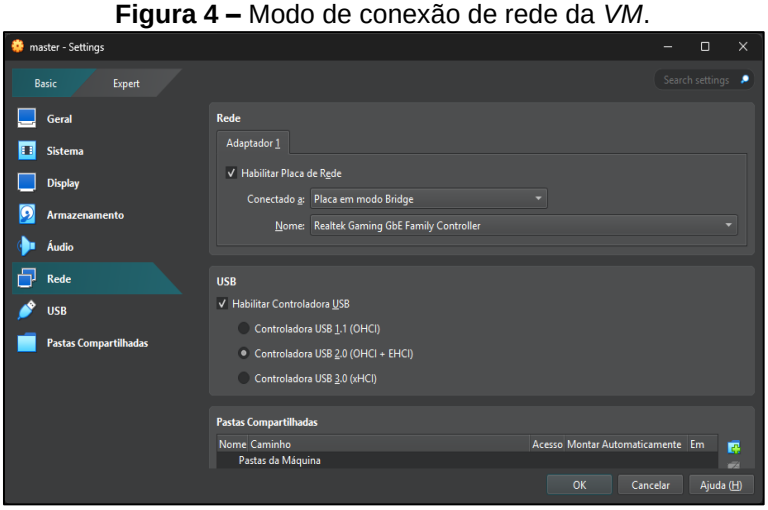
Fonte: Autores, 2025.

3º Passo: Alterar o modo de conexão de rede das Máquinas Virtuais

Em seguida, selecione a *VM* e, na sequência, clique na opção **Configurações**. Na janela seguinte, clique em **Rede** e, no campo **Conectado a**, depois, selecione a

opção **Placa em modo Bridge** e, por fim, clique em **OK**, para concluir essa etapa de configuração, como mostra a Figura 4.

Figura 4 – Modo de conexão de rede da *VM*.



Fonte: Autores, 2025.

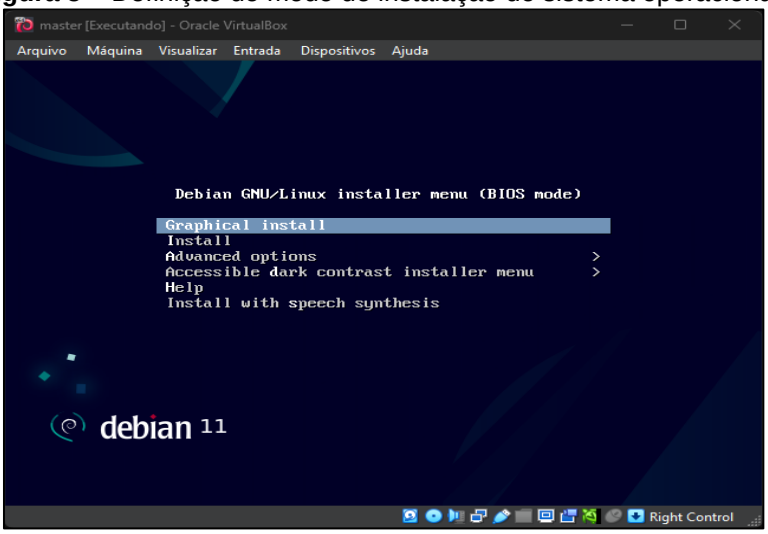
2.2 Instalação do Sistema Operacional

Nesta seção, será descrito o processo de instalação dos sistemas operacionais nas *VMs* do *cluster*.

1º Passo: Escolher o modo de instalação do sistema operacional

Para escolher o modo de instalação do sistema operacional, selecione a *VM* e, em seguida, clique em **Iniciar**. Após a inicialização da *VM*, escolha a opção **Graphical install**, utilizando as setas do teclado e, na sequência, pressione **Enter**, para iniciar o processo de instalação, como mostra a Figura 5.

Figura 5 – Definição do modo de instalação do sistema operacional.

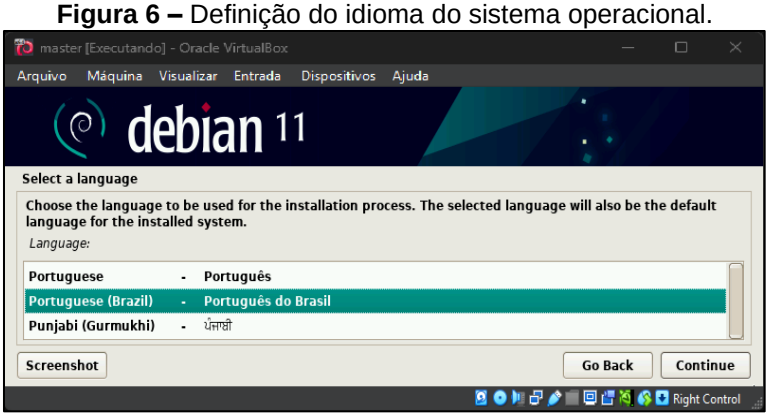


Fonte: Autores, 2025.

2º Passo: Escolher o idioma e a configuração do teclado do sistema operacional

Continuando o processo de instalação do sistema operacional, escolha o idioma, clicando na opção **Português do Brasil** e, em seguida, pressione **Enter** ou clique em **Continue**, para avançar no processo de instalação do sistema operacional, como mostra a Figura 6.

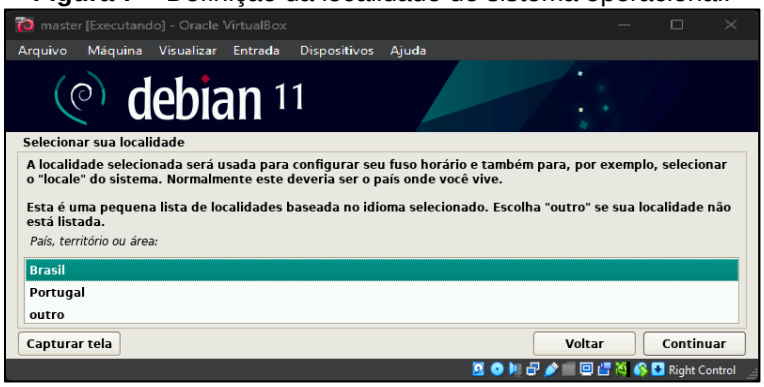
Figura 6 – Definição do idioma do sistema operacional.



Fonte: Autores, 2025.

Na sequência, escolha o país **Brasil**, como a localidade e, em seguida, pressione **Enter** ou clique em **Continuar**, para avançar no processo de configuração, como mostra a Figura 7.

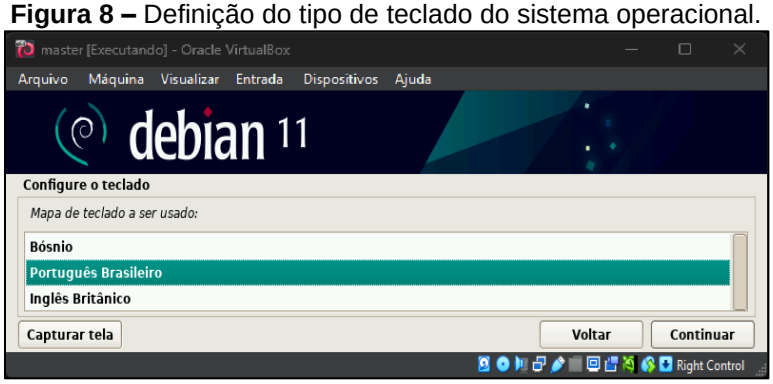
Figura 7 – Definição da localidade do sistema operacional.



Fonte: Autores, 2025.

Na tela seguinte, escolha o tipo de teclado, clicando na opção **Português Brasileiro** e, para finalizar o processo, pressione **Enter** ou clique em **Continuar**, como mostra a Figura 8.

Figura 8 – Definição do tipo de teclado do sistema operacional.

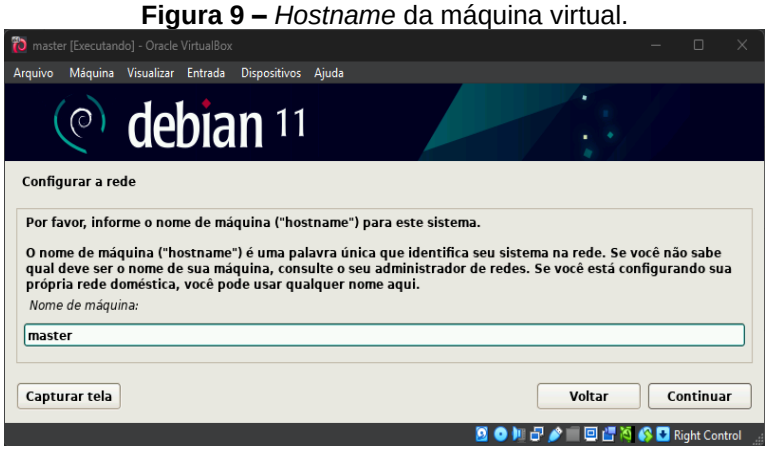


Fonte: Autores, 2025.

3º Passo: Definir o *Hostname* da Máquina Virtual (*VM*)

Na próxima tela de configuração das *VMs*, defina o ***hostname*** (nome da máquina virtual) para cada uma delas, lembrando que os nomes das *VMs* deste projeto serão: **master**, **slave1** e **slave2**. Após a inserção dos ***hostnames***, pressione **Enter** e, em seguida, pressione **Enter** novamente ou clique em **Continuar**, para avançar neste processo, como mostra a Figura 9.

Figura 9 – *Hostname* da máquina virtual.

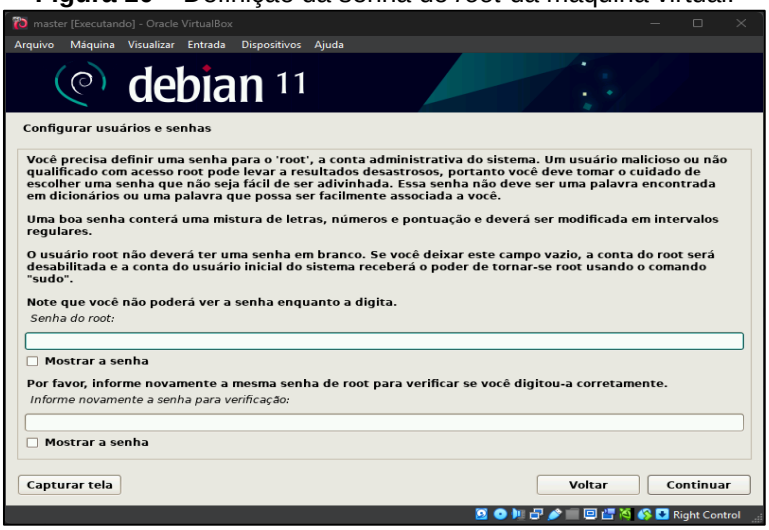


Fonte: Autores, 2025.

4º Passo: Definir a senha do *Root*

Neste passo, mantenha a **Senha do root** em branco e, em seguida, pressione **Enter** ou clique em **Continuar**, para seguir no processo de configuração das *VMs*, como mostra a Figura 10.

Figura 10 – Definição da senha do *root* da máquina virtual.

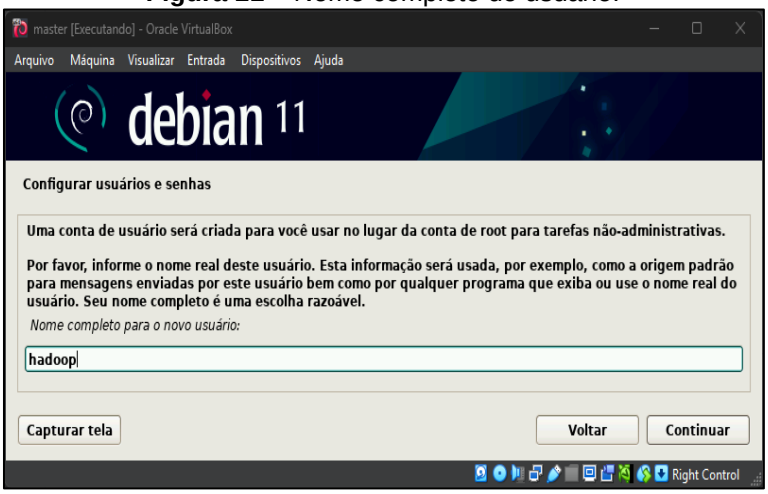


Fonte: Autores, 2025.

5º Passo: Definir o nome completo do usuário

Na tela seguinte, no campo referente ao **Nome completo do usuário**, insira o nome **hadoop** e, na sequência, pressione **Enter** ou clique em **Continuar**, para prosseguir com o processo de configuração da *VM*, como mostra a Figura 11.

Figura 11 – Nome completo do usuário.

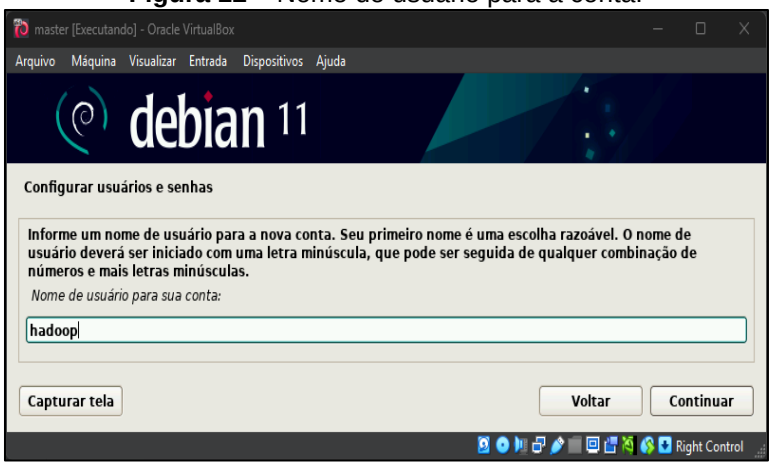


Fonte: Autores, 2025.

6º Passo: Definir o nome de usuários e senhas

Na sequência, na tela de configuração de usuários e senhas, no campo referente ao **Nome de usuário para sua conta**, insira o nome **hadoop** e, em seguida, pressione **Enter** ou clique em **Continuar**, para prosseguir com o processo de configuração do usuário, como mostra a Figura 12.

Figura 12 – Nome do usuário para a conta.

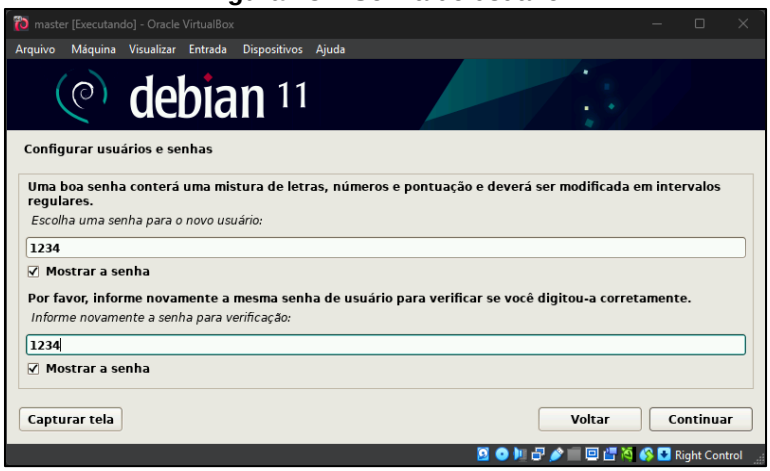


Fonte: Autores, 2025.

7º Passo: Definir a senha do usuário

Na próxima tela, é necessário definir uma senha para o usuário, portanto, no campo referente a escolha da senha para o novo usuário, digite **1234** (senha) e, na sequência, repita a senha utilizada, para verificação, por fim, pressione **Enter** ou clique em **Continuar**, para avançar nas configurações da *VM*, como mostra a Figura 13.

Figura 13 – Senha do usuário.

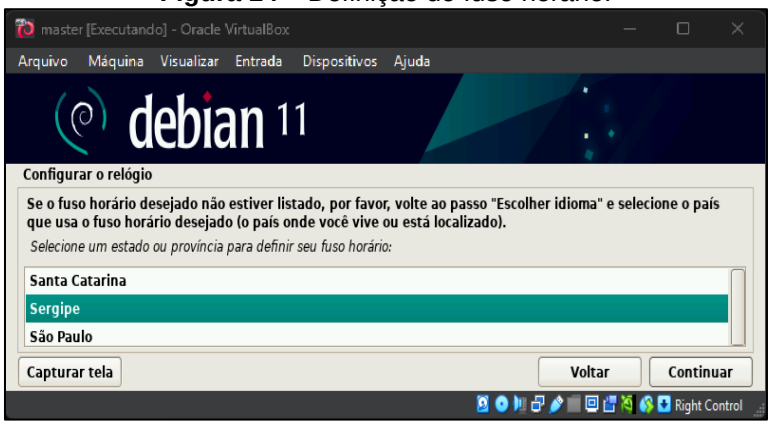


Fonte: Autores, 2025.

8º Passo: Escolher o fuso horário

Seguindo no processo de configuração da *VM*, escolha um estado (**Sergipe**) para definir o fuso horário da máquina e, em seguida, pressione **Enter** ou clique em **Continuar**, para avançar, como mostra a Figura 14.

Figura 14 – Definição do fuso horário.

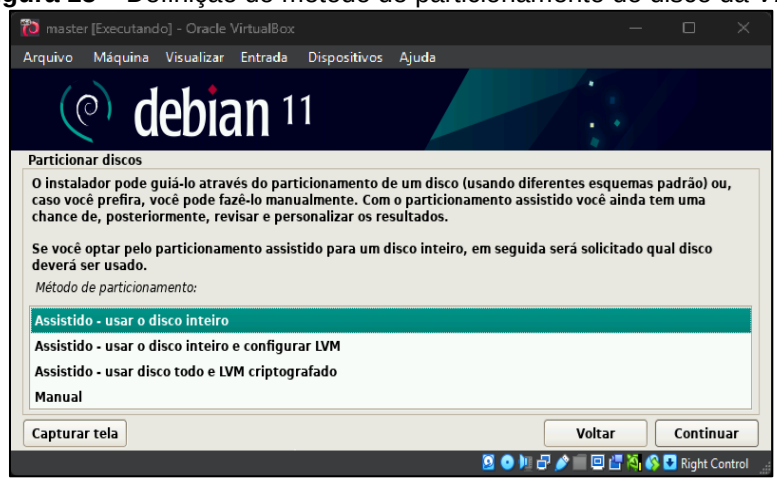


Fonte: Autores, 2025.

9º Passo: Realizar o particionamento do disco

No método de particionamento dos discos da *VM*, escolha a opção **Assistido - usar o disco inteiro** e, em seguida, pressione **Enter** ou clique em **Continuar**, como mostra a Figura 15.

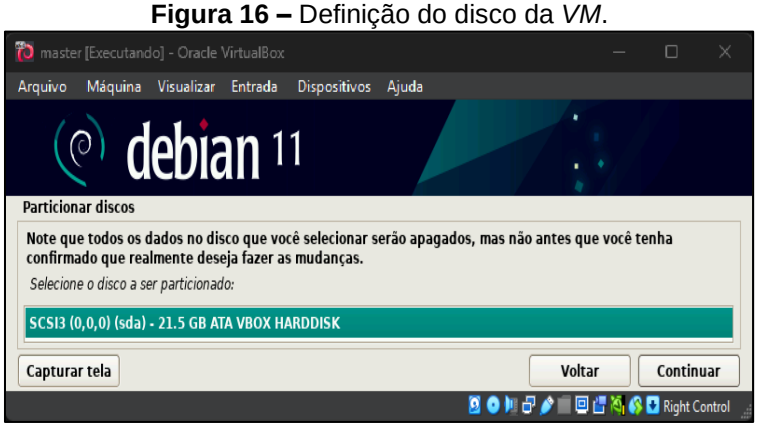
Figura 15 – Definição do método de particionamento do disco da *VM*.



Fonte: Autores, 2025.

Para selecionar o disco a ser utilizado pela *VM*, é preciso escolher o disco que foi definido anteriormente e, na sequência, pressionar **Enter** ou clicar em **Continuar**, para avançar, como mostra a Figura 16.

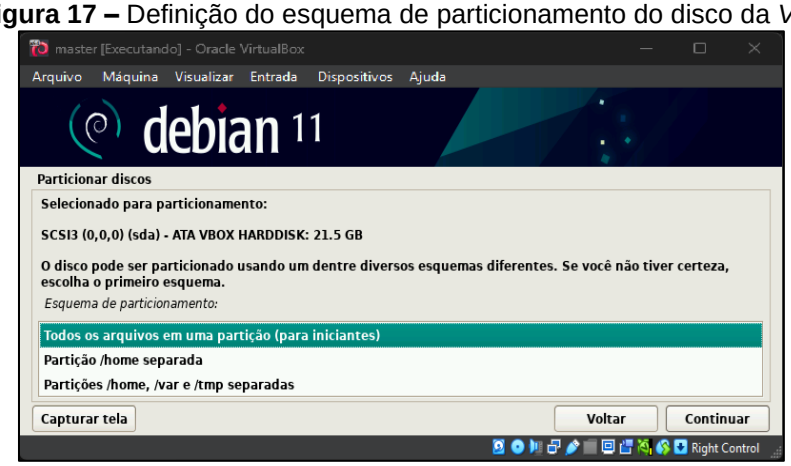
Figura 16 – Definição do disco da *VM*.



Fonte: Autores, 2025.

Em seguida, na opção do esquema de particionamento do disco, escolha **Todos os arquivos em uma partição** e, depois, pressione **Enter** ou clique em **Continuar**, para avançar, como mostra a Figura 17.

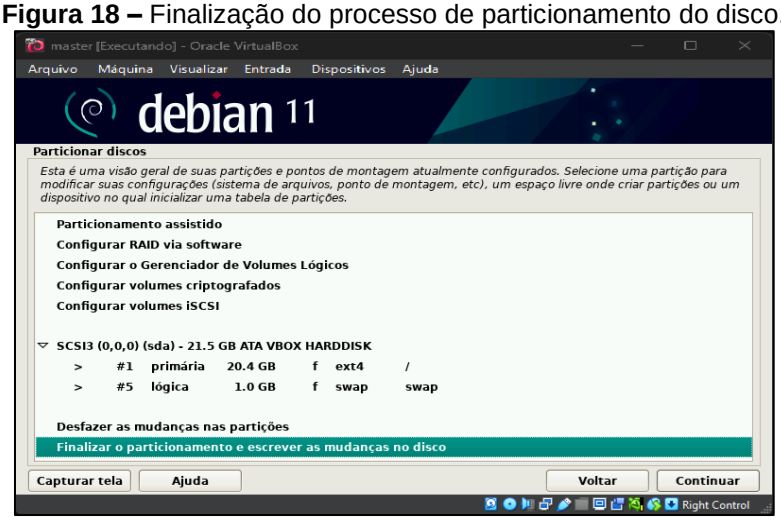
Figura 17 – Definição do esquema de particionamento do disco da *VM*.



Fonte: Autores, 2025.

Por fim, para finalizar o processo de particionamento do disco, escolha a opção **Finalizar o particionamento e escrever as mudanças no disco** e, na sequência, pressione **Enter** ou clique em **Continuar**, como mostra a Figura 18.

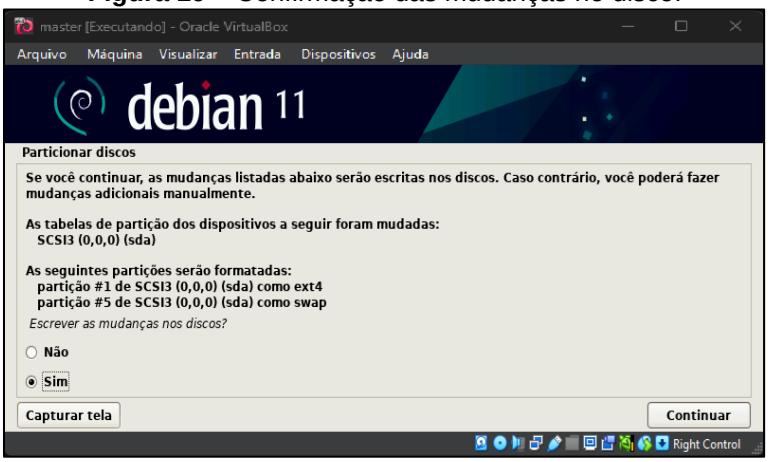
Figura 18 – Finalização do processo de particionamento do disco.



Fonte: Autores, 2025.

Continuando, escolha a opção **Sim**, para escrever as mudanças no disco e, depois, pressione **Enter** ou clique em **Continuar**, para finalizar o processo, como mostra a Figura 19.

Figura 19 – Confirmação das mudanças no disco.

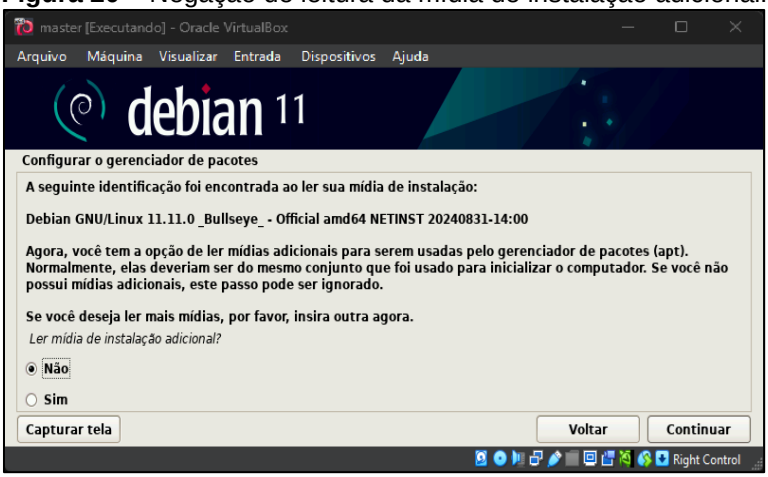


Fonte: Autores, 2025.

10º Passo: Escolher o repositório

Na tela seguinte, para escolher o repositório de pacotes de instalação da *VM*, escolha a opção **Não**, para indicar que não deseja ler uma mídia de instalação adicional e, em seguida, pressione **Enter** ou clique em **Continuar**, para avançar, como mostra a Figura 20.

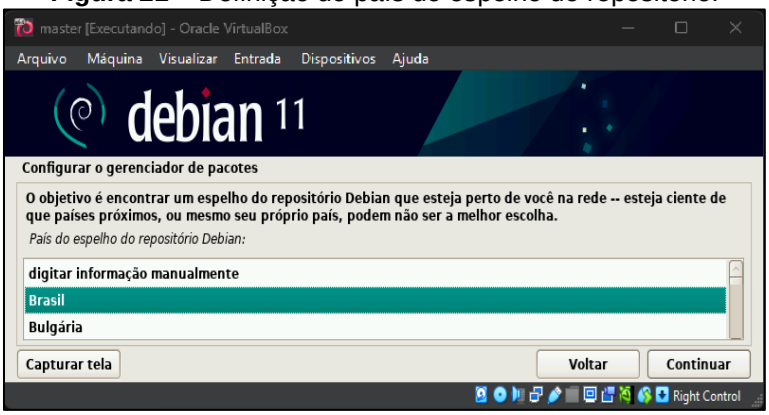
Figura 20 – Negação de leitura da mídia de instalação adicional.



Fonte: Autores, 2025.

Na sequência, escolha o **Brasil** como país do espelho do repositório e, em seguida, pressione **Enter** ou clique em **Continuar**, para definir o local do repositório, como mostra a Figura 21.

Figura 21 – Definição do país do espelho do repositório.



Fonte: Autores, 2025.

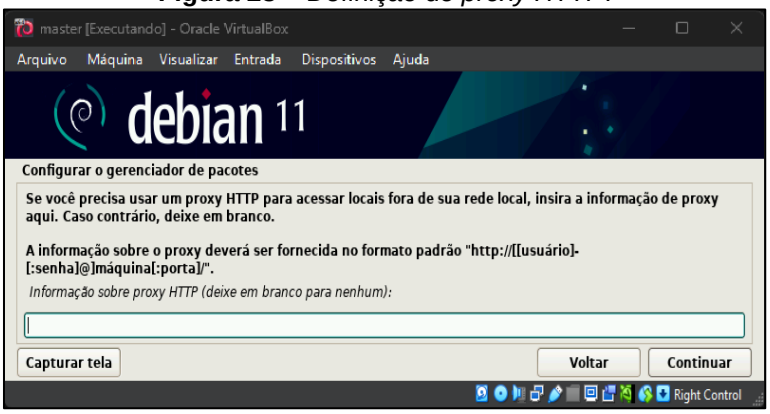
Na tela seguinte, escolha a opção **deb.debian.org**, como espelho do repositório e, em seguida, pressione **Enter** ou clique em **Continuar**, para seguir com o processo de definição do país do espelho, como mostra a Figura 22.

Figura 22 – Definição do espelho do repositório.

Fonte: Autores, 2025.

Na sequência, mantenha o campo **proxy HTTP** em branco e, em seguida, pressione **Enter** ou clique em **Continuar**, para avançar na configuração do *proxy*, como mostra a Figura 23.

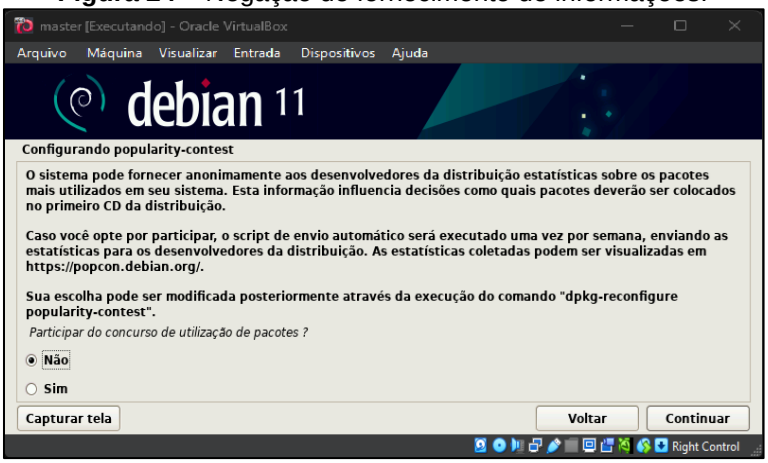
Figura 23 – Definição do *proxy HTTP*.



Fonte: Autores, 2025.

Em seguida, escolha a opção **Não,** para indicar que não deseja fornecer informações sobre os pacotes utilizados e, para finalizar, pressione **Enter** ou clique em **Continuar**, como mostra a Figura 24.

Figura 24 – Negação do fornecimento de informações.



Fonte: Autores, 2025.

11º Passo: Selecionar *softwares* adicionais

Na tela de **Seleção de *software***, utilize a tecla de **Espaço** para desmarcar todas as opções de *softwares* disponíveis, para que nenhum *software* adicional seja instalado e, em seguida, pressione **Enter** ou clique em **Continuar**, para finalizar esse processo, como mostra a Figura 25.

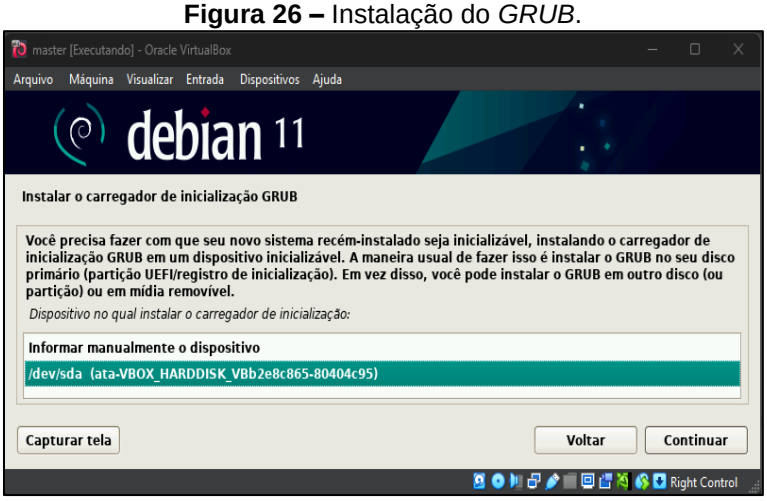
Figura 25 – *Softwares* adicionais para instalação.

Fonte: Autores, 2025.

12º Passo: Instalar o *GRUB*

Para **Instalar o carregador de inicialização GRUB**, escolha a opção **Sim**, para indicar que deseja instalar o *GRUB* (gerenciador de inicialização de sistemas operacionais) e, em seguida, pressione **Enter** ou clique em **Continuar**. Na sequência, escolha o disco em que o *GRUB* deverá ser instalado e, depois, pressione **Enter** ou clique em **Continuar**, para finalizar, como mostra a Figura 26.

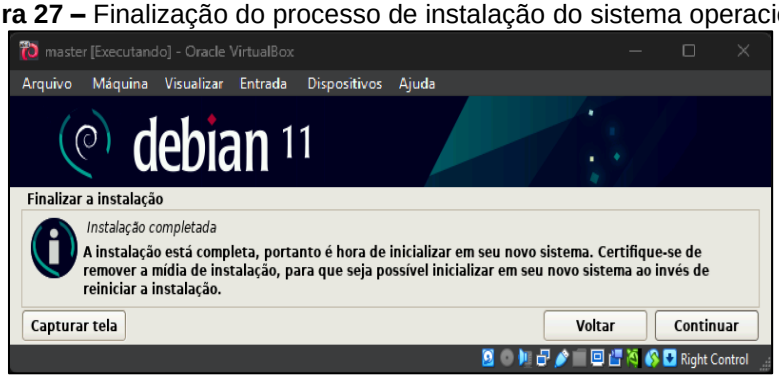
Figura 26 – Instalação do *GRUB*.



Fonte: Autores, 2025.

Na tela seguinte, pressione **Enter** ou clique em **Continuar**, para finalizar o processo de instalação do sistema operacional e reinicializar a *VM*, como mostra a Figura 27.

Figura 27 – Finalização do processo de instalação do sistema operacional.



Fonte: Autores, 2025.

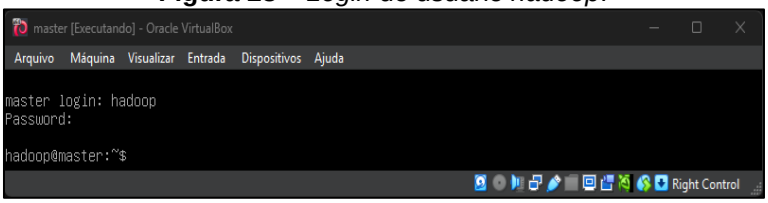
2.3 Configuração do *Apache Hadoop*

Nesta seção, será descrito o processo de configuração da biblioteca de paralelismo *Apache Hadoop* nas *VMs* do *cluster*. Essas configurações são essenciais para permitir o bom funcionamento do sistema.

1º Passo: Entrar na conta do usuário *hadoop*

Após o processo de reinicialização da *VM*, selecione a *VM master* e, em seguida, clique em **Iniciar**. Depois do *start* da *VM master*, entre no modo terminal e, na sequência, digite o nome do usuário e senha (usuário: **hadoop** e, em seguida, pressione **Enter** / senha: **1234** e, na sequência, pressione **Enter** novamente). Ao digitar a senha, não será possível visualizá-la, por uma questão de segurança do sistema operacional *Linux*, como mostra a Figura 28.

Figura 28 – *Login* do usuário *hadoop*.

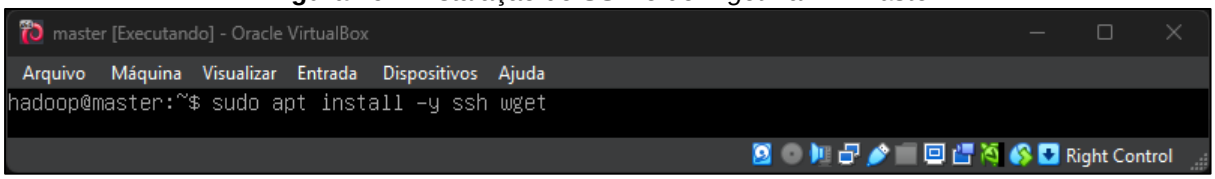


Fonte: Autores, 2025.

2º Passo: Instalar o *SSH* e o *Wget*

Para realizar a instalação do *SSH* (*Secure Shell*) e do *Wget* (utilitário de código aberto utilizado para baixar arquivos da internet), no *prompt* (terminal) de comando da *VM master*, execute a seguinte linha de comando: **sudo apt install -y ssh wget**, como mostra a Figura 29.

Figura 29 – Instalação do *SSH* e do *Wget* na *VM master*.



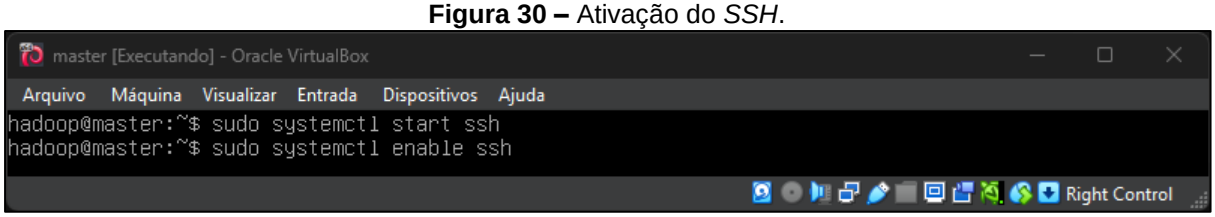
Fonte: Autores, 2025.

Como a linha de comando utiliza o comando **sudo**, para ser executado com permissões de administrador da máquina, será necessário digitar a senha (1234) do usuário de *root*. Lembrando que durante um período de 15 minutos, não será necessário digitar a senha novamente para executar comandos que envolva esse usuário. Além disso, essa linha de comando deverá ser executada nas três *VMs* do *cluster*.

3º Passo: Iniciar e ativar o *SSH*

Para inicializar o serviço *SSH*, execute a seguinte linha de comando no modo terminal da *VM master*: **sudo systemctl start ssh**. Em seguida, execute a linha de comando: **sudo systemctl enable ssh**, para ativar o serviço *SSH*, como mostra a Figura 30. Lembrando que esses dois comandos deverão ser executados nas três *VMs* do *cluster*.

Figura 30 – Ativação do *SSH*.

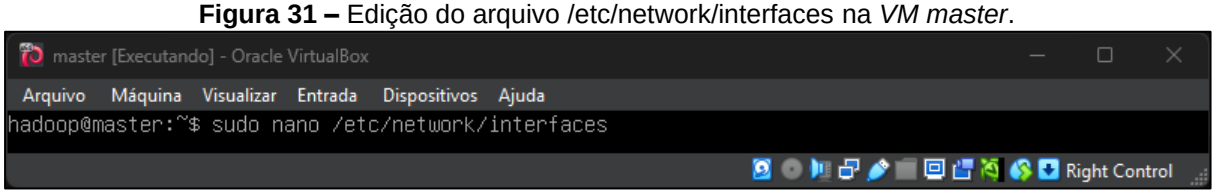


Fonte: Autores, 2025.

4º Passo: Alterar os endereços *IP* das *VMs*

No modo terminal de cada *VM*, execute a seguinte linha de comando: **sudo nano /etc/network/interfaces**, para abrir o editor de texto e alterar os endereços *IPs* das respectivas *VMs*, como mostra a Figura 31.

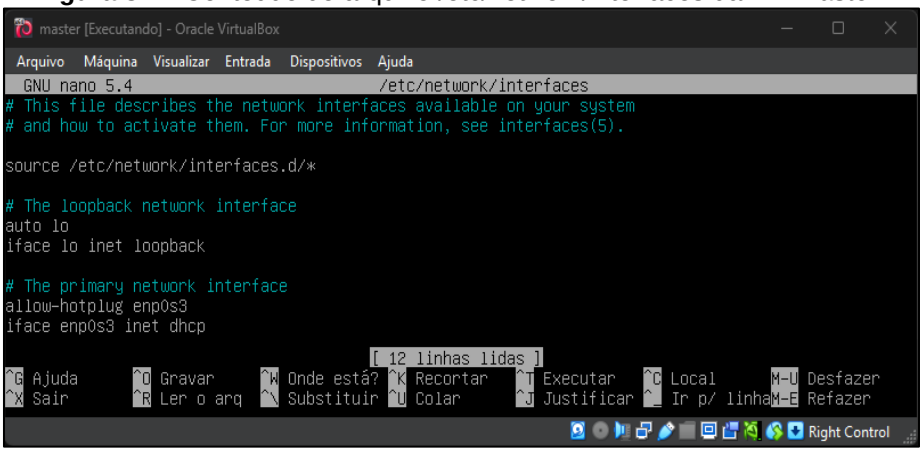
Figura 31 – Edição do arquivo /etc/network/interfaces na *VM master*.



Fonte: Autores, 2025.

Na sequência, será possível visualizar o conteúdo do arquivo ***interfaces***. Para isso, utilize as setas do teclado para posicionar o cursor e, em seguida, remova as duas últimas linhas desse arquivo, como mostra a Figura 32.

Figura 32 – Conteúdo do arquivo /etc/network/interfaces da *VM master*.

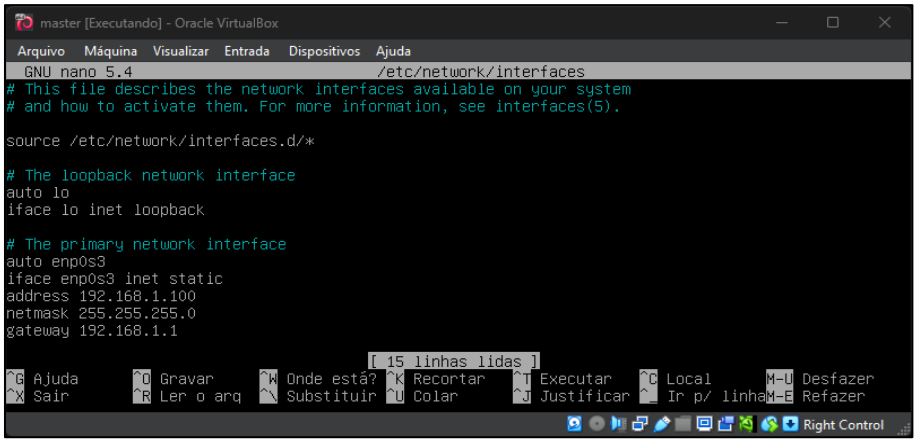


Fonte: Autores, 2025.

Dando continuidade ao processo de configuração dos endereços *IPs* nas *VMs*, posicione o cursor no final do arquivo e digite as informações de configuração abaixo, como mostra a Figura 33.

```
auto enp0s3
iface enp0s3 inet static
address 192.168.1.100
netmask 255.255.255.0
gateway 192.168.1.1
```

Figura 33 – Conteúdo atualizado do arquivo /etc/network/interfaces da *VM master*.

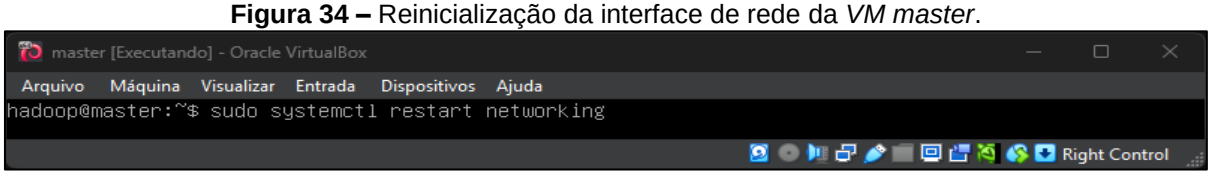


Fonte: Autores, 2025.

Após adicionar as informações, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

Para aplicar as configurações que foram alteradas, execute a seguinte linha de comando no modo terminal de cada *VM*: **sudo systemctl restart networking**, como mostra a Figura 34.

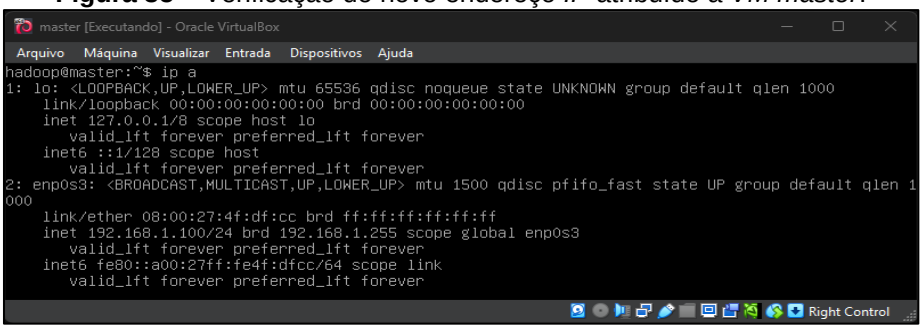
Figura 34 – Reinicialização da interface de rede da *VM master*.



Fonte: Autores, 2025.

Na sequência, execute a seguinte linha de comando no modo terminal da *VM master,* para verificar se as configurações do endereço *IP* foram alteradas corretamente: **ip a**, como mostra a Figura 35.

Figura 35 – Verificação do novo endereço *IP* atribuído a *VM master*.



Fonte: Autores, 2025.

Lembrando que as alterações dos endereços *IPs* deverão ocorrer em todas as *VMs* do *cluster* (*master* e *slaves*), utilizando os respectivos endereços *IPs* estabelecidos pelo projeto: **master: 192.168.1.100 / slave1: 192.168.1.101 / slave2: 192.168.1.102.**

5º Passo: Acessar remotamente as *VMs* utilizando o *SSH*

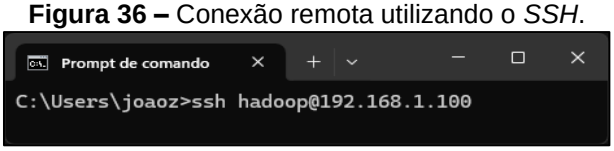
No seu computador de trabalho, abra o modo terminal (*prompt*) e execute a seguinte linha de comando: **ssh hadoop@192.168.1.100**. Na sequência, digite **yes** e pressione **Enter**, para confirmar que deseja se conectar remotamente a *VM master*.

Em seguida, será necessário digitar a senha do usuário *hadoop*, que foi definida durante o processo de instalação. Ressaltando que esse processo deverá ser realizado nas três *VMs* do *cluster*, utilizando seus respectivos endereços *IPs*, a fim de testar a conexão remota entre elas.

A conexão remota, utilizando o modo terminal do computador, facilitará a execução dos próximos passos de configuração da biblioteca *Apache Hadoop*, permitindo copiar e colar os comandos de configuração, de forma a agilizar o processo de configuração. Para encerrar a conexão remota, basta digitar, no modo terminal de cada *VM*, o comando **exit** e, em seguida, pressionar **Enter**.

A Figura 36 mostra a execução do comando para a conexão remota com a *VM master*, a partir da máquina de trabalho do usuário que está configurando o *cluster*.

Figura 36 – Conexão remota utilizando o *SSH*.



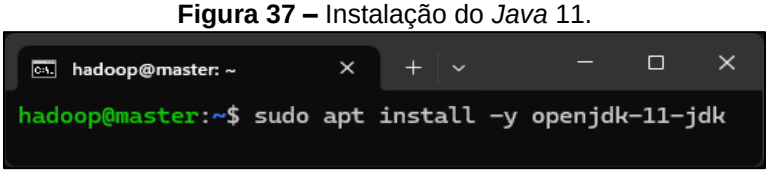
Fonte: Autores, 2025.

6º Passo: Instalar o *Java* 11

Continuando a configuração do *cluster*, a partir da sua máquina de trabalho, acesse remotamente as *VMs*, uma a uma e, em seguida, execute a seguinte linha de comando no terminal de cada uma delas: **sudo apt install -y openjdk-11-jdk**, como

mostra a Figura 37. Lembrando que esse comando deverá ser executado nas três *VMs* do *cluster*.

Figura 37 – Instalação do *Java* 11.



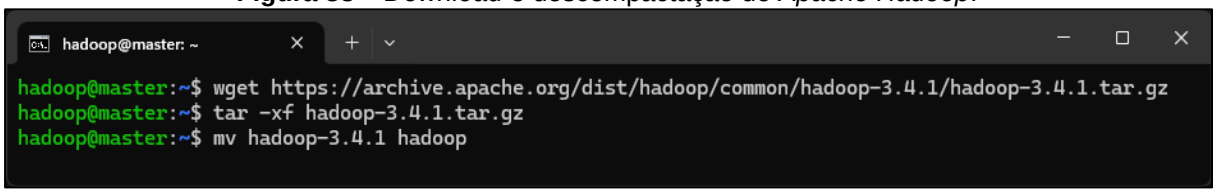
Fonte: Autores, 2025.

7º Passo: Baixar o *Apache Hadoop* 3.4.1

Para baixar o *Apache Hadoop*, acesse remotamente cada *VM* e, no modo terminal de cada uma delas, execute a seguinte linha de comando: **wget https://archive.apache.org/dist/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz**.

Em seguida, para descompactar o arquivo que foi baixado, execute a seguinte linha de comando no modo terminal de cada *VM*: **tar -xf hadoop-3.4.1.tar.gz**. Na sequência, execute o seguinte comando: **mv hadoop-3.4.1 hadoop**, para renomear a pasta. Os comandos utilizados podem ser visualizados na Figura 38.

Figura 38 – *Download* e descompactação do *Apache Hadoop*.

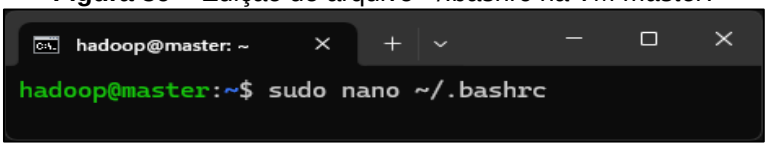


Fonte: Autores, 2025.

8º Passo: Definir a variável de ambiente *HADOOP_HOME* e editar o *PATH*

Na sequência, acesse remotamente, via *SSH*, cada uma das três *VMs* e, execute o seguinte comando no modo terminal: **sudo nano ~/.bashrc**, como mostra a Figura 39.

Figura 39 – Edição do arquivo ~/.bashrc na *VM master*.

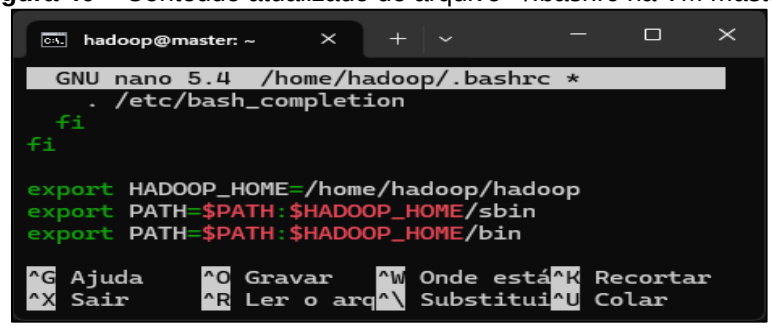


Fonte: Autores, 2025.

Após abrir o editor de texto, posicione o cursor no final do arquivo e digite as informações abaixo, como mostra a Figura 40.

```
export HADOOP_HOME=/home/hadoop/hadoop
export PATH=$PATH:$HADOOP_HOME/sbin
export PATH=$PATH:$HADOOP_HOME/bin
```

Figura 40 – Conteúdo atualizado do arquivo ~/.bashrc na *VM master*.

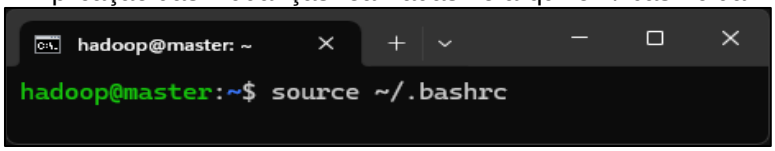


Fonte: Autores, 2025.

Depois de adicionar as informações no arquivo, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

Na sequência, execute a seguinte linha de comando no modo terminal de cada *VM*, para aplicar as modificações realizadas no arquivo: **source ~/.bashrc**, como mostra a Figura 41.

Figura 41 – Aplicação das mudanças realizadas no arquivo ~/.bashrc da *VM master*.



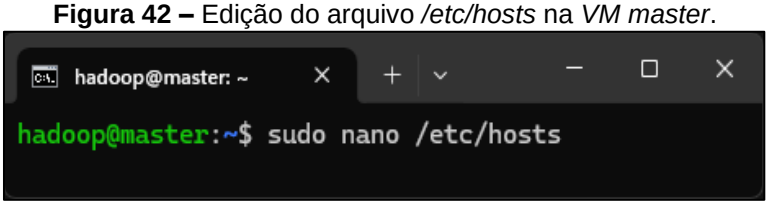
Fonte: Autores, 2025.

Lembrando que a alteração do arquivo **~/.bashrc** deverá ser realizada nas três *VMs* do *cluster*.

9º Passo: Editar o arquivo */etc/hosts*

Para editar o arquivo ***/etc/hosts***, acesse remotamente cada *VM*, via *SSH* e, no modo terminal de cada uma delas, execute a seguinte linha de comando: **sudo nano /etc/hosts**, para abrir o arquivo **hosts** e editá-lo, como mostra a Figura 42.

Figura 42 – Edição do arquivo */etc/hosts* na *VM master*.

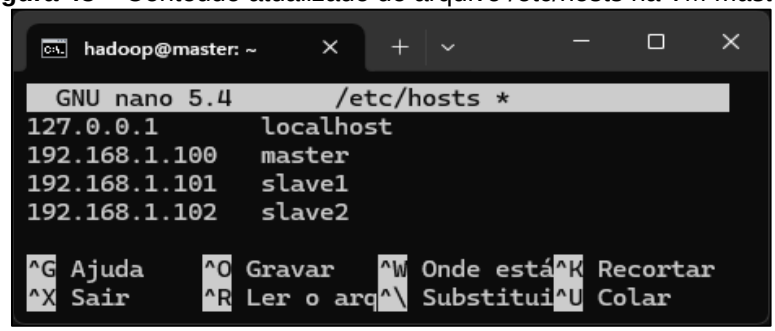


Fonte: Autores, 2025.

Com o editor de texto aberto, apague todo o conteúdo do arquivo e digite as informações abaixo, como mostra a Figura 43.

```
127.0.0.1       localhost
192.168.1.100   master
192.168.1.101   slave1
192.168.1.102   slave2
```

Figura 43 – Conteúdo atualizado do arquivo /etc/hosts na *VM master*.



Fonte: Autores, 2025.

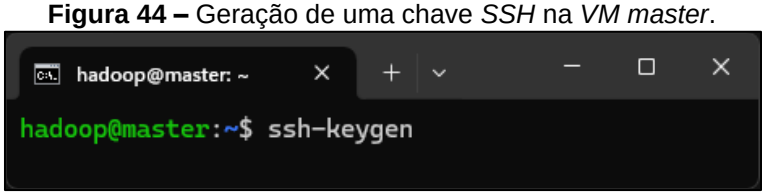
Após adicionar as informações, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

Lembrando que a edição do arquivo **/etc/hosts** deverá ser realizada em todas as *VMs* do *cluster* (*master* e *slaves*).

10º Passo: Gerar uma chave *SSH*

Na sequência, é preciso gerar uma chave *SSH*, para isso, execute a seguinte linha de comando no modo terminal da *VM master*: **ssh-keygen**, como mostra a Figura 44.

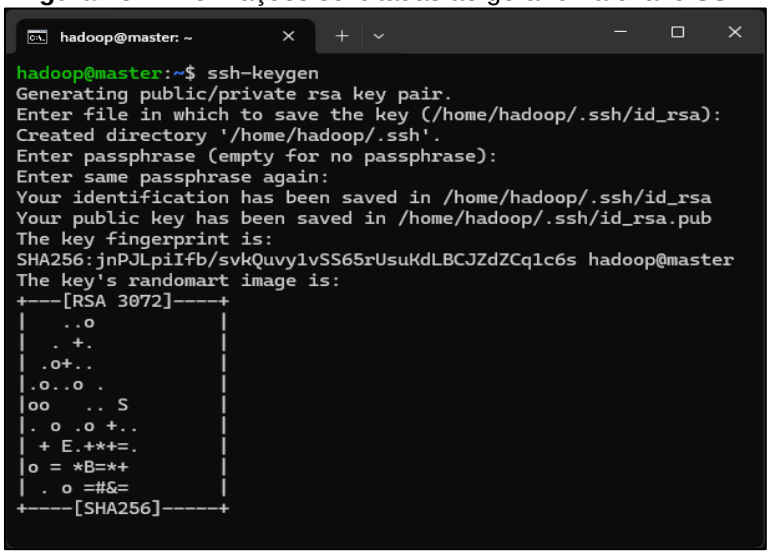
Figura 44 – Geração de uma chave *SSH* na *VM master*.



Fonte: Autores, 2025.

Em seguida, serão solicitadas algumas informações, como o local para salvar a **chave** e uma **senha** de proteção. No entanto, basta pressionar **Enter** em todas as solicitações, para aceitar os valores predefinidos, como mostra a Figura 45.

Figura 45 – Informações solicitadas ao gerar uma chave *SSH*.

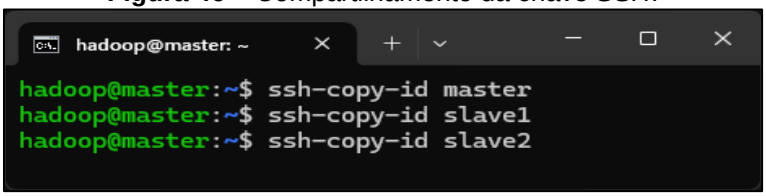


Fonte: Autores, 2025.

11º Passo: Compartilhar a chave *SSH*

Na *VM master*, no modo terminal, execute as seguintes linhas de comandos para compartilhar a chave *SSH*: **ssh-copy-id master**, **ssh-copy-id slave1** e **ssh-copy-id slave2**, respectivamente, como mostra a Figura 46.

Figura 46 – Compartilhamento da chave *SSH*.



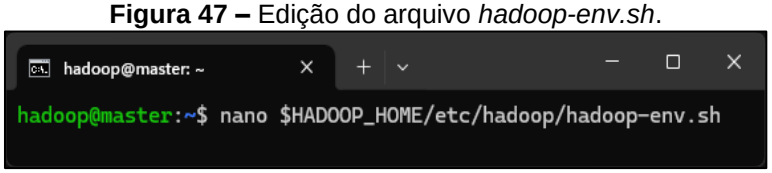
Fonte: Autores, 2025.

Durante esse processo, será necessário digitar "**yes**" e, em seguida, pressionar **Enter** para confirmar que deseja se conectar remotamente na *VM*, bem como digitar a senha do usuário ***hadoop***.

12º Passo: Editar o arquivo *hadoop-env.sh*

Na sequência, é necessário editar o arquivo ***hadoop-env.sh***, para isso, execute a seguinte linha de comando no modo terminal da *MV master*: **nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh**, como mostra a Figura 47.

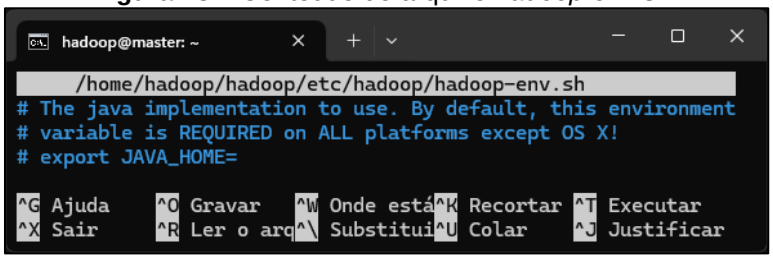
Figura 47 – Edição do arquivo *hadoop-env.sh*.



Fonte: Autores, 2025.

Após abrir o editor de texto, pressione **CTRL+W** e digite a seguinte linha de comando: **export JAVA_HOME** e, em seguida, pressione **Enter** para localizar a linha que deverá ser editada, como mostra a Figura 48.

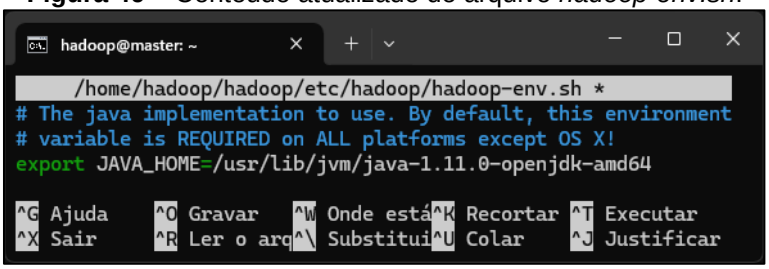
Figura 48 – Conteúdo do arquivo *hadoop-env.sh*.



Fonte: Autores, 2025.

Em seguida, remova o sinal “#” do início da linha e, após o sinal de “**=**”, digite a seguinte linha de comando: **/usr/lib/jvm/java-1.11.0-openjdk-amd64**, como mostra a Figura 49.

Figura 49 – Conteúdo atualizado do arquivo *hadoop-env.sh*.



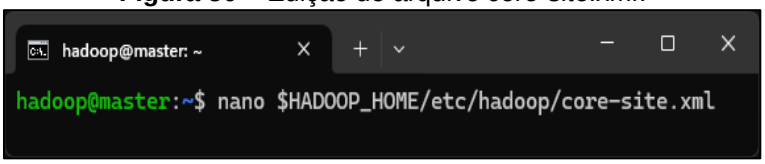
Fonte: Autores, 2025.

Após adicionar a informação no arquivo, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

13º Passo: Editar o arquivo *core-site.xml*

Para editar o arquivo ***core-site.xml***, no modo terminal da *VM master*, execute a seguinte linha de comando: **nano $HADOOP_HOME/etc/hadoop/core-site.xml**, como mostra a Figura 50.

Figura 50 – Edição do arquivo *core-site.xml*.

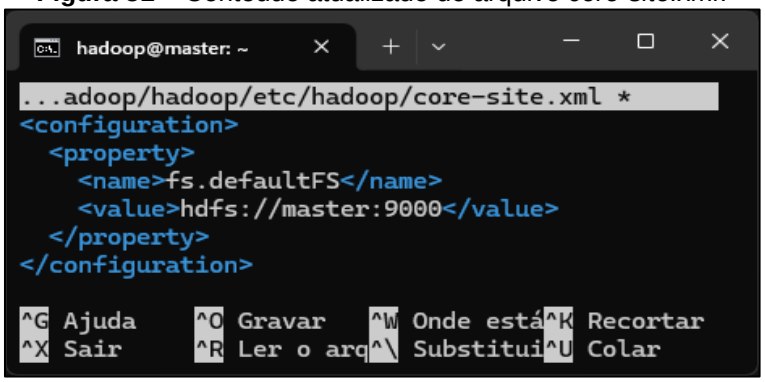


Fonte: Autores, 2025.

Após abrir o editor de texto, digite as informações abaixo, como mostra a Figura 51.

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://master:9000</value>
</property>
```

Figura 51 – Conteúdo atualizado do arquivo *core-site.xml*.



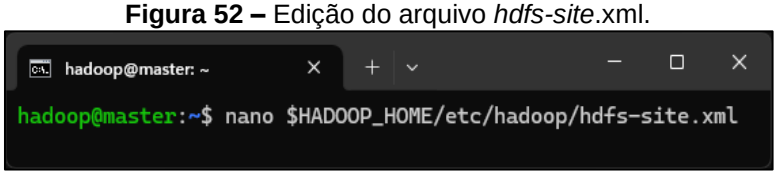
Fonte: Autores, 2025.

Depois de adicionar as informações no arquivo, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

14º Passo: Editar o arquivo *hdfs-site.xml*

Na sequência, execute a seguinte linha de comando no modo terminal da *VM master*: **nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml**, como mostra a Figura 52.

Figura 52 – Edição do arquivo *hdfs-site*.xml.

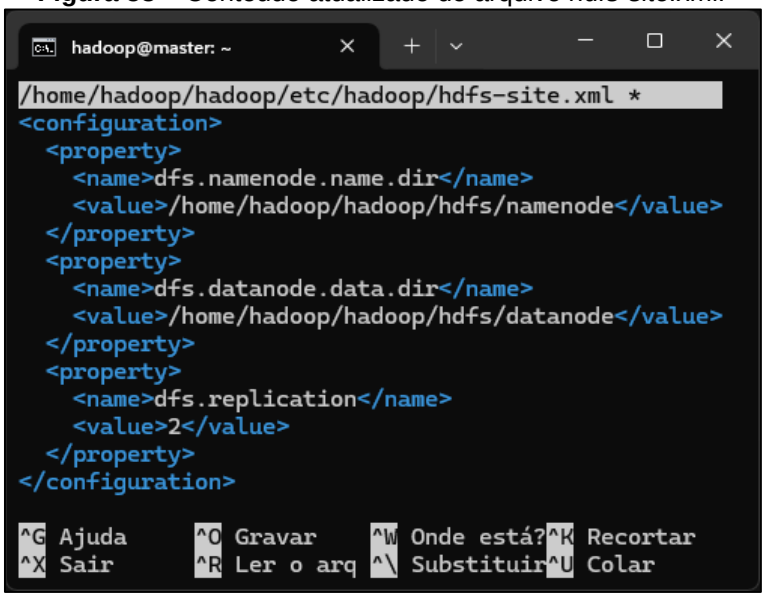


Fonte: Autores, 2025.

Após abrir o editor de texto, digite as informações abaixo, como mostra a Figura 53.

```
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/hadoop/hadoop/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/home/hadoop/hadoop/hdfs/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>2</value>
</property>
```

Figura 53 – Conteúdo atualizado do arquivo *hdfs-site.xml.*



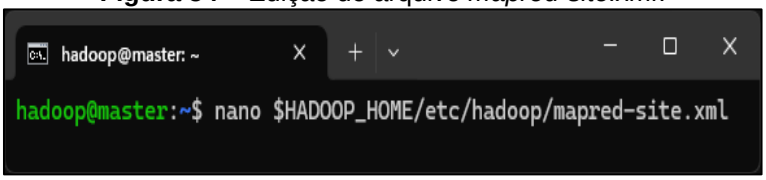
Fonte: Autores, 2025.

Depois de adicionar as informações, pressione **CTRL+S** para salvar o arquivo e **CTRL+X** para sair do editor de texto.

15º Passo: Editar o arquivo *mapred-site.xml*

Para abrir o arquivo ***mapred-site.xml*** e editá-lo, no modo terminal da *VM master*, execute a seguinte linha de comando: **nano $HADOOP_HOME/etc/hadoop/mapred-site.xml**, como mostra a Figura 54.

Figura 54 – Edição do arquivo *mapred-site.xml*.



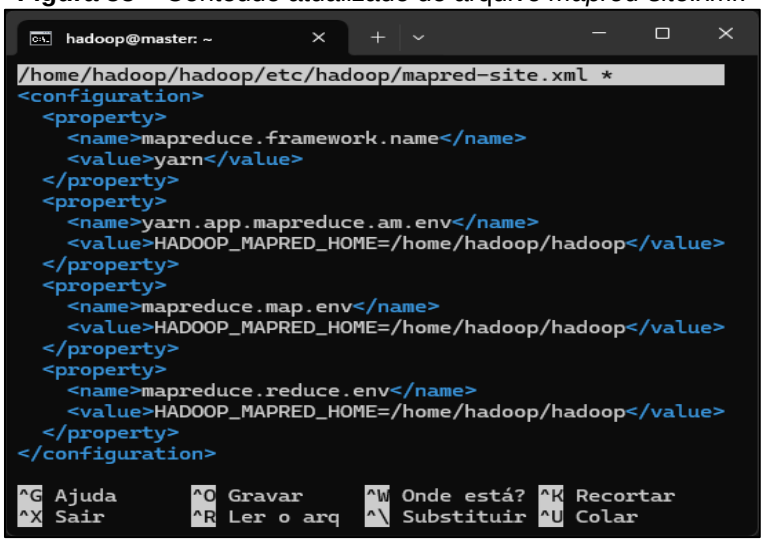
Fonte: Autores, 2025.

Após abrir o editor de texto, digite as informações abaixo, como mostra a Figura 55.

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
<property>
```

```
  <name>yarn.app.mapreduce.am.env</name>
  <value>HADOOP_MAPRED_HOME=/home/hadoop/hadoop</value>
</property>
<property>
  <name>mapreduce.map.env</name>
  <value>HADOOP_MAPRED_HOME=/home/hadoop/hadoop</value>
</property>
<property>
  <name>mapreduce.reduce.env</name>
  <value>HADOOP_MAPRED_HOME=/home/hadoop/hadoop</value>
</property>
```

Figura 55 – Conteúdo atualizado do arquivo *mapred-site.xml*.



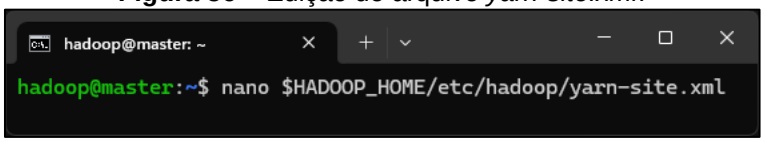
Fonte: Autores, 2025.

Em seguida, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

16º Passo: Editar o arquivo *yarn-site.xml*

Para abrir o arquivo ***yarn-site.xml*** e editá-lo, no modo terminal da *VM master*, execute a seguinte linha de comando: **nano $HADOOP_HOME/etc/hadoop/yarn-site.xml**, como mostra a Figura 56.

Figura 56 – Edição do arquivo *yarn-site.xml*.

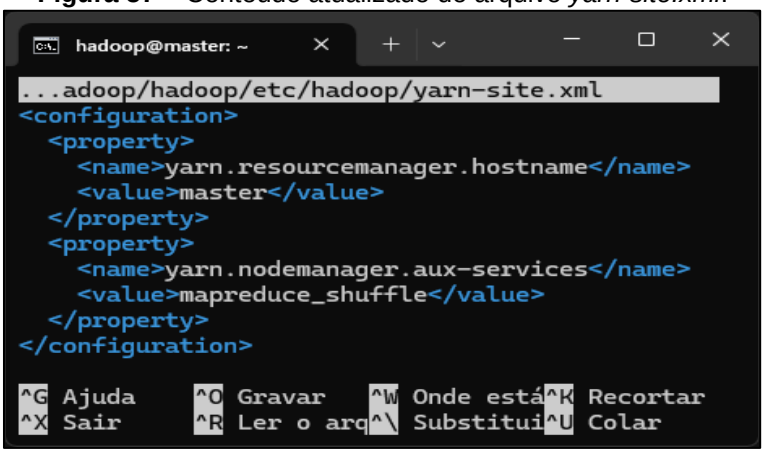


Fonte: Autores, 2025.

Após abrir o editor de texto, digite as informações abaixo, como mostra a Figura 57.

```
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>master</value>
 </property>
 <property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
 </property>
```

Figura 57 – Conteúdo atualizado do arquivo *yarn-site.xml*.



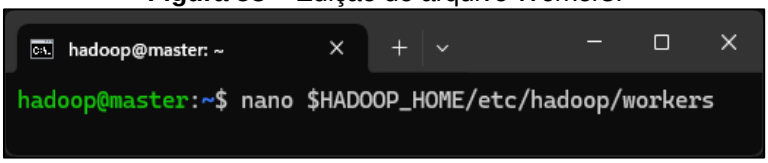
Fonte: Autores, 2025.

Na sequência, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor.

17º Passo: Editar o arquivo *Workers*

Em seguida, para abrir e editar o arquivo ***workers***, entre no modo terminal da *VM master* e, na sequência, execute a seguinte linha de comando: **nano $HADOOP_HOME/etc/hadoop/workers**, como mostra a Figura 58.

Figura 58 – Edição do arquivo *Workers*.

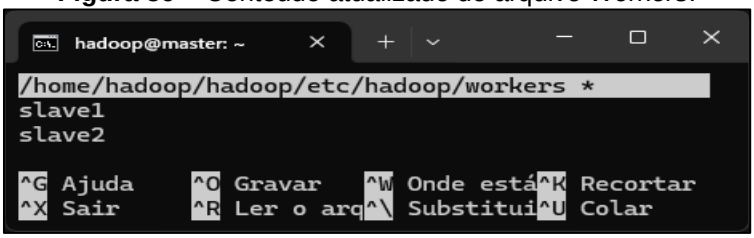


Fonte: Autores, 2025.

Após abrir o editor de texto, apague o nome **localhost** e digite as informações abaixo, como mostra a Figura 59.

```
slave1
slave2
```

Figura 59 – Conteúdo atualizado do arquivo *Workers*.



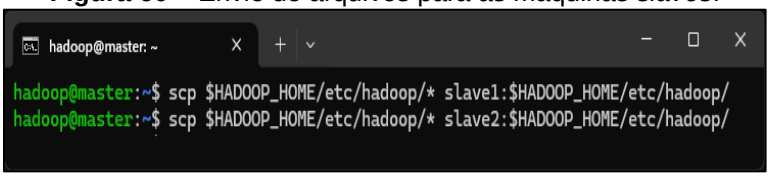
Fonte: Autores, 2025.

Em seguida, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

18º Passo: Enviar os arquivos editados para os *slaves*

Para enviar os arquivos editados para as *VMs slaves*, no modo terminal da *VM master*, execute as seguintes linhas de comandos, respectivamente: **scp $HADOOP_HOME/etc/hadoop/* slave1:$HADOOP_HOME/etc/hadoop/** e **scp $HADOOP_HOME/etc/hadoop/* slave2:$HADOOP_HOME/etc/hadoop/**, como mostra a Figura 60.

Figura 60 – Envio de arquivos para as máquinas *slaves*.



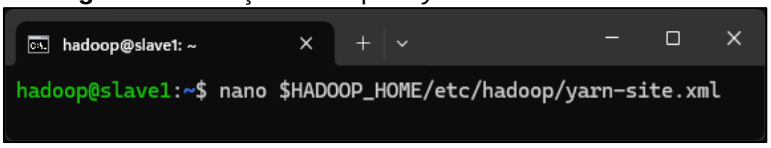
Fonte: Autores, 2025.

19º Passo: Adicionar mais informações no arquivo *yarn-site.xml* nos *slaves*

Para adicionar mais informações no arquivo ***yarn-site.xml***, nas *VMs*, entre no modo terminal de cada uma delas e, em seguida, execute a seguinte linha de

comando: **nano $HADOOP_HOME/etc/hadoop/yarn-site.xml**, como mostra a Figura 61.

Figura 61 – Edição do arquivo *yarn-site.xml* nos *slaves*.

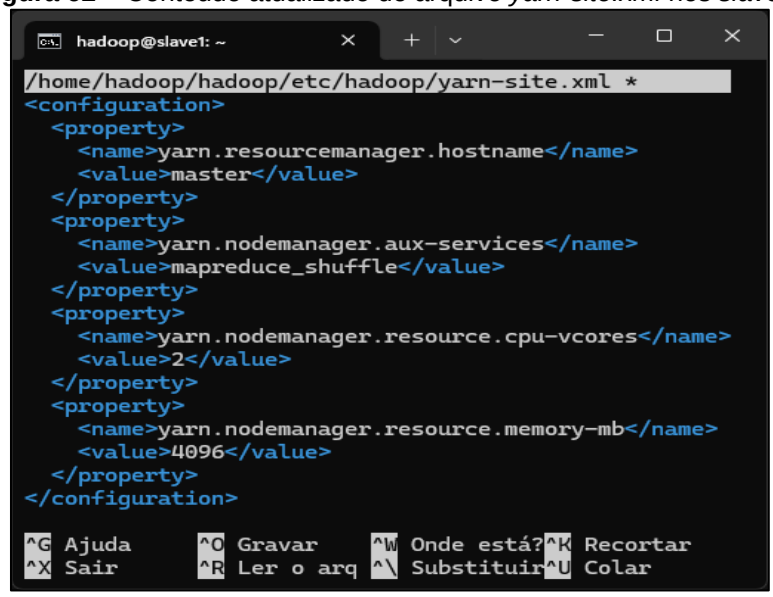


Fonte Autores, 2025.

Na sequência, complete o conteúdo do arquivo digitando as informações abaixo, como mostra a Figura 62.

```
<property>
  <name>yarn.nodemanager.resource.cpu-vcores</name>
  <value>2</value>
</property>
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>4096</value>
</property>
```

Figura 62 – Conteúdo atualizado do arquivo *yarn-site.xml* nos *slaves*.



Fonte: Autores, 2025.

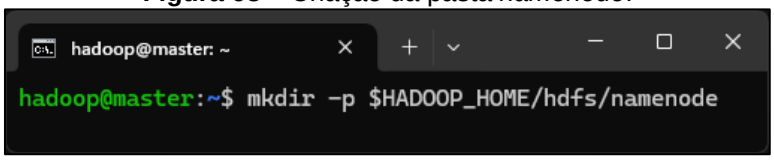
Após adicionar as informações, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto. Lembrando que essas informações deverão ser adicionadas nas duas *VMs slaves*. A conexão remota para as *VMs slaves* deverá ser realizada via

SSH. Para isso, no modo terminal da *VM master*, execute a seguinte linha de comando, para se conectar a cada uma das *VMs*: **ssh slave1** ou **ssh slave2**.

20º Passo: Criar a pasta *namenode* no *master*

Para criar a pasta *namenode* na *VM master*, execute a seguinte linha de comando no modo terminal: **mkdir -p $HADOOP_HOME/hdfs/namenode**, como mostra a Figura 63.

Figura 63 – Criação da pasta *namenode*.

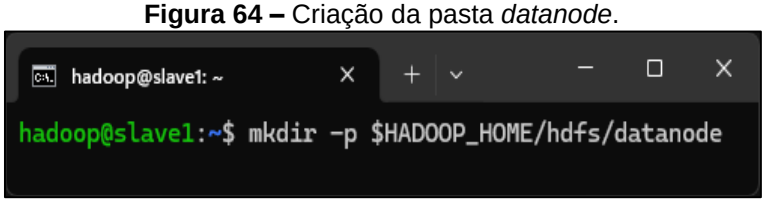


Fonte: Autores, 2025.

21º Passo: Criar a pasta *datanode* nos *slaves*

Para criar a pasta ***datanode*** nas *VMs slaves*, no modo terminal de cada uma delas, execute a seguinte linha de comando: **mkdir -p $HADOOP_HOME/hdfs/datanode**, como mostra a Figura 64.

Figura 64 – Criação da pasta *datanode*.

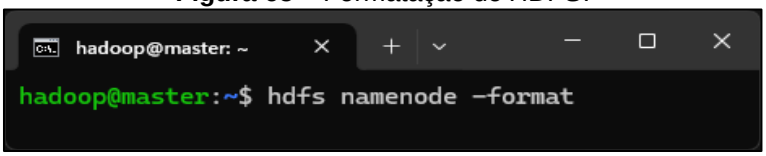


Fonte: Autores, 2025.

22º Passo: Formatar o *HDFS*

Para formatar o *HDFS*, no modo terminal da *VM master*, execute a seguinte linha de comando: **hdfs namenode -format**, como mostra a Figura 65.

Figura 65 – Formatação do *HDFS*.

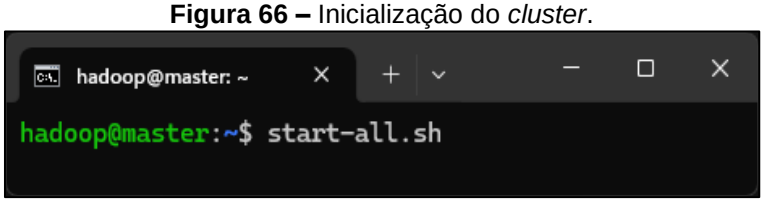


Fonte: Autores, 2025.

23º Passo: Iniciar o *cluster*

Para iniciar o *cluster*, execute o seguinte comando no modo terminal da *VM master*: **start-all.sh**, como mostra a Figura 66.

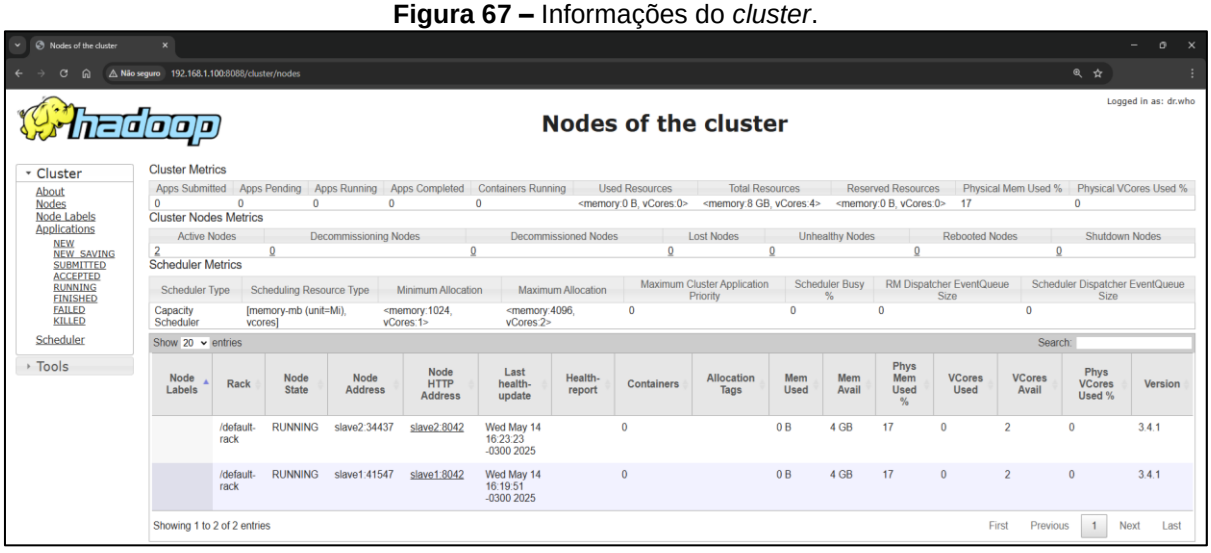
Figura 66 – Inicialização do *cluster*.



Fonte: Autores, 2025.

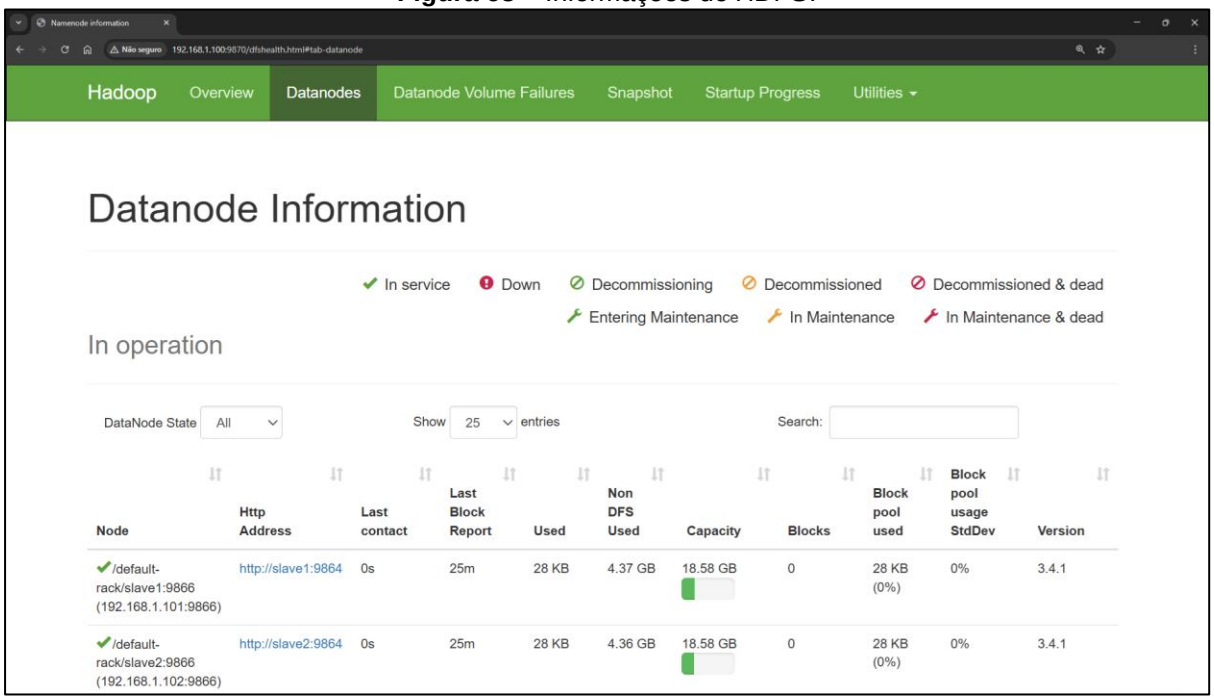
O monitoramento do *cluster* poderá ser realizado através das seguintes *URLs*: **http://192.168.1.100:8088** ou **http://192.168.1.100:9870**, como mostram as Figuras 67 e 68.

Figura 67 – Informações do *cluster*.



Fonte: Autores, 2025.

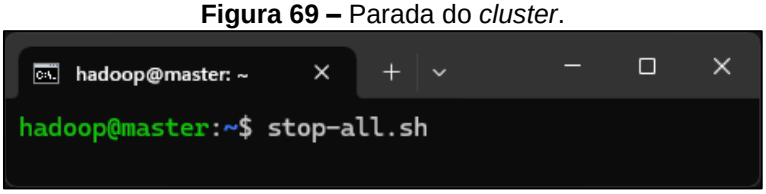
Figura 68 – Informações do *HDFS*.



Fonte: Autores, 2025.

Caso exista a necessidade de “**parar**” o *cluster*, execute o seguinte comando no modo terminal da *VM master*: **stop-all.sh**, como mostra a Figura 69.

Figura 69 – Parada do *cluster*.



Fonte: Autores, 2025.

Capítulo 3

FERRAMENTA DE GERENCIAMENTO DO *CLUSTER*

Neste capítulo, será descrito o processo de instalação e configuração do *Zabbix Server*, uma das ferramentas mais utilizadas no gerenciamento e monitoramento de *clusters*.

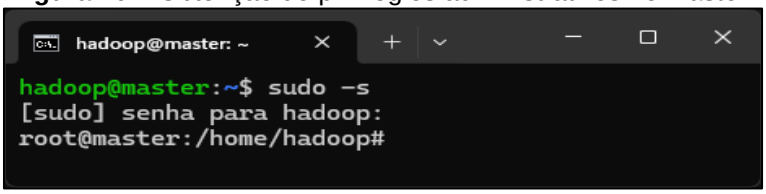
3.1 Instalação do *Zabbix*

Esta seção descreve o processo de instalação do *software Zabbix Server*, uma ferramenta de gerenciamento e monitoramento utilizada para a coleta de dados (métricas) dos dispositivos em uma rede. No caso deste projeto, das *VMs* que fazem parte do *cluster*.

1º Passo: Obter privilégios administrativos na *VM master*

No modo terminal da *VM master*, execute a seguinte linha de comando: **sudo -s** e, em seguida, digite a senha "**1234**", como mostra a Figura 70.

Figura 70 – Obtenção de privilégios administrativos no *master*.



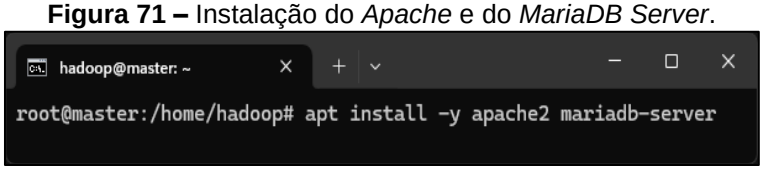
Fonte: Autores, 2025.

Esse processo permitirá que os próximos comandos sejam executados com a permissão de administrador, sem a necessidade de digitar novamente o comando **sudo**, antes de cada linha de comando. Para sair desse modo privilegiado, basta digitar o comando "**exit",** no terminal da *VM master* e, em seguida, pressionar **Enter**.

2º Passo: Instalar o *Apache* e o *MariaDB Server*

Na *VM master*, no modo terminal, com o privilégio de administrador, execute a seguinte linha de comando: **apt install -y apache2 mariadb-server**, para instalar o ***Apache*** e o banco de dados **MariaDB**, como mostra a Figura 71.

Figura 71 – Instalação do *Apache* e do *MariaDB Server*.

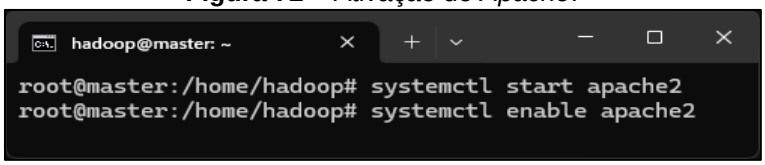


Fonte: Autores, 2025.

3º Passo: Iniciar e ativar o *Apache*

Para iniciar o *Apache*, execute a seguinte linha de comando no modo terminal da *VM master*: **systemctl start apache2**. Em seguida, execute o comando: **systemctl enable apache2**, para ativar o *Apache*, como mostra a Figura 72.

Figura 72 – Ativação do *Apache*.

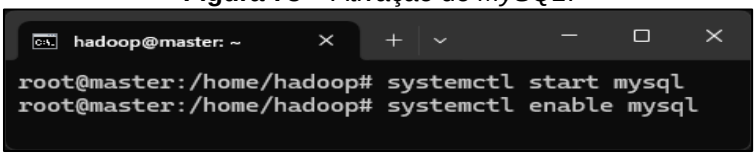


Fonte: Autores, 2025.

4º Passo: Iniciar e ativar o *MySQL*

O próximo passo é iniciar o *MySQL*, para isso, execute o seguinte comando no modo terminal da *MV master*: **systemctl start mysql**. Na sequência, ainda no modo terminal da *VM master*, execute a seguinte linha de comando: **systemctl enable mysql**, para ativar o *MySQL*, como mostra a Figura 73.

Figura 73 – Ativação do *MySQL*.



Fonte: Autores, 2025.

5º Passo: Baixar o pacote do *Zabbix* na *VM master*

Na sequência, no modo terminal da *VM master*, execute a seguinte linha de comando para baixar o pacote do *Zabbix*, diretamente de um repositório: **wget https://repo.zabbix.com/zabbix/6.0/debian/pool/main/z/zabbix-release/zabbix-release_6.0-4+debian11_all.deb**, como mostra a Figura 74.

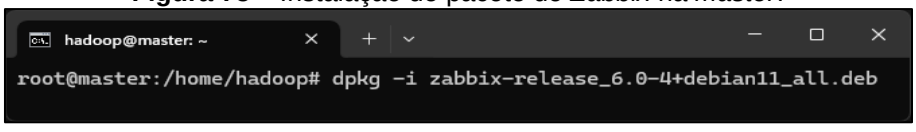
Figura 74 – *Download* do pacote do *Zabbix* na *master*.

Fonte: Autores, 2025.

6º Passo: Instalar o pacote do *Zabbix* na máquina *master*

Para instalar o pacote do *Zabbix*, no modo terminal da *VM master*, execute o seguinte comando: **dpkg -i zabbix-release_6.0-4+debian11_all.deb**, como mostra a Figura 75.

Figura 75 – Instalação do pacote do *Zabbix* na *master*.

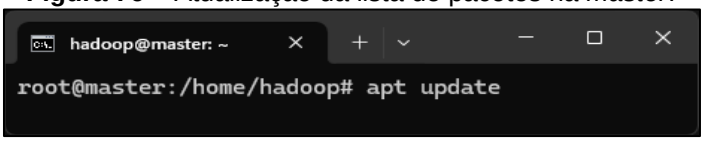


Fonte: Autores, 2025.

7º Passo: Atualizar a lista de pacotes no *master*

Em seguida, no modo terminal da *VM master*, execute a seguinte linha de comando para atualizar a lista de pacotes do *Zabbix*: **apt update**, como mostra a Figura 76.

Figura 76 – Atualização da lista de pacotes na *master*.

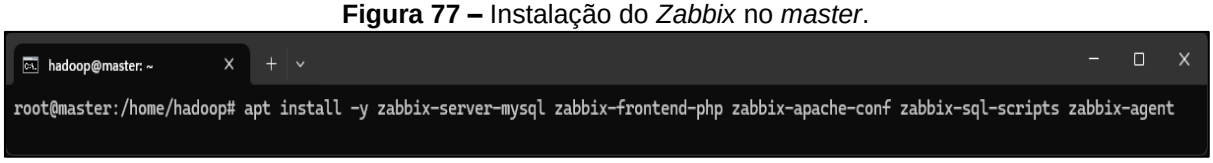


Fonte: Autores, 2025.

8º Passo: Instalar o servidor, agente e a interface *web* do *Zabbix* no *master*

Para instalar o servidor, o agente e a interface *Web* da ferramenta *Zabbix*, no modo terminal da *VM master*, execute a seguinte linha de comando: **apt install -y zabbix-server-mysql zabbix-frontend-php zabbix-apache-conf zabbix-sql-scripts zabbix-agent**, como mostra a Figura 77.

Figura 77 – Instalação do *Zabbix* no *master*.

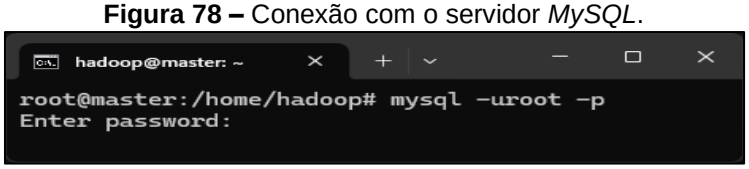


Fonte: Autores, 2025.

9º Passo: Conectar-se ao servidor *MySQL*

Para se conectar ao servidor *MySQL*, no modo terminal da *VM master*, execute a seguinte linha de comando: **mysql -uroot -p**. Quando for solicitada a senha, pressione **Enter** para prosseguir com o processo, como mostra a Figura 78.

Figura 78 – Conexão com o servidor *MySQL*.

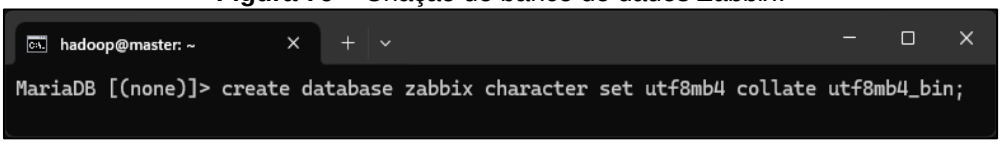


Fonte: Autores, 2025.

10º Passo: Criar o banco de dados *zabbix*

Para criar o banco de dados do *Zabbix*, no modo terminal da *VM master*, execute a seguinte linha de comando: **create database zabbix character set utf8mb4 collate utf8mb4_bin;**, como mostra a Figura 79.

Figura 79 – Criação do banco de dados *Zabbix*.

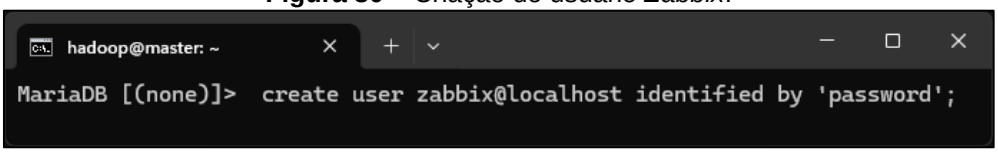


Fonte: Autores, 2025.

11º Passo: Criar o usuário *zabbix*

Em seguida, para criar o usuário "***zabbix***", execute a seguinte linha de comando no modo terminal da *VM master*: **create user zabbix@localhost identified by 'password';**, como mostra a Figura 80.

Figura 80 – Criação do usuário *Zabbix*.

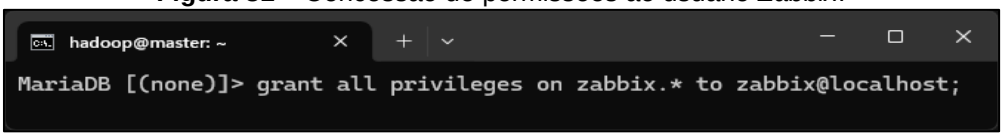


Fonte: Autores, 2025.

12º Passo: Conceder permissões ao usuário *zabbix*

No modo terminal da *VM master*, execute a seguinte linha de comando para conceder permissões ao usuário "***zabbix***": **grant all privileges on zabbix.* to zabbix@localhost;**, como mostra a Figura 81.

Figura 81 – Concessão de permissões ao usuário *Zabbix*.

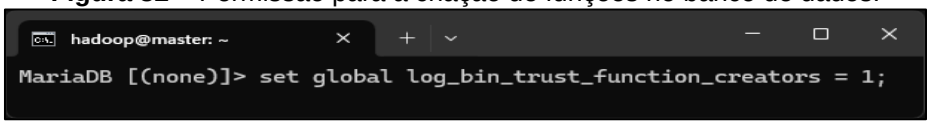


Fonte: Autores, 2025.

13º Passo: Permitir a criação de funções no banco de dados

Para permitir a criação de funções no banco de dados, execute a seguinte linha de comando no terminal da *VM master*: **set global log_bin_trust_function_creators = 1;**, como mostra a Figura 82.

Figura 82 – Permissão para a criação de funções no banco de dados.

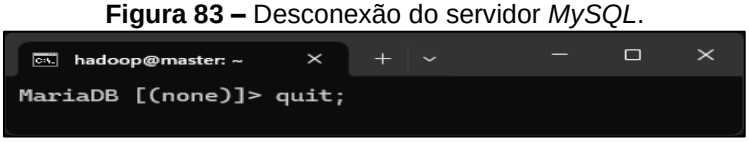


Fonte: Autores, 2025.

14º Passo: Desconectar-se do servidor *MySQL*

Para sair do banco de dados e se desconectar do servidor *MySQL*, no modo terminal da *VM master*, execute a seguinte linha de comando: **quit;**, como mostra a Figura 83.

Figura 83 – Desconexão do servidor *MySQL*.

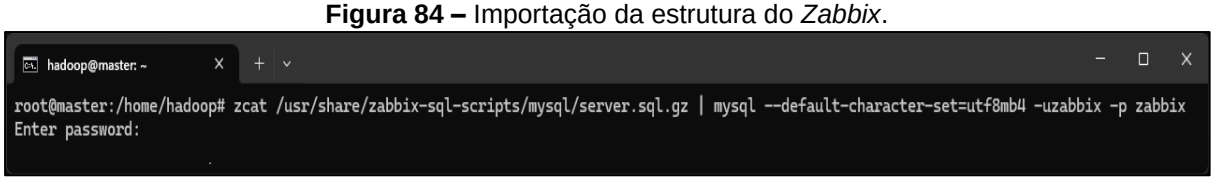


Fonte: Autores, 2025.

15º Passo: Importar a estrutura do *zabbix*

Em seguida, no modo terminal da máquina *master*, execute a seguinte linha de comando para importar a estrutura do *Zabbix*: **zcat /usr/share/zabbix-sql-scripts/mysql/server.sql.gz | mysql --default-character-set=utf8mb4 -uzabbix -p zabbix** e, em seguida, digite a senha “**password***”*, como mostra a Figura 84.

Figura 84 – Importação da estrutura do *Zabbix*.

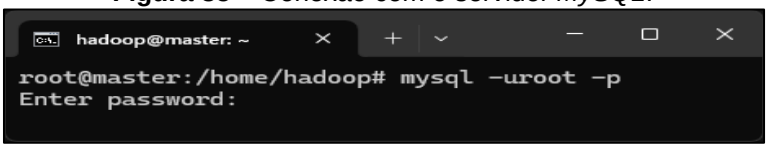


Fonte: Autores, 2025.

16º Passo: Conectar-se ao servidor *MySQL*

Para se conectar ao servidor *MySQL*, execute a seguinte linha de comando no modo terminal da *VM master*: **mysql -uroot -p**. Quando for solicitada a senha, pressione **Enter** para continuar com o processo, como mostra a Figura 85.

Figura 85 – Conexão com o servidor *MySQL*.

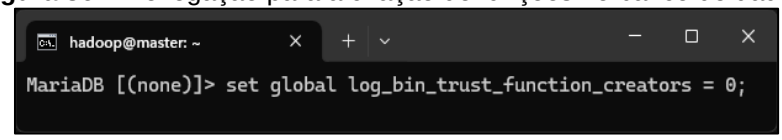


Fonte: Autores, 2025.

17º Passo: Revogar a criação de funções no banco de dados

No modo terminal da *VM master*, execute a seguinte linha de comando: **set global log_bin_trust_function_creators = 0;**, para revogar a criação de funções no banco de dados, como mostra a Figura 86.

Figura 86 – Revogação para a criação de funções no banco de dados.

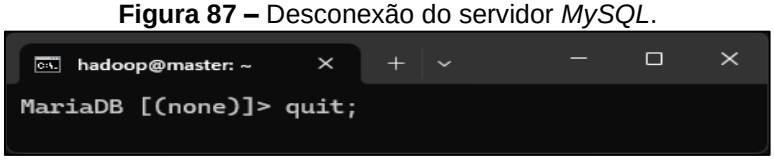


Fonte: Autores, 2025.

18º Passo: Desconectar-se do servidor *MySQL*

Para se desconectar do servidor *MySQL*, execute a seguinte linha de comando no modo terminal da *VM master*: **quit;**, como mostra a Figura 87.

Figura 87 – Desconexão do servidor *MySQL*.

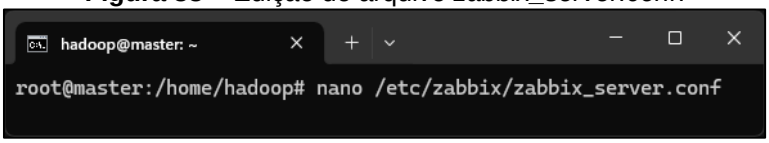


Fonte: Autores, 2025.

19º Passo: Editar o arquivo *zabbix_server.conf*

No modo terminal da *VM master*, execute o seguinte comando para editar o arquivo ***zabbix_server.conf***: **nano /etc/zabbix/zabbix_server.conf**, como mostra a Figura 88.

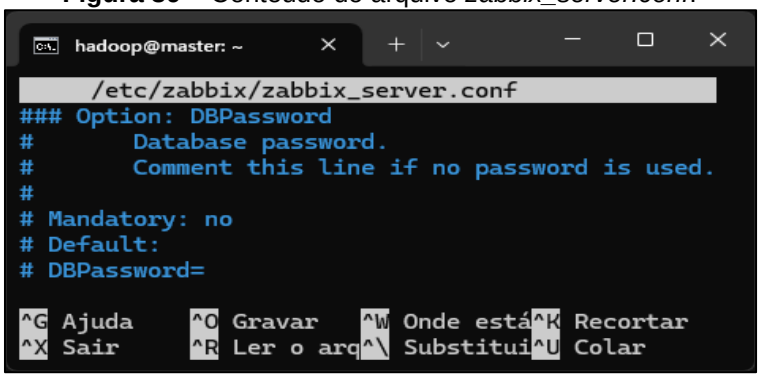
Figura 88 – Edição do arquivo *zabbix_server.conf*.



Fonte: Autores, 2025.

Após abrir o editor de texto, pressione **CTRL+W**, digite a linha **DBPassword=** e, em seguida, pressione **Enter** para localizar a linha que deverá ser editada, como mostra a Figura 89.

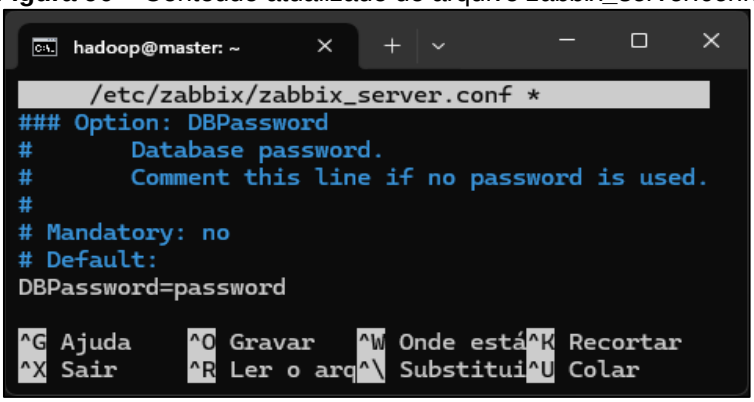
Figura 89 – Conteúdo do arquivo *zabbix_server.conf*.



Fonte: Autores, 2025.

Na sequência, remova o “**#**” do início da linha e, após o sinal de “**=**”, digite a seguinte informação: **password**, como mostra a Figura 90.

Figura 90 – Conteúdo atualizado do arquivo *zabbix_server.conf*.



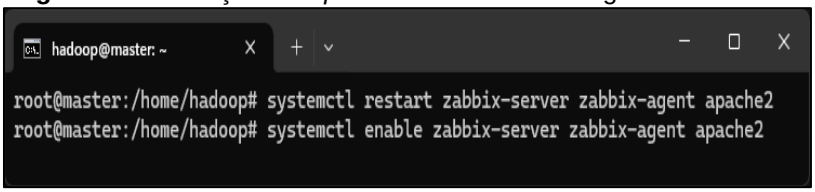
Fonte: Autores, 2025.

Após a adição da informação no arquivo, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

20º Passo: Reiniciar e ativar os serviços do *Zabbix* e do *Apache*

Para reiniciar os serviços do *Zabbix* e do *Apache*, execute a seguinte linha de comando no modo terminal da máquina *master*: **systemctl restart zabbix-server zabbix-agent apache2**. Em seguida, execute o comando: **systemctl enable zabbix-server zabbix-agent apache2**, para ativar os serviços novamente, como mostra a Figura 91.

Figura 91 – Ativação do *Apache* e do servidor e agente do *Zabbix*.

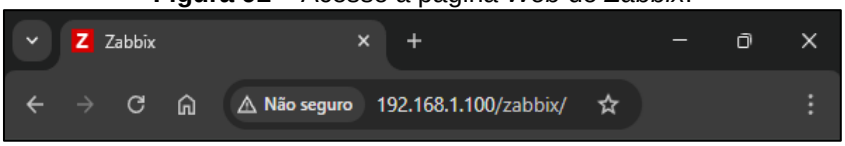


Fonte: Autores, 2025.

21º Passo: Acessar a página *Web* do *Zabbix*

Para acessar a página *Web* do *Zabbix*, utilize a seguinte *URL*: **http://192.168.1.100/zabbix**, como ilustra a Figura 92.

Figura 92 – Acesso à página *Web* do *Zabbix*.

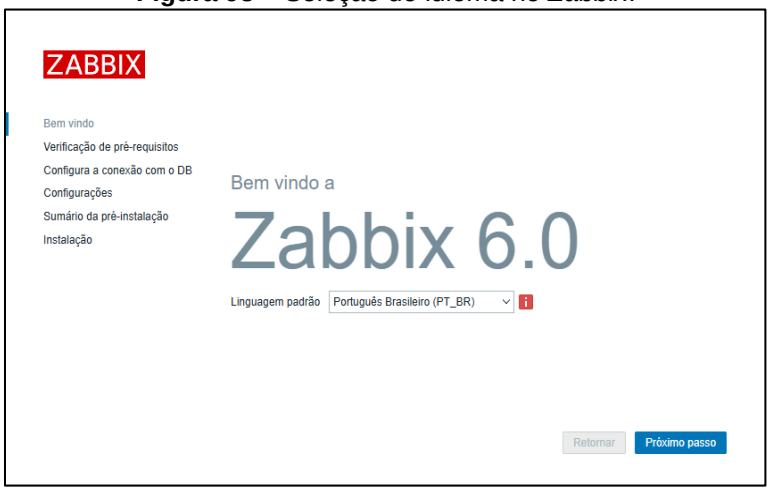


Fonte: Autores, 2025.

22º Passo: Escolher o idioma do *Zabbix*

Na sequência, é preciso escolher o idioma que o *Zabbix* irá trabalhar, para isso, selecione a opção **Português Brasileiro** e, em seguida, clique em **Próximo passo**, para continuar, como mostra a Figura 93.

Figura 93 – Seleção do idioma no *Zabbix*.

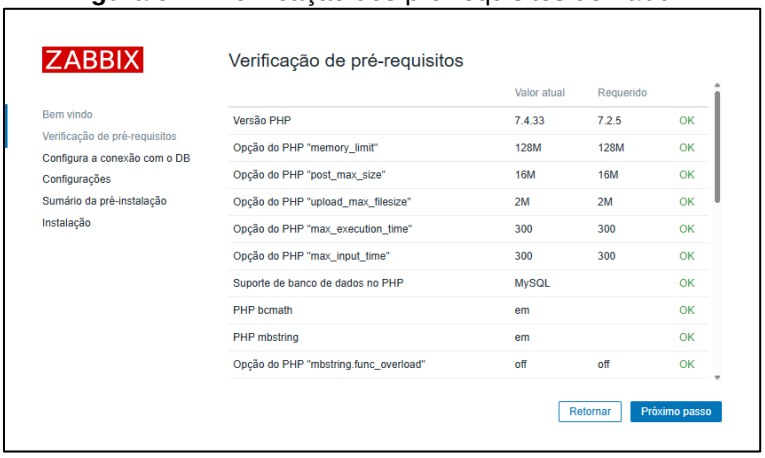


Fonte: Autores, 2025.

23º Passo: Verificar os pré-requisitos do *Zabbix*

O próximo passo é efetuar uma verificação dos pré-requisitos necessários para o funcionamento do *Zabbix*. Caso todos os pré-requisitos estejam "**OK**", clique em **Próximo passo**, para continuar com o processo, como mostra a Figura 94.

Figura 94 – Verificação dos pré-requisitos do *Zabbix*.

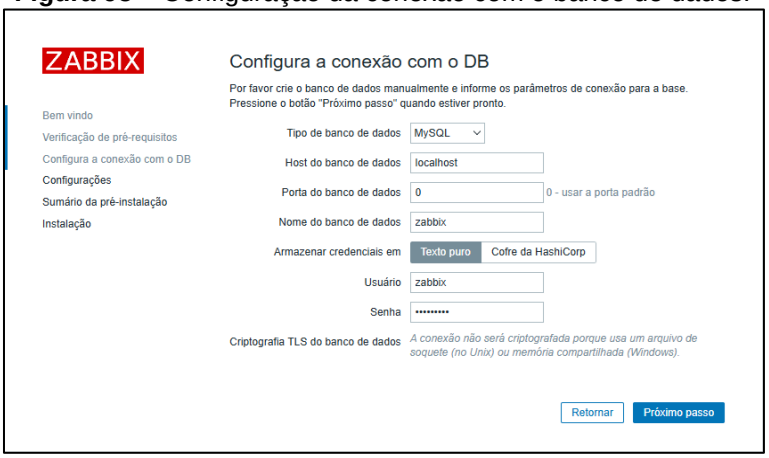


Fonte: Autores, 2025.

24º Passo: Configurar a conexão com o banco de dados

Em seguida, para estabelecer a conexão com o banco de dados do *Zabbix*, digite **password** no campo **Senha** e, em seguida, clique em **Próximo passo**, para continuar, como mostra a Figura 95.

Figura 95 – Configuração da conexão com o banco de dados.

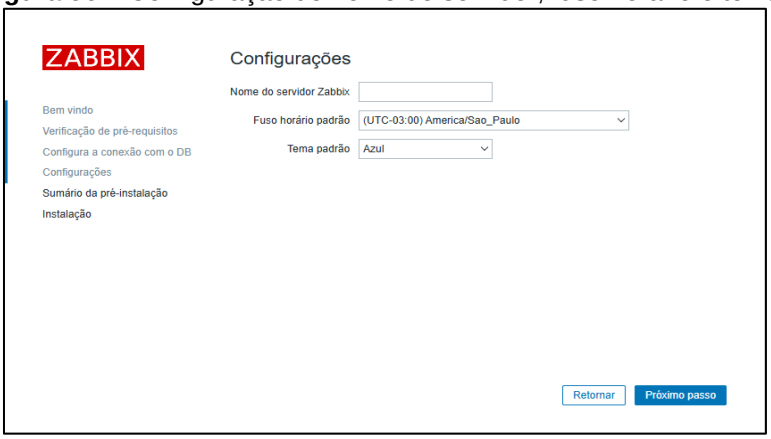


Fonte: Autores, 2025.

25º Passo: Configurar nome do servidor, fuso horário e tema

Se desejar, defina um nome para o servidor *Zabbix*, depois, altere o fuso horário ou o tema da interface. Em seguida, clique em **Próximo passo**, para continuar, como mostra a Figura 96.

Figura 96 – Configuração do nome do servidor, fuso horário e tema.

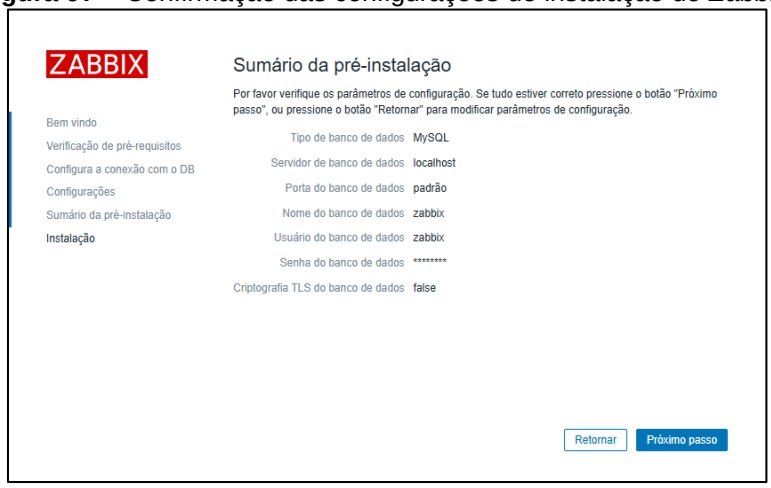


Fonte: Autores, 2025.

26º Passo: Confirmar as configurações de instalação do *Zabbix*

Na tela seguinte, confirme as configurações de instalação do *Zabbix*, clicando em **Próximo passo**, como mostra a Figura 97.

Figura 97 – Confirmação das configurações de instalação do *Zabbix*.

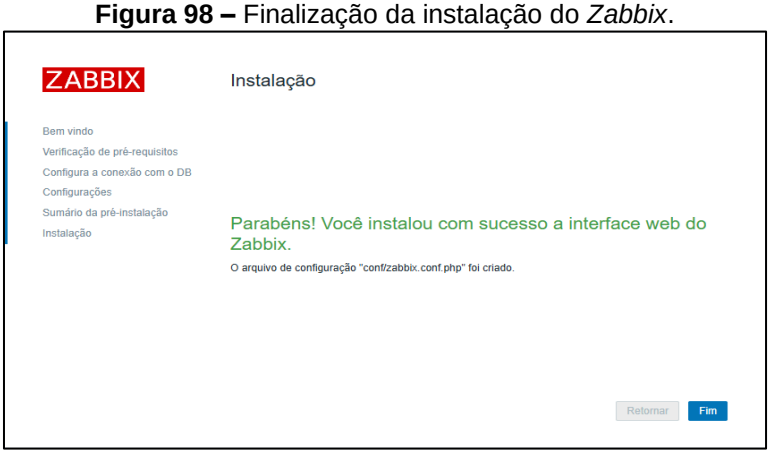


Fonte: Autores, 2025.

27º Passo: Finalizar a instalação do *Zabbix*

Para finalizar o processo de instalação do *Zabbix*, clique no botão **Fim**, como mostra a Figura 98.

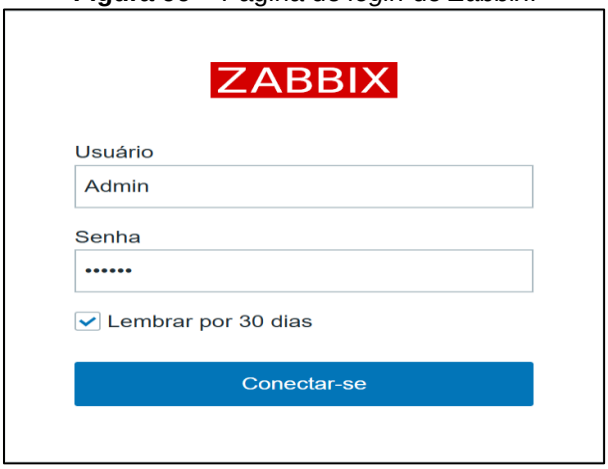
Figura 98 – Finalização da instalação do *Zabbix*.



Fonte: Autores, 2025.

Em seguida, para acessar a página *Web* do *Zabbix*, digite a seguinte *URL*: **http://192.168.1.100/zabbix**. Ao acessar a página, digite **Admin**, no campo do **Usuário** e **zabbix**, no campo **Senha** e, em seguida, clique em **Conectar-se**, como mostra a Figura 99.

Figura 99 – Página de *login* do *Zabbix*.

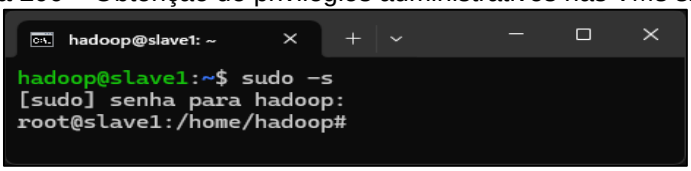


Fonte: Autores, 2025.

28º Passo: Obter privilégios administrativos nos *slaves*

Para obter os privilégios administrativos nas *VMs slaves*, execute a seguinte linha de comando no terminal de cada uma delas: **sudo -s** e, em seguida, digite a senha **1234**, como mostra a Figura 100.

Figura 100 – Obtenção de privilégios administrativos nas *VMs slaves*.



Fonte: Autores, 2025.

29º Passo: Baixar o pacote do repositório do *Zabbix* nas máquinas *slaves*

No modo terminal das *VMs slaves*, execute a seguinte linha de comando: **wget https://repo.zabbix.com/zabbix/6.0/debian/pool/main/z/zabbix-release/zabbix-release_6.0-4+debian11_all.deb**, para baixar o pacote do repositório do *Zabbix*, como mostra a Figura 101.

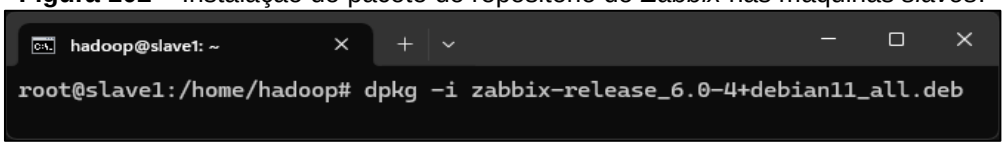
Figura 101 – *Download* do pacote do repositório do *Zabbix* nas máquinas *slaves*.

Fonte: Autores, 2025.

30º Passo: Instalar o pacote do repositório do *Zabbix* nas máquinas *slaves*

Para instalar o pacote do repositório do *Zabbix* nas máquinas *slaves*, execute a seguinte linha de comando no modo terminal de cada uma delas: **dpkg -i zabbix-release_6.0-4+debian11_all.deb**, como mostra a Figura 102.

Figura 102 – Instalação do pacote do repositório do *Zabbix* nas máquinas *slaves*.

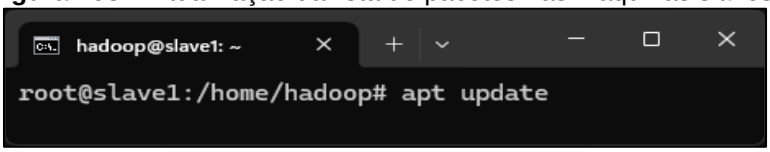


Fonte: Autores, 2025.

31º Passo: Atualizar a lista de pacotes nas máquinas *slaves*

No modo terminal de cada uma das máquinas *slaves*, execute a seguinte linha de comando para atualizar a lista de pacotes: **apt update**, como mostra a Figura 103.

Figura 103 – Atualização da lista de pacotes nas máquinas *slaves*.



Fonte: Autores, 2025.

32º Passo: Instalar o agente do *Zabbix* nas máquinas *slaves*

Para instalar o agente do *Zabbix* nas *VMs slaves*, no modo terminal de cada máquina *slave*, execute a seguinte linha de comando: **apt install -y zabbix-agent**, como mostra a Figura 104.

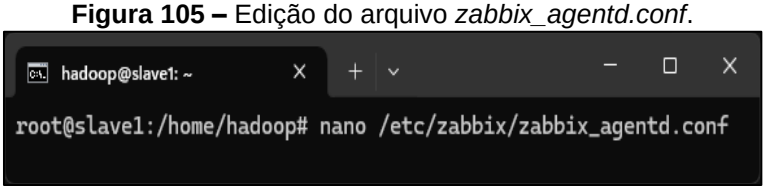
Figura 104 – Instalação do agente *Zabbix* nas máquinas *slaves*.

Fonte: Autores, 2025.

33º Passo: Editar o arquivo *zabbix_agentd.conf*

No modo terminal das máquinas *slaves*, execute a seguinte linha de comando para editar o arquivo ***zabbix_agentd.conf***: **nano /etc/zabbix/zabbix_agentd.conf**, como mostra a Figura 105.

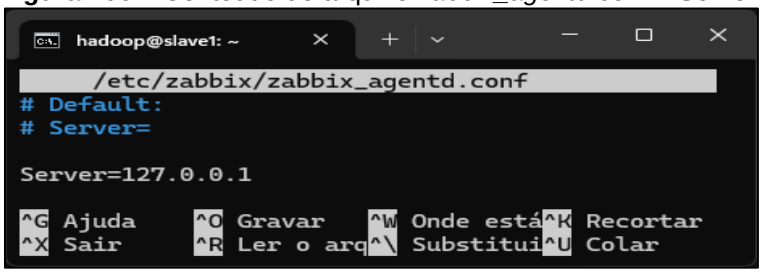
Figura 105 – Edição do arquivo *zabbix_agentd.conf*.



Fonte: Autores, 2025.

Após abrir o editor de texto, pressione **CTRL+W**, digite **Server=127.0.0.1** e, em seguida, pressione **Enter** para localizar a linha que deverá ser editada, como mostra a Figura 106.

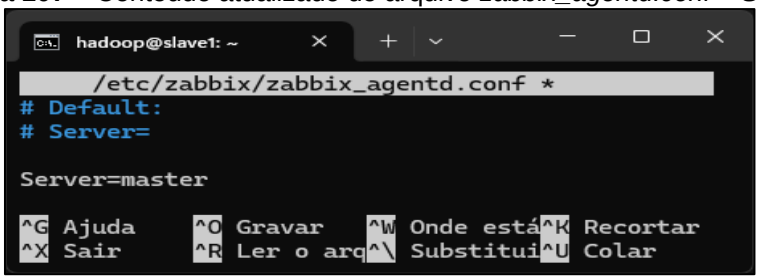
Figura 106 – Conteúdo do arquivo *zabbix_agentd.conf – Server*.



Fonte: Autores, 2025.

Na sequência, substitua o endereço *IP* **127.0.0.1** pela palavra ***master***, como mostra a Figura 107.

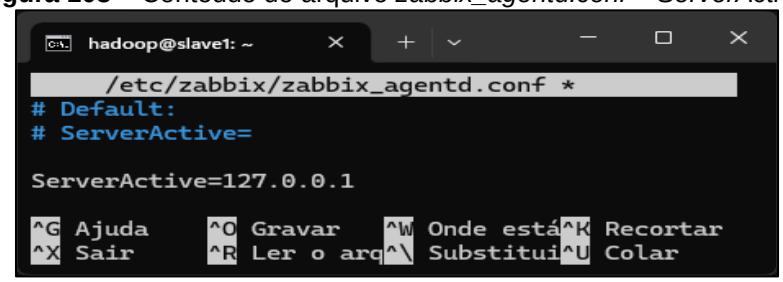
Figura 107 – Conteúdo atualizado do arquivo *zabbix_agentd.conf – Server*.



Fonte: Autores, 2025.

Em seguida, pressione **CTRL+W** novamente, digite **ServerActive=127.0.0.1** e, em seguida, pressione **Enter** para localizar a linha que deverá ser editada, como mostra a Figura 108.

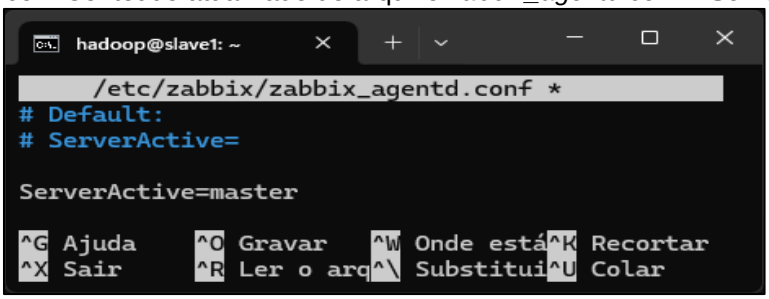
Figura 108 – Conteúdo do arquivo *zabbix_agentd.conf – ServerActive*.



Fonte: Autores, 2025.

Após localizar a linha, substitua o endereço *IP* **127.0.0.1** pela palavra ***master***, como mostra a Figura 109.

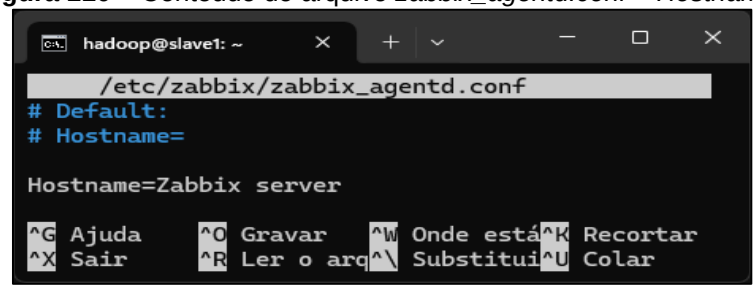
Figura 109 – Conteúdo atualizado do arquivo *zabbix_agentd.conf – ServerActive*.



Fonte: Autores, 2025.

Na sequência, pressione **CTRL+W**, digite **Hostname=Zabbix server** e, em seguida, pressione **Enter** para localizar a linha que deverá ser editada, como mostra a Figura 110.

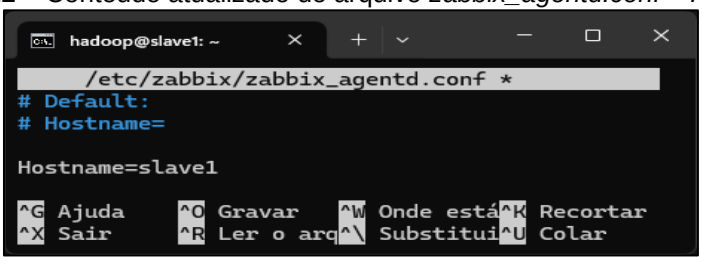
Figura 110 – Conteúdo do arquivo *zabbix_agentd.conf – Hostname*.



Fonte: Autores, 2025.

Continuando o processo, substitua o nome **Zabbix server** por **slave1** ou **slave2**, conforme a *VM* que estiver sendo configurada no momento, como mostra a Figura 111.

Figura 111 – Conteúdo atualizado do arquivo *zabbix_agentd.conf – Hostname*.



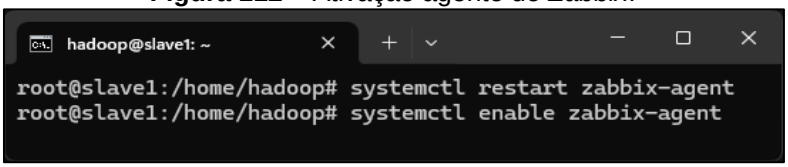
Fonte: Autores, 2025.

Depois de realizar as substituições no arquivo, pressione **CTRL+S** para salvar e **CTRL+X** para sair do editor de texto.

34º Passo: Reiniciar e ativar o agente do *Zabbix*

Para reiniciar o serviço do agente do *Zabbix*, execute a seguinte linha de comando no terminal de cada uma das máquinas *slaves*: **systemctl restart zabbix-agent**. Em seguida, ainda no modo terminal das *slaves*, execute o comando: **systemctl enable zabbix-agent**, para ativar o serviço, como mostra a Figura 112.

Figura 112 – Ativação agente do *Zabbix*.

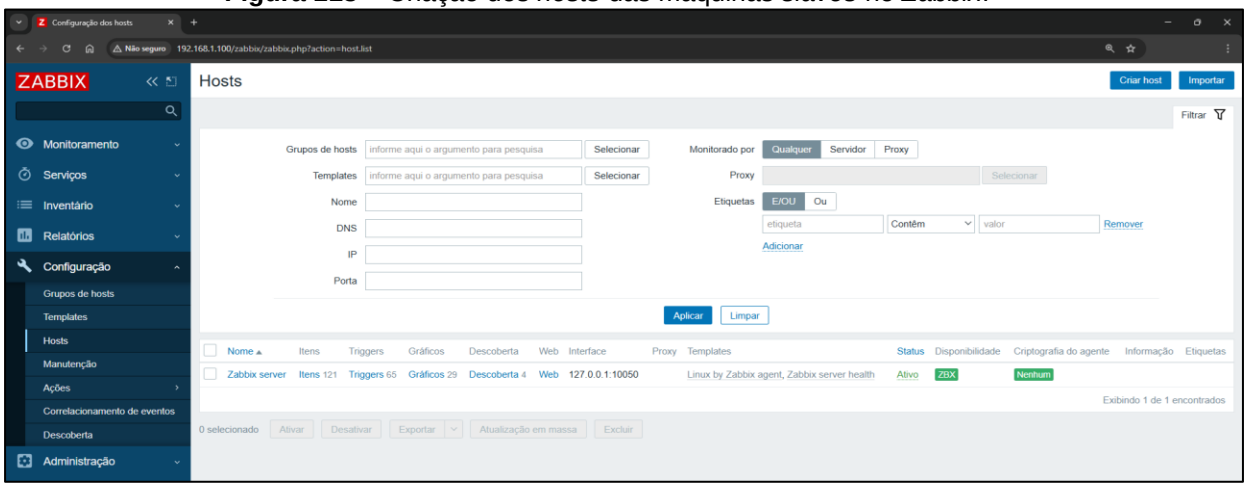


Fonte: Autores, 2025.

35º Passo: Criar os *hosts* das máquinas *slaves* no *Zabbix*

Na sequência, para criar os *hosts* das *VMs slaves*, acesse a página principal do *Zabbix*, através do navegador. No menu lateral esquerdo, clique em **Configuração**, em seguida, clique em **Hosts** e, por fim, clique em **Criar host**, localizado no canto superior direito da página, como mostra a Figura 113.

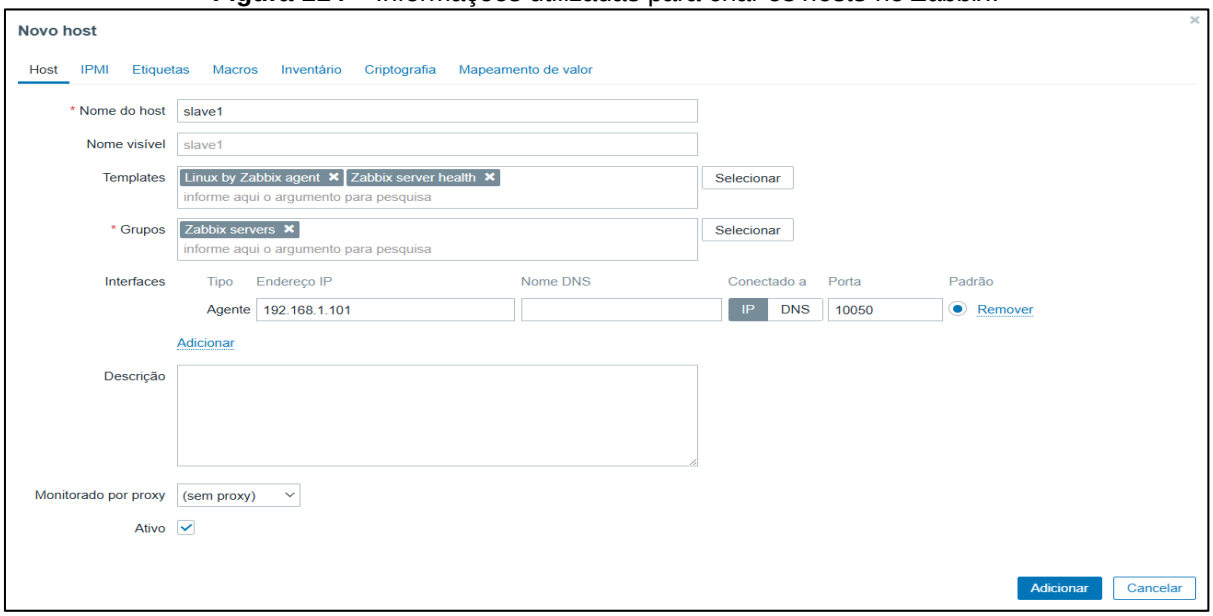
Figura 113 – Criação dos *hosts* das máquinas *slaves* no *Zabbix*.



Fonte: Autores, 2025.

Ao abrir a janela de criação dos *hosts*, digite **slave1** ou **slave2**, no campo **Nome do host**, de acordo com a máquina que está sendo configurada. Em **Templates**, digite **Linux by Zabbix agent** e **Zabbix server health**, pressionando **Enter** após digitar cada um. No campo **Grupos**, digite **Zabbix servers** e pressione **Enter**. Em seguida, em **Interfaces**, clique em **Adicionar**, selecione a opção **Agente** e substitua o endereço *IP* **127.0.0.1** por **192.168.1.101** ou **192.168.1.102**, de acordo com a máquina *slave* que está sendo configurada no momento. Por fim, clique no botão **Adicionar**, como mostra a Figura 114.

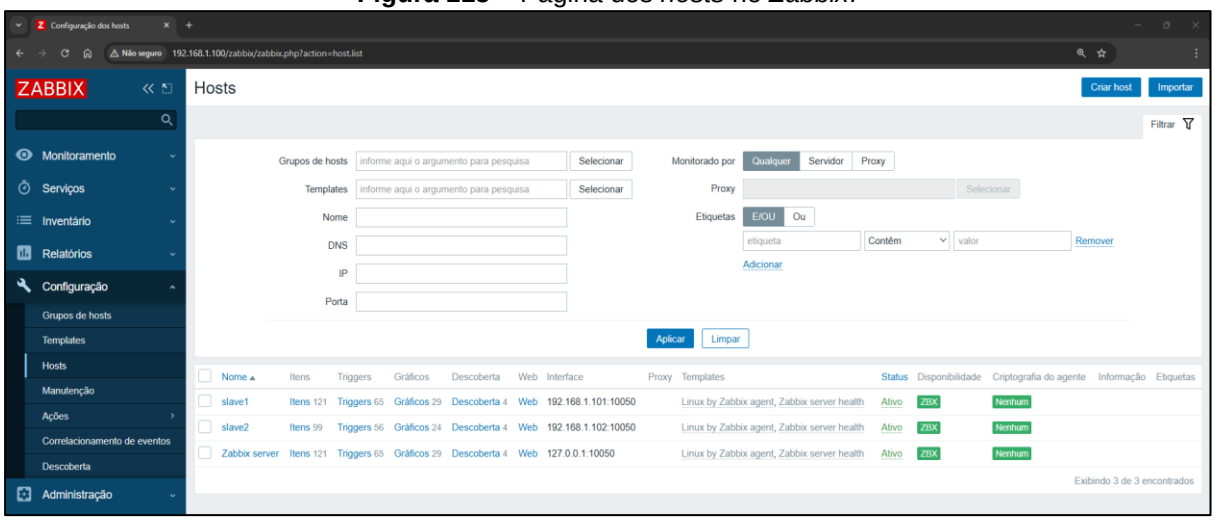
Figura 114 – Informações utilizadas para criar os *hosts* no *Zabbix*.



Fonte: Autores, 2025.

Em seguida, após adicionar os *hosts* das *VMs slaves*, será possível visualizá-los na página dos *hosts* disponíveis no *Zabbix*, como mostra a Figura 115.

Figura 115 – Página dos *hosts* no *Zabbix*.



Fonte: Autores, 2025.

Capítulo 4

VISUALIZAÇÃO GRÁFICA DAS MÉTRICAS DO *CLUSTER*

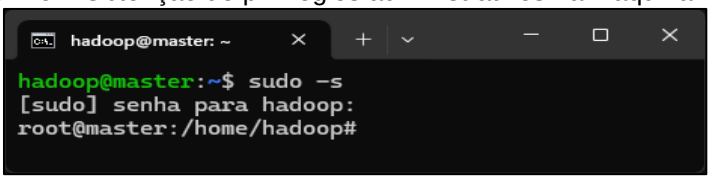
Neste capítulo, será descrito o processo de instalação e configuração do *Grafana*, uma ferramenta visual que permite apresentar informações relevantes das métricas coletadas durante a execução dos testes de *benchmark* do *cluster*, de maneira consolidada e organizada, facilitando a análise e tomada de decisões.

4.1 Instalação do *Grafana*

1º Passo: Obter privilégios administrativos no *master*

No modo terminal da *VM master*, execute a seguinte linha de comando para obter os privilégios de administrador: **sudo -s** e, em seguida, digite a senha **1234**, como mostra a Figura 116.

Figura 116 – Obtenção de privilégios administrativos na máquina *master*.

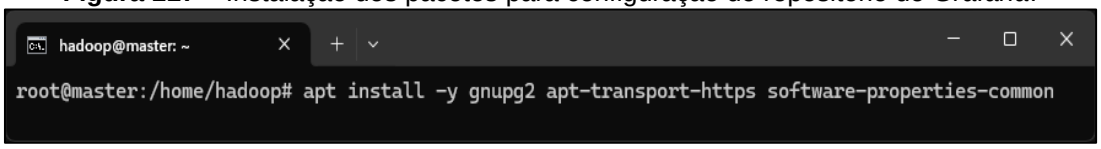


Fonte: Autores, 2025.

2º Passo: Instalar os pacotes de configuração do repositório do *Grafana*

Para fazer a instalação dos pacotes de configuração do *Grafana*, no modo terminal da máquina *master*, execute a seguinte linha de comando: **apt install -y gnupg2 apt-transport-https software-properties-common**, como mostra a Figura 117.

Figura 117 – Instalação dos pacotes para configuração do repositório do *Grafana*.

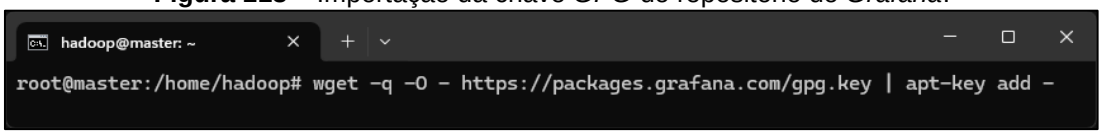


Fonte: Autores, 2025.

3º Passo: Importar a chave *GPG* do repositório do *Grafana*

Em seguida, para importar a chave *GPG* do repositório do *Grafana*, no modo terminal da máquina *master*, execute a seguinte linha de comando: **wget -q -O - https://packages.grafana.com/gpg.key | apt-key add -**, como mostra a Figura 118.

Figura 118 – Importação da chave *GPG* do repositório do *Grafana*.

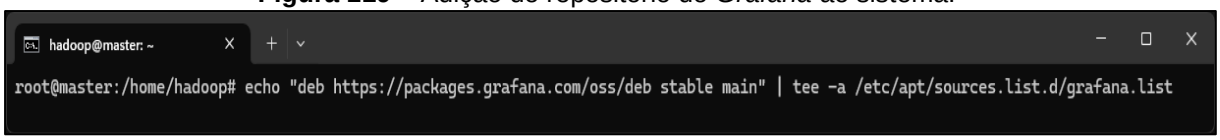


Fonte: Autores, 2025.

4º Passo: Adicionar o repositório do *Grafana* ao sistema

No modo terminal da máquina *master*, execute a seguinte linha de comando para adicionar o repositório do *Grafana* ao sistema: **echo "deb https://packages.grafana.com/oss/deb stable main" | tee -a /etc/apt/sources.list.d/grafana.list**, como mostra a Figura 119.

Figura 119 – Adição do repositório do *Grafana* ao sistema.

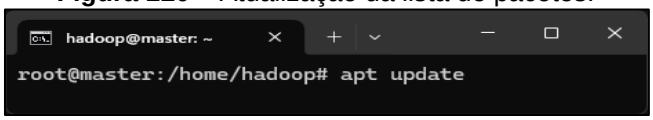


Fonte: Autores, 2025.

5º Passo: Atualizar a lista de pacotes

Na sequência, para atualizar a lista de pacotes, no modo terminal da máquina *master*, execute a seguinte linha de comando: **apt update**, como mostra a Figura 120.

Figura 120 – Atualização da lista de pacotes.

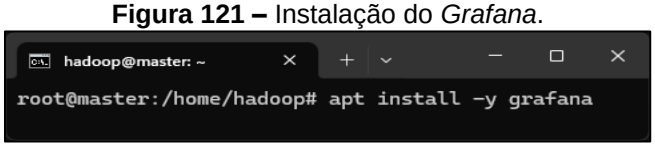


Fonte: Autores, 2025.

6º Passo: Instalar o *Grafana*

Para instalar o *Grafana*, no modo terminal da máquina *master*, execute a seguinte linha de comando: **apt install -y grafana**, como mostra a Figura 121.

Figura 121 – Instalação do *Grafana*.

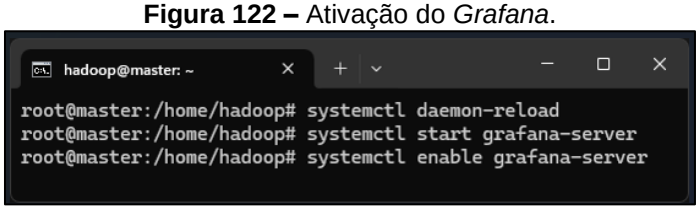


Fonte: Autores, 2025.

7º Passo: Iniciar e ativar o *Grafana*

Continuando com o processo, é necessário recarregar as configurações realizadas, para isso, execute a seguinte linha de comando no terminal da máquina *master*: **systemctl daemon-reload**. Na sequência, ainda no modo terminal da máquina *master*, execute os comandos: **systemctl start grafana-server**, para iniciar o serviço do *Grafana*, e **systemctl enable grafana-server**, para ativar o serviço, como mostra a Figura 122.

Figura 122 – Ativação do *Grafana*.



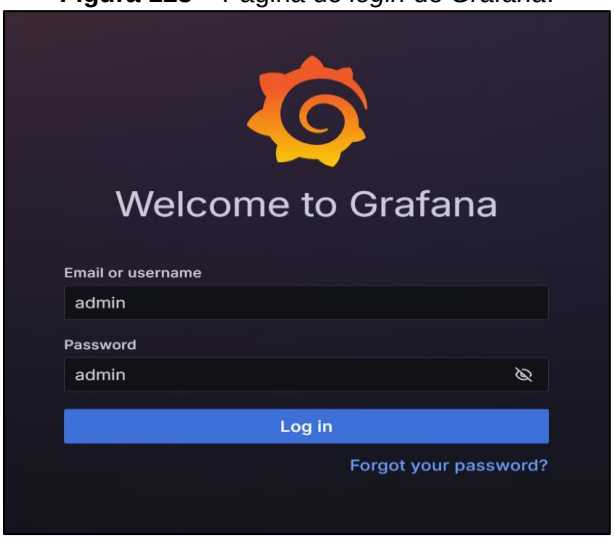
Fonte: Autores, 2025.

8º Passo: Acessar a página *Web* do *Grafana*

Em seguida, para acessar a página *Web* do *Grafana*, digite a seguinte *URL*: **192.168.1.100:3000**.

Na tela de *login*, digite **admin** nos campos **Username** e **Password** e, em seguida, clique em ***Log in***, como mostra a Figura 123.

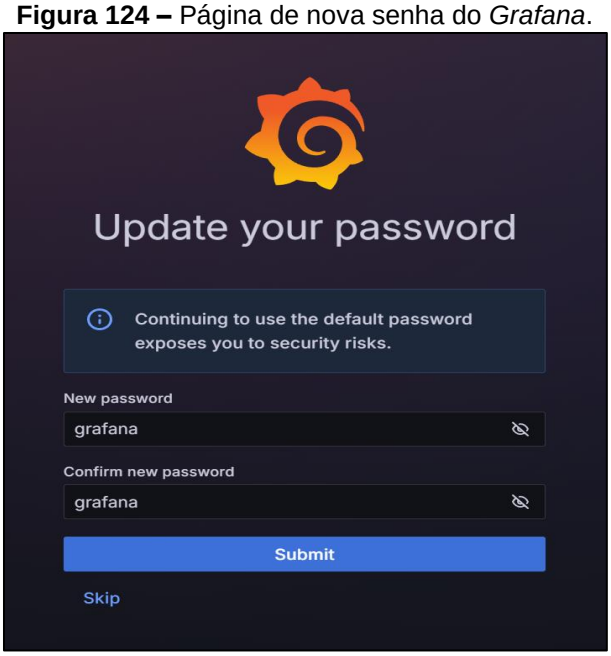
Figura 123 – Página de *login* do *Grafana*.



Fonte: Autores, 2025.

Na sequência, será solicitada a definição de uma nova senha. Digite, por exemplo, a senha **grafana** e repita para confirmar, em seguida, clique em **Submit**, como mostra a Figura 124.

Figura 124 – Página de nova senha do *Grafana*.



Fonte: Autores, 2025.

Capítulo 5

INTEGRAÇÃO ENTRE O *ZABBIX* E O *GRAFANA*

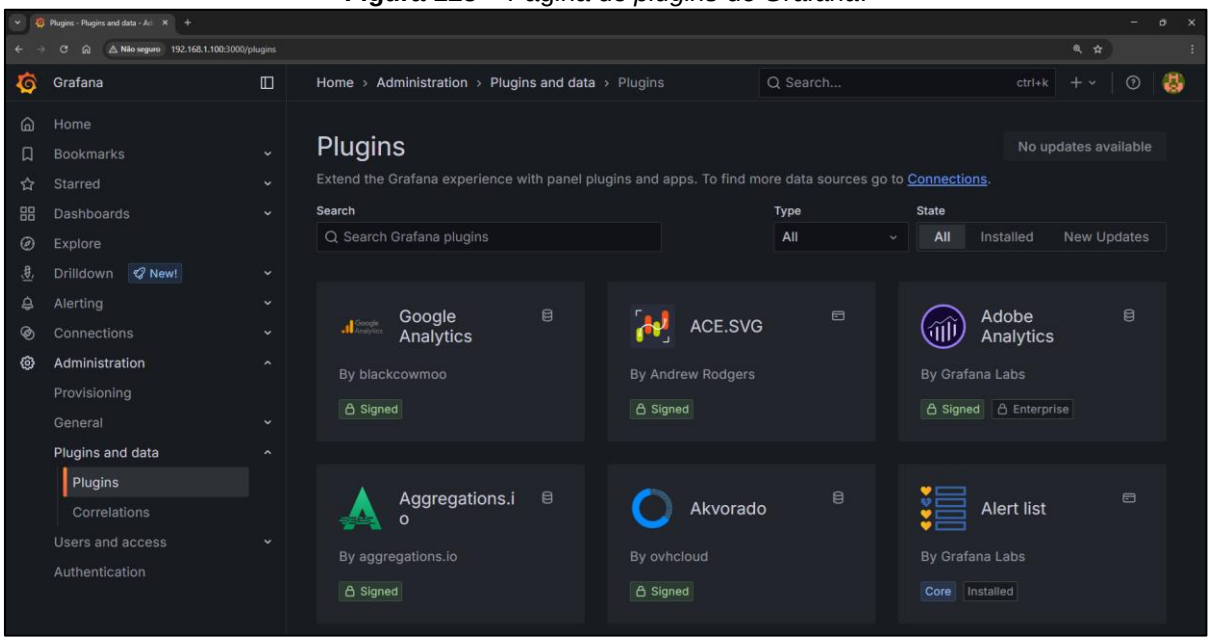
5.1 Integração entre os *softwares*

Esta seção descreve o processo de integração entre a ferramenta de monitoramento e gerenciamento do *cluster*, o *Zabbix* e a ferramenta de visualização gráfica, para confecção de *dashboards*, o *Grafana*.

1º Passo: Acessar a página de *plugins* do *Grafana*

Para acessar a página de *plugins* do *Grafana*, através do navegador, no menu lateral esquerdo, clique em **Administration**, em seguida em **Plugins and data** e, por fim, clique em **Plugins**, como mostra a Figura 125.

Figura 125 – Página de *plugins* do *Grafana*.

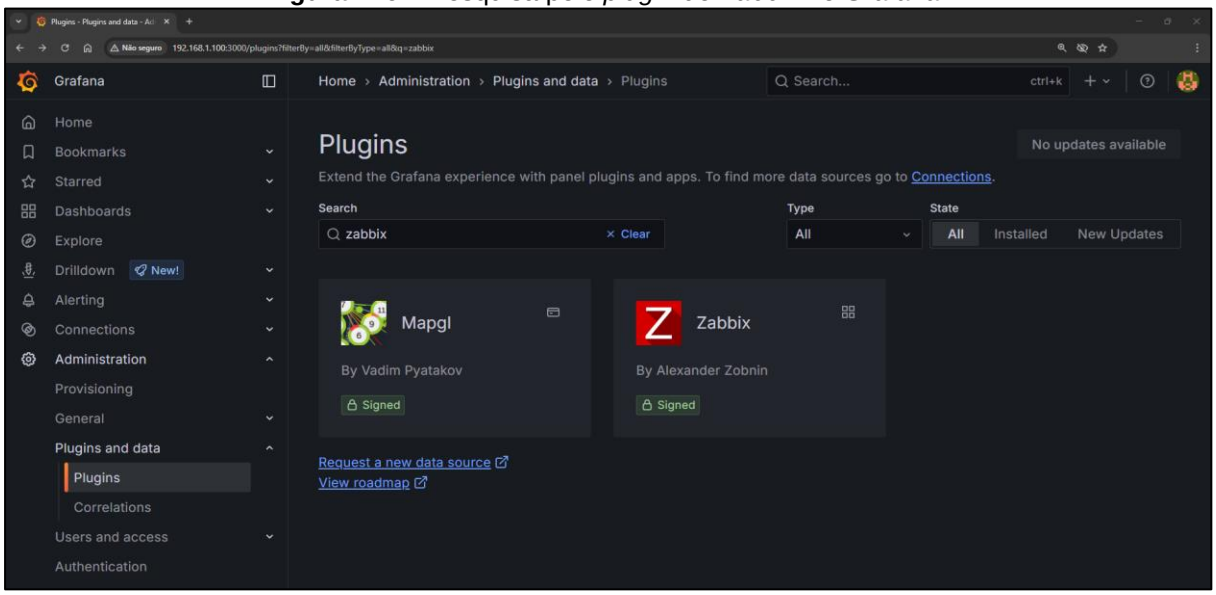


Fonte: Autores, 2025.

2º Passo: Instalar o *plugin* do *Zabbix* no *Grafana*

Na tela seguinte, pesquise por *Zabbix* e, em seguida, clique na opção do *Zabbix* que for exibida, como mostra a Figura 126.

Figura 126 – Pesquisa pelo *plugin* do *Zabbix* no *Grafana*.



Fonte: Autores, 2025.

Após o carregamento da página, clique em **Install** para realizar a instalação. Em seguida, clique em **Enable** para ativar o *plugin*, como mostra a Figura 127.

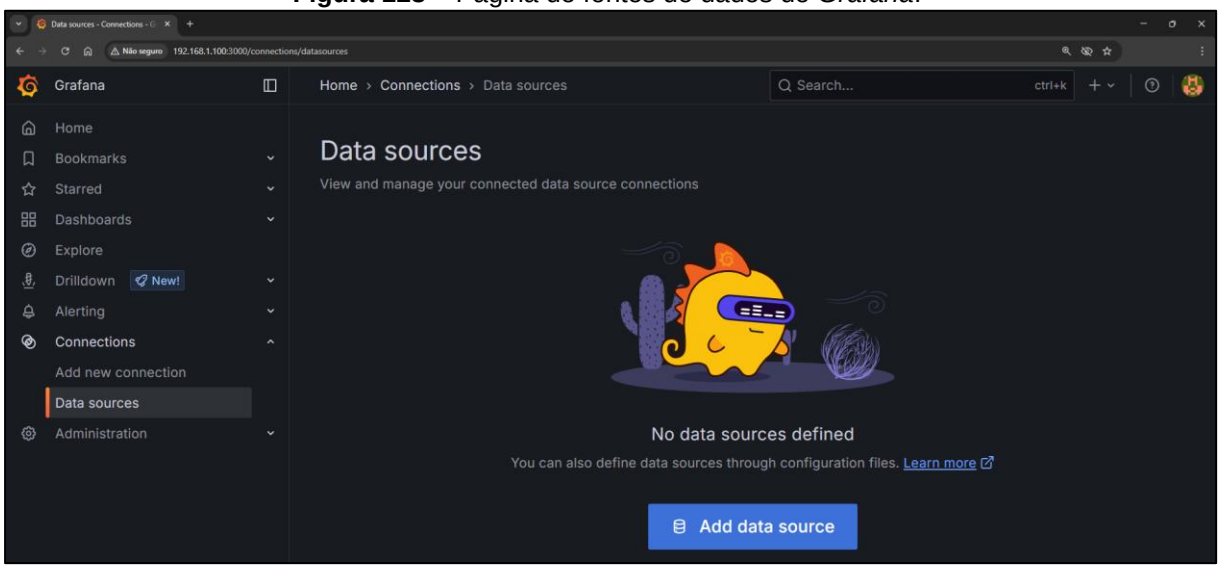
Figura 127 – Instalação e ativação do *plugin* do *Zabbix* no Grafana.

Fonte: Autores, 2025.

3º Passo: Adicionar o *Zabbix* como uma fonte de dados

Continuando o processo, no menu lateral esquerdo, clique em **Connections**, e, em seguida, em **Data sources** e, por fim, clique na opção **Add data source**, como mostra a Figura 128.

Figura 128 – Página de fontes de dados do *Grafana*.



Fonte: Autores, 2025.

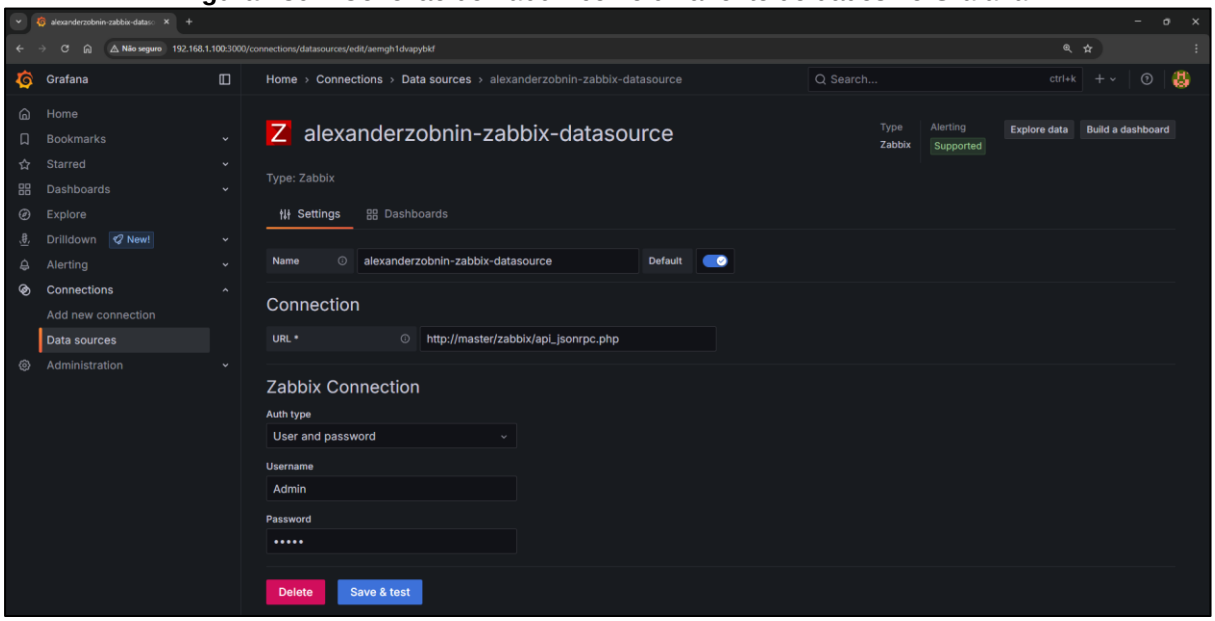
Em seguida, pesquise por *Zabbix* e, na sequência, clique na opção do *Zabbix* que for exibida, como mostra a Figura 129.

Figura 129 – Pesquisa pela fonte de dados *Zabbix* no *Grafana*.

Fonte: Autores, 2025.

Ao carregar a página no navegador, digite a seguinte *URL*: **http://master/zabbix/api_jsonrpc.php**, no campo **URL**. Na sequência, no campo **Username**, digite **Admin**, e em **Password**, digite **zabbix**. Por fim, clique no botão **Save & test**, como mostra a Figura 130.

Figura 130 – Conexão do *Zabbix* como uma fonte de dados no *Grafana*.



Fonte: Autores, 2025.

Capítulo 6

ALGORITMOS DE *BENCHMARK*

Nesta seção serão apresentados os algoritmos de *benchmark* utilizados nos testes de desempenho do *cluster*.

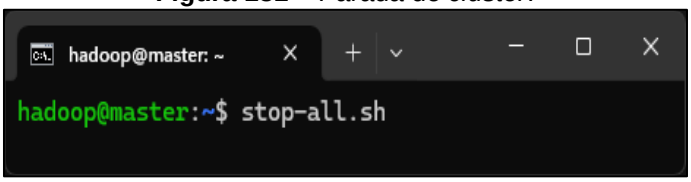
6.1 Algoritmo *Pi* (*quasi-Monte Carlo*)

Para a realização dos testes envolvendo o algoritmo *Pi*, foram utilizados 40 mapas e 10 milhões de amostras. Os comandos descritos nesta seção deverão ser executados no modo terminal da máquina *master*.

1º Passo: Parar o *Cluster*

O processo de execução do algoritmo começa com a “parada” do *cluster*. Esse cenário ocorre sempre que um novo teste tiver que ser realizado no *cluster*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **stop-all.sh**. Além disso, essa ação é necessária para limpar o histórico exibido na interface *web* do *cluster*, disponível através da porta 8088, como mostra a Figura 131.

Figura 131 – Parada do *cluster*.

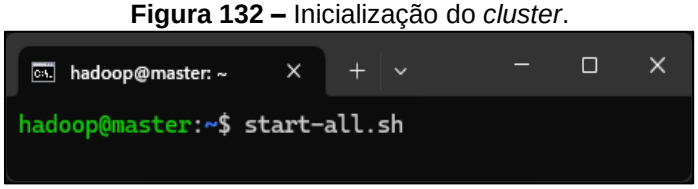


Fonte: Autores, 2025.

2º Passo: Iniciar o *Cluster*

Na sequência, será necessário reiniciar o *cluster*. Para isso, no modo terminal da máquina *master*, execute a seguinte linha de comando: **start-all.sh**, como mostra a Figura 132.

Figura 132 – Inicialização do *cluster*.

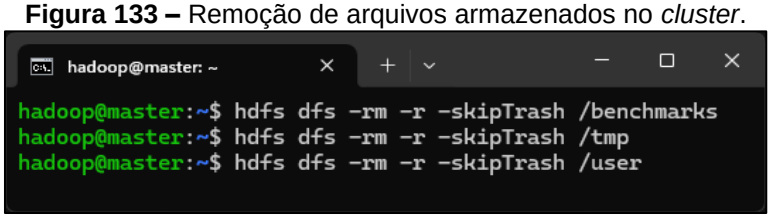


Fonte: Autores, 2025.

3º Passo: Apagar os arquivos armazenados no *Cluster*

Continuando, é preciso apagar os arquivos armazenados no *cluster*, referentes a outros testes, caso tenham sido realizados. Para isso, execute os seguintes comandos no modo terminal da máquina *master*, respectivamente: **hdfs dfs -rm -r -skipTrash /benchmarks**; **hdfs dfs -rm -r -skipTrash /tmp;** e **hdfs dfs -rm -r -skipTrash /user**, como mostra a Figura 133.

Figura 133 – Remoção de arquivos armazenados no *cluster*.

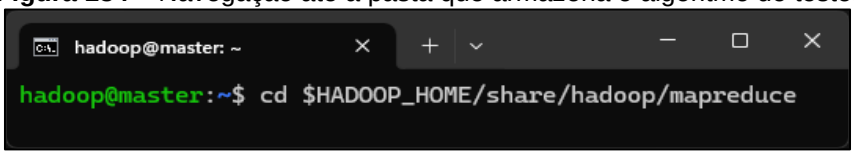


Fonte: Autores, 2025.

4º Passo: Acessar a pasta que armazena os algoritmos de testes do *Hadoop*

Em seguida, é preciso acessar a pasta onde se encontram os algoritmos de teste do *Apache Hadoop*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **cd $HADOOP_HOME/share/hadoop/mapreduce**, como mostra a Figura 134.

Figura 134 – Navegação até a pasta que armazena o algoritmo do teste.

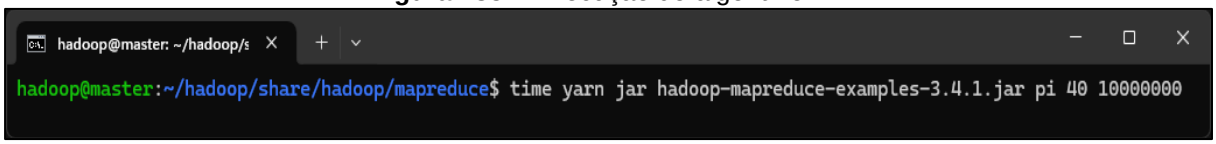


Fonte: Autores, 2025.

5º Passo: Executar o algoritmo *Pi*

Após entrar na pasta de testes do *Hadoop*, execute a seguinte linha de comando no modo terminal da máquina *master*: **time yarn jar hadoop-mapreduce-examples-3.4.1.jar pi 40 10000000**, para a execução do algoritmo *Pi*, como mostra a Figura 135.

Figura 135 – Execução do algoritmo *Pi*.

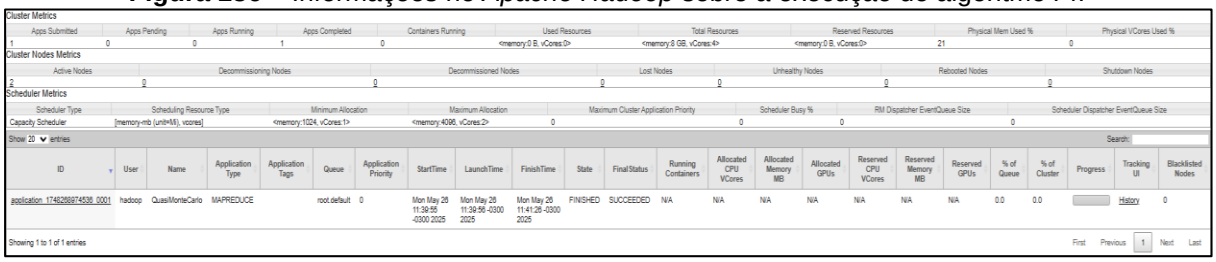


Fonte: Autores, 2025.

Lembrando que o comando **time**, quando escrito na frente da linha de comando, permite que seja verificado o tempo de execução do teste. Esse parâmetro é fundamental para avaliar o comportamento do *cluster* durante a execução dos testes de *benchmark*, bem como compará-lo com outros cenários de testes.

Na Figura 136, observa-se todas as informações referentes a execução do algoritmo *Pi* diretamente da página de métricas do *Apache Hadoop*, bem como os recursos de *hardware* utilizados durante a execução dos testes. Nela, é possível verificar a quantidade de nós (*nodes*) ativos, no campo **Active Nodes**, além da data e hora de início e término da execução dos testes, indicadas nos campos **StartTime** e **FinishTime**. Além disso, observa-se se o teste foi concluído com sucesso ou não, como pode ser verificado no campo **FinalStatus**. Por fim, no campo **Total Resources**, é possível verificar os recursos de *hardware* disponíveis durante os testes no *cluster*, incluindo os núcleos de processamento (4 núcleos) e a memória *RAM* (8 *GByte*).

Figura 136 – Informações no *Apache Hadoop* sobre a execução do algoritmo *Pi*.



Fonte: Autores, 2025.

A Figura 137 mostra a utilização do espaço em disco nos nós (*nodes*) do *cluster*, após a execução dos testes envolvendo o algoritmo *Pi*.

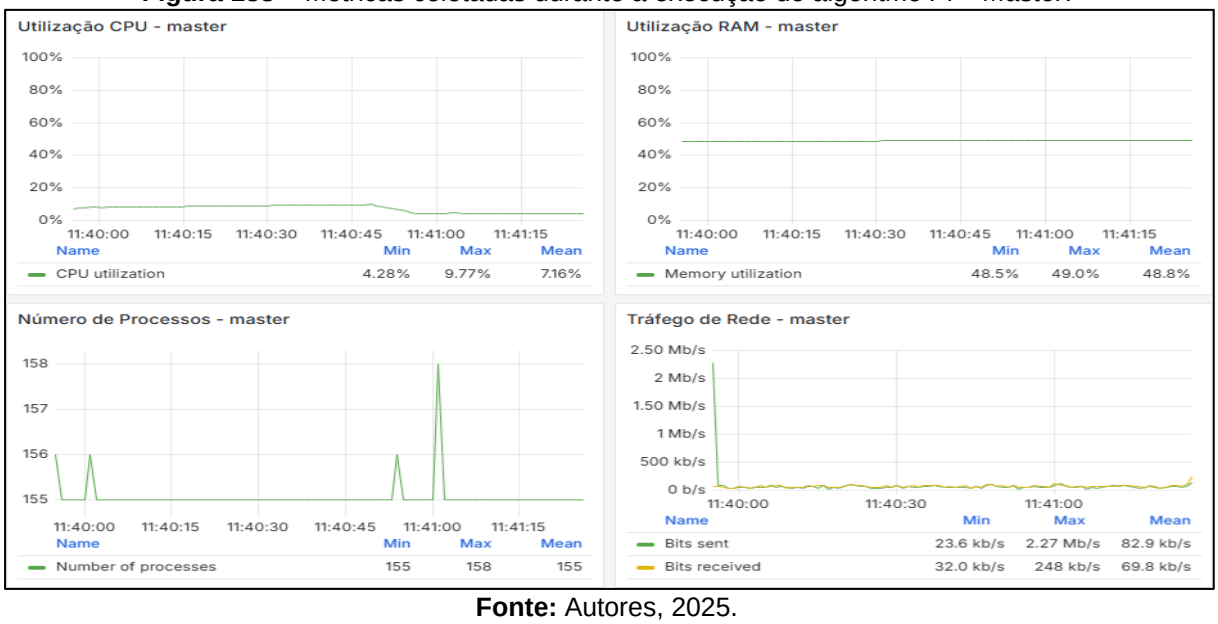
Figura 137 – Capacidade do *cluster* após a execução do algoritmo *Pi*.

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Block pool usage StdDev	Version
/default-rack/slave1:9866 (192.168.1.101:9866)	http://slave1:9864	1s	66m	524 KB	4.4 GB	18.58 GB	3	524 KB (0%)	0%	3.4.1
/default-rack/slave2:9866 (192.168.1.102:9866)	http://slave2:9864	0s	21m	524 KB	4.39 GB	18.58 GB	3	524 KB (0%)	0%	3.4.1

Fonte: Autores, 2025.

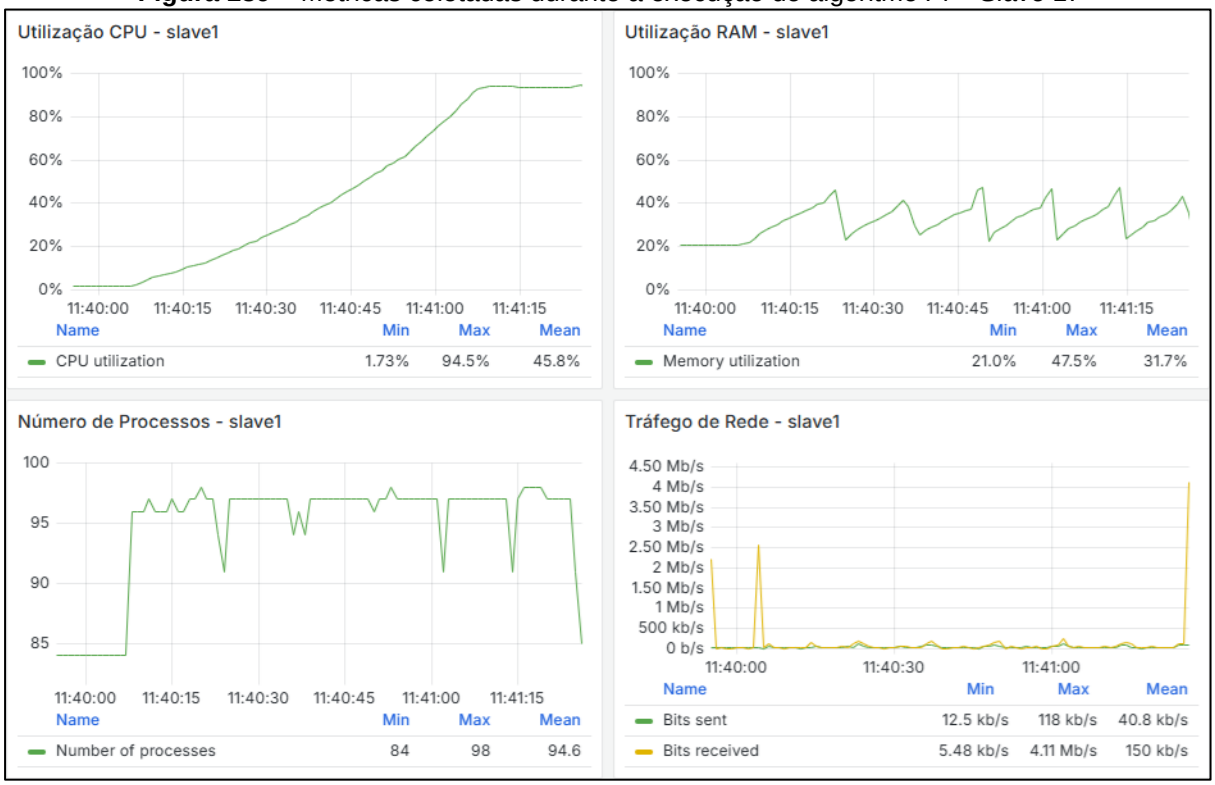
Já as Figuras 138, 139 e 140, respectivamente, apresentam os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução dos testes envolvendo o algoritmo *Pi*.

Figura 138 – Métricas coletadas durante a execução do algoritmo *Pi – Master*.



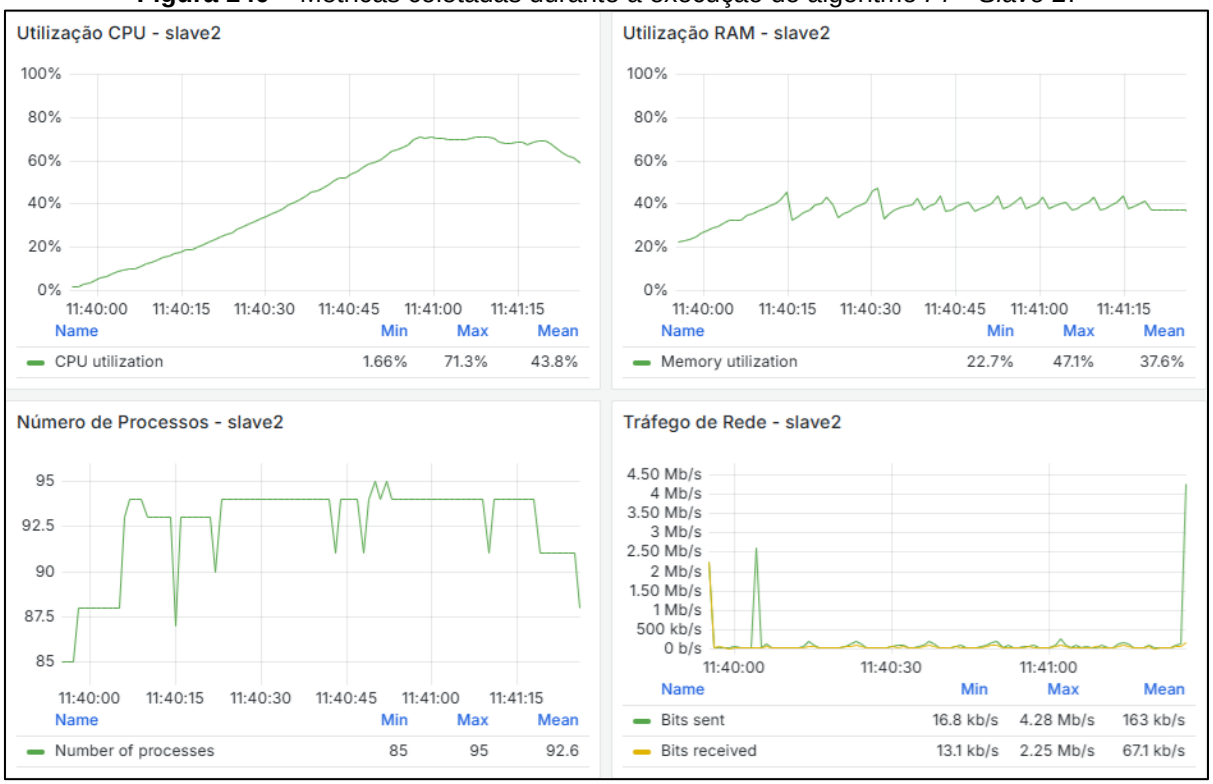
Fonte: Autores, 2025.

Figura 139 – Métricas coletadas durante a execução do algoritmo *Pi – Slave 1*.



Fonte: Autores, 2025.

Figura 140 – Métricas coletadas durante a execução do algoritmo *Pi – Slave 2*.



Fonte: Autores, 2025.

6.2 Algoritmos “*Tera*” (*TeraGen - TeraSort - TeraValidate*)

Nesta seção, serão apresentados os procedimentos para a realização dos testes envolvendo os algoritmos “*Tera*”. Para os testes, foi utilizado um arquivo com tamanho de 1 *GByte*. Lembrando que os comandos descritos a seguir deverão ser executados no modo terminal da máquina *master*.

1º Passo: Parar o *Cluster*

O processo de execução do algoritmo começa com a “parada” do *cluster*. Esse cenário ocorre sempre que um novo teste tiver que ser realizado no *cluster*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **stop-all.sh**. Além disso, essa ação é necessária para limpar o histórico exibido na interface *web* do *cluster*, disponível através da porta 8088, como mostra a Figura 141.

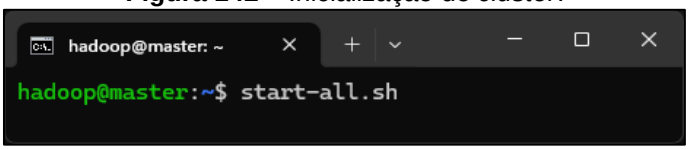
Figura 141 – Parada do *cluster*.

Fonte: Autores, 2025.

2º Passo: Iniciar o *Cluster*

Em seguida, será necessário reiniciar o *cluster*. Para isso, no modo terminal da máquina *master*, execute a seguinte linha de comando: **start-all.sh**, como mostra a Figura 142.

Figura 142 – Inicialização do *cluster*.

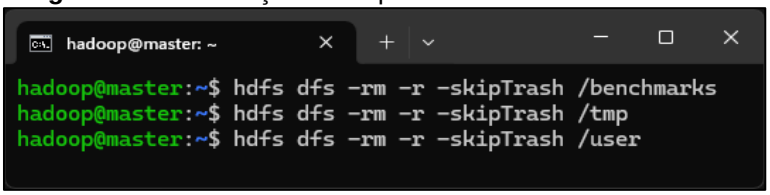


Fonte: Autores, 2025.

3º Passo: Apagar os arquivos armazenados no *Cluster*

Continuando, é preciso apagar os arquivos armazenados no *cluster*, referentes a outros testes, caso tenham sido realizados. Para isso, execute os seguintes comandos no modo terminal da máquina *master*, respectivamente: **hdfs dfs -rm -r -skipTrash /benchmarks**; **hdfs dfs -rm -r -skipTrash /tmp;** e **hdfs dfs -rm -r -skipTrash /user**, como mostra a Figura 143.

Figura 143 – Remoção de arquivos armazenados no *cluster*.

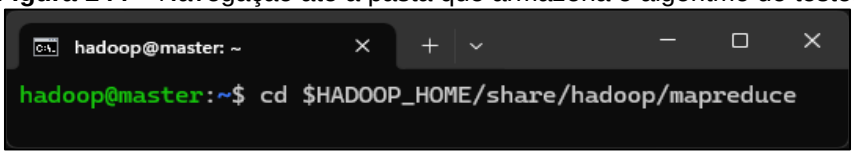


Fonte: Autores, 2025.

4º Passo: Acessar a pasta que armazena os algoritmos de testes do *Hadoop*

Em seguida, é preciso acessar a pasta onde se encontram os algoritmos de teste do *Apache Hadoop*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **cd $HADOOP_HOME/share/hadoop/mapreduce**, como mostra a Figura 144.

Figura 144 – Navegação até a pasta que armazena o algoritmo do teste.



Fonte: Autores, 2025.

5º Passo: Executar o algoritmo *TeraGen*

Após entrar na pasta de testes do *Hadoop*, execute a seguinte linha de comando no modo terminal da máquina *master*: **time yarn jar hadoop-mapreduce-examples-3.4.1.jar teragen 10737418 /user/teragen**, como mostra a Figura 145.

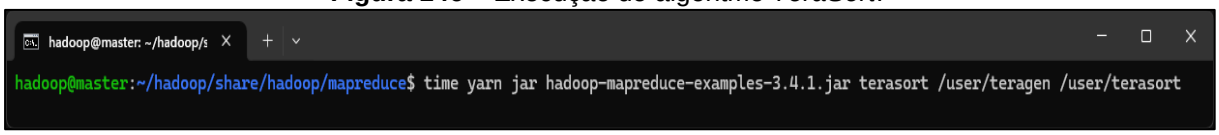
Figura 145 – Execução do algoritmo *TeraGen*.

Fonte: Autores, 2025.

6º Passo: Executar o algoritmo *TeraSort*

Em seguida, execute a seguinte linha de comando no modo terminal da máquina *master*: **time yarn jar hadoop-mapreduce-examples-3.4.1.jar terasort /user/teragen /user/terasort**, para a execução do teste *TeraSort*, como mostra a Figura 146.

Figura 146 – Execução do algoritmo *TeraSort*.

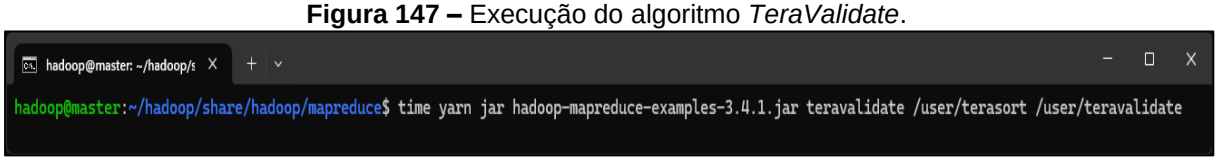


Fonte: Autores, 2025.

7º Passo: Executar o algoritmo *TeraValidate*

Na sequência, para concluir com os testes dos algoritmos "*Tera*", no modo terminal da máquina *master*, execute a seguinte linha de comando: **time yarn jar hadoop-mapreduce-examples-3.4.1.jar teravalidate /user/terasort /user/teravalidate**, para executar o algoritmo *TeraValidate*, como mostra a Figura 147.

Figura 147 – Execução do algoritmo *TeraValidate*.

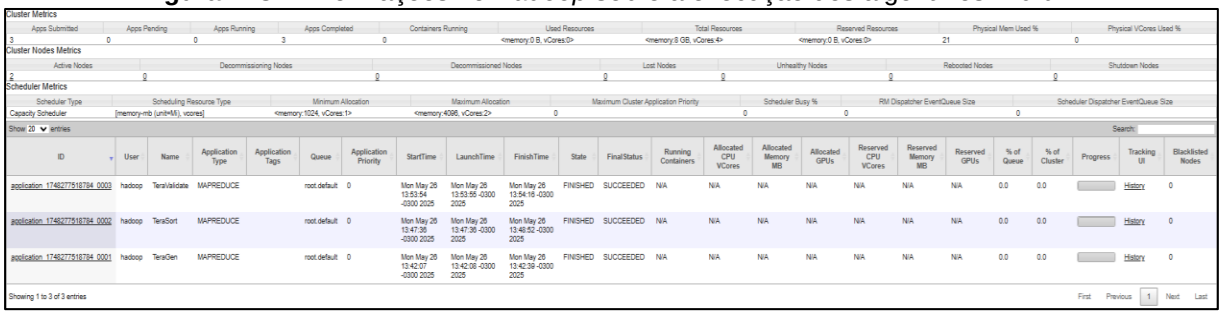


Fonte: Autores, 2025.

Na Figura 148, observa-se todas as informações referentes a execução dos algoritmos "*Tera*" diretamente da página de métricas do *Apache Hadoop*, bem como os recursos de *hardware* utilizados durante a execução dos testes. Nela, é possível

verificar a quantidade de nós (*nodes*) ativos, no campo **Active Nodes**, além da data e hora de início e término da execução dos testes, indicadas nos campos **StartTime** e **FinishTime**. Além disso, observa-se se o teste foi concluído com sucesso ou não, como pode ser verificado no campo **FinalStatus**. Por fim, no campo **Total Resources**, é possível verificar os recursos de *hardware* disponíveis durante os testes no *cluster*, incluindo os núcleos de processamento (4 núcleos) e a memória *RAM* (8 *GByte*).

Figura 148 – Informações no *Hadoop* sobre a execução dos algoritmos *"Tera"*.



Fonte: Autores, 2025.

A Figura 149 mostra a utilização do espaço em disco nos nós (*nodes*) do *cluster*, após a execução dos testes envolvendo os algoritmos "*Tera*".

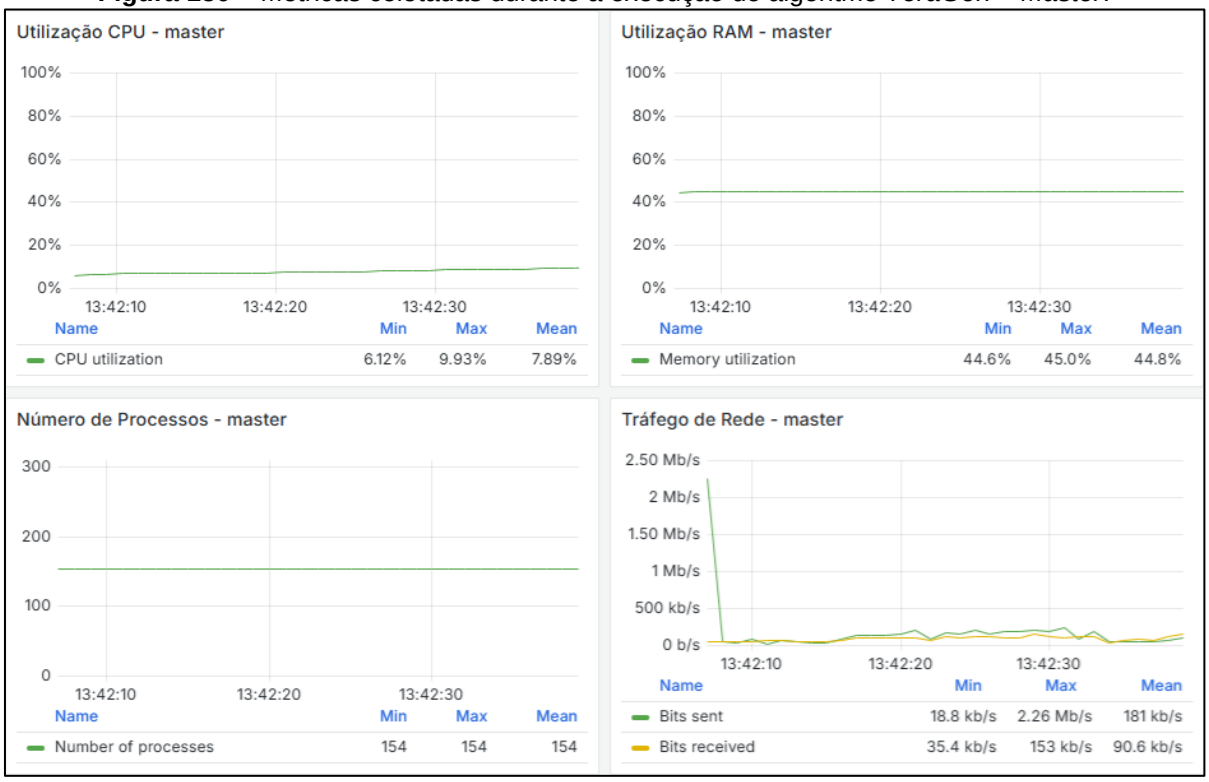
Figura 149 – Capacidade do *cluster* após a execução do algoritmo "*Tera"*.

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Block pool usage StdDev	Version
✔/default-rack/slave1:9866 (192.168.1.101:9866)	http://slave1:9864	1s	21m	1.01 GB	4.41 GB	18.58 GB	15	1.01 GB (5.43%)	0%	3.4.1
✔/default-rack/slave2:9866 (192.168.1.102:9866)	http://slave2:9864	0s	21m	2.02 GB	4.39 GB	18.58 GB	26	2.02 GB (10.86%)	0%	3.4.1

Fonte: Autores, 2025.

Nas Figuras 150, 151 e 152, observa-se os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução do teste envolvendo o algoritmo *TeraGen*.

Figura 150 – Métricas coletadas durante a execução do algoritmo *TeraGen – Master*.

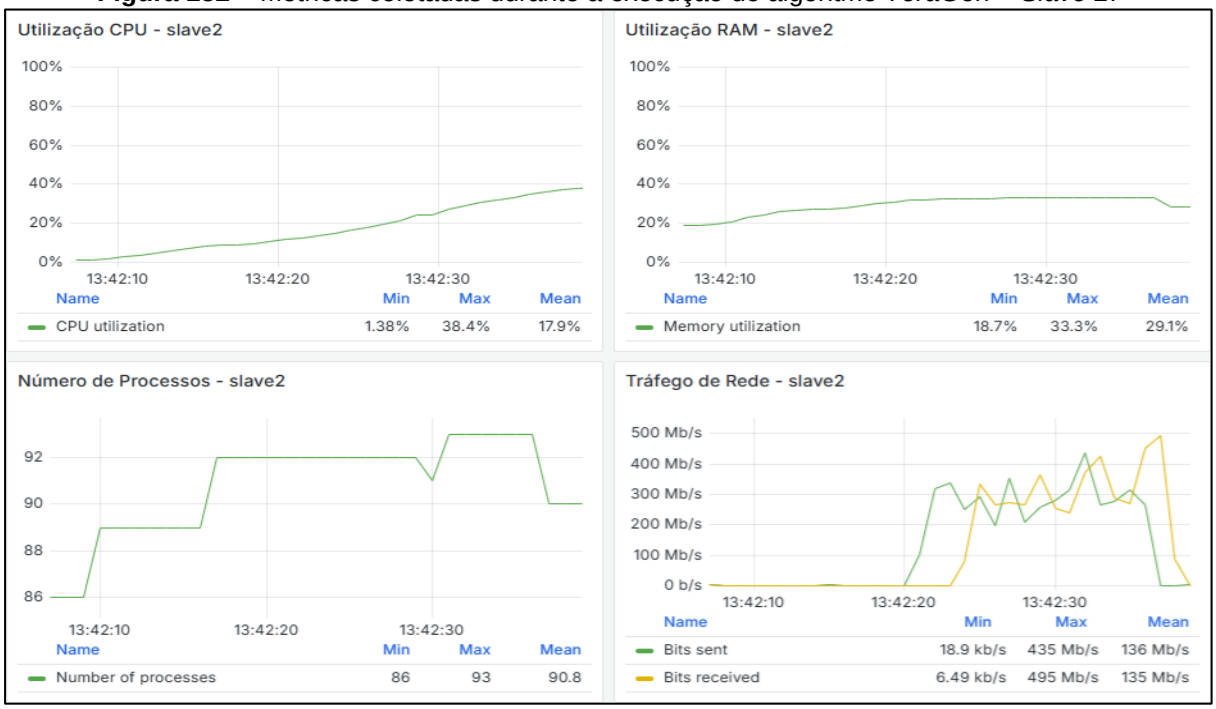


Fonte: Autores, 2025.

Figura 151 – Métricas coletadas durante a execução do algoritmo *TeraGen – Slave 1*.

Fonte: Autores, 2025.

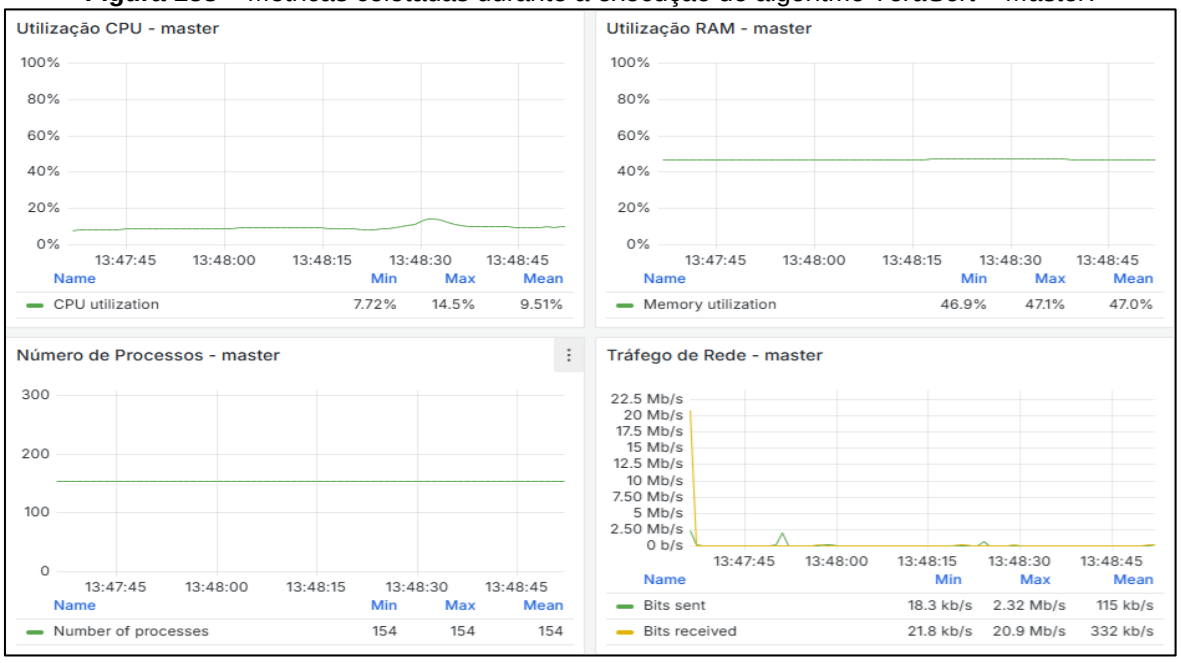
Figura 152 – Métricas coletadas durante a execução do algoritmo *TeraGen – Slave 2*.



Fonte: Autores, 2025.

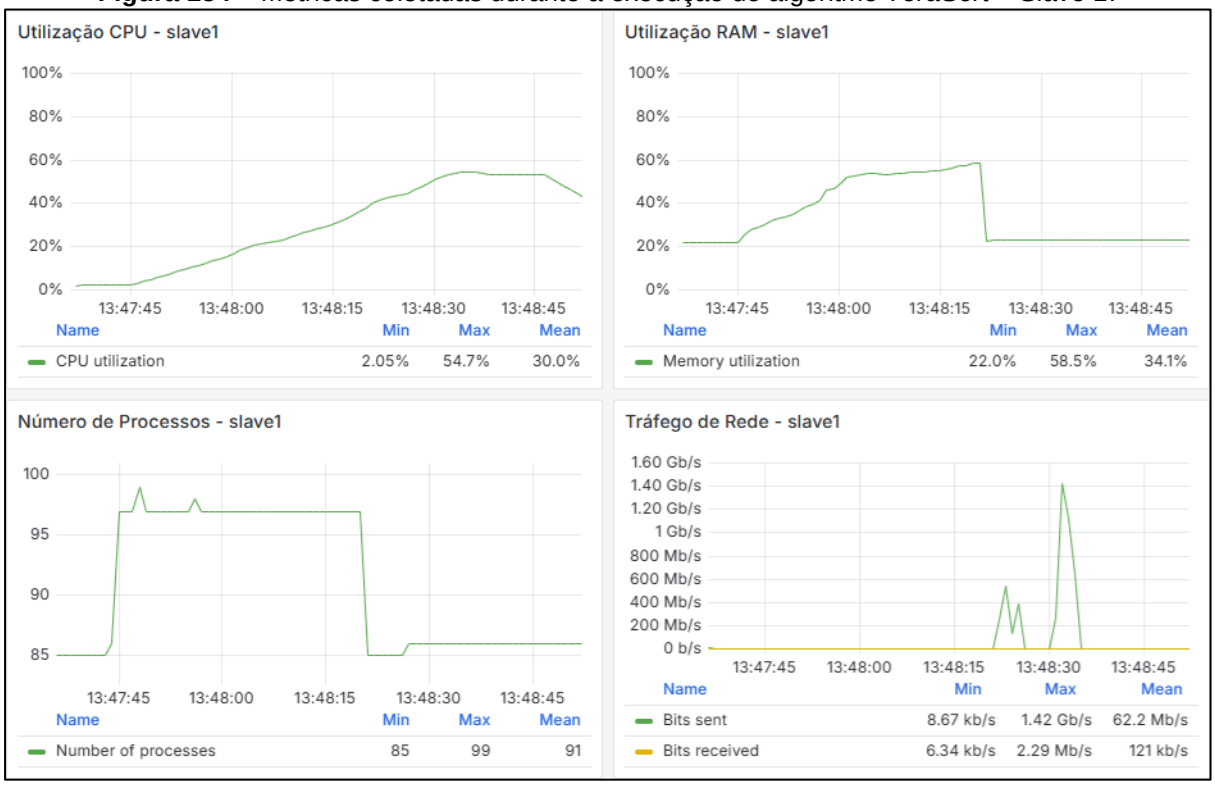
Já as Figuras 153, 154 e 155, apresentam os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução do teste envolvendo o algoritmo *TeraSort*.

Figura 153 – Métricas coletadas durante a execução do algoritmo *TeraSort – Master*.



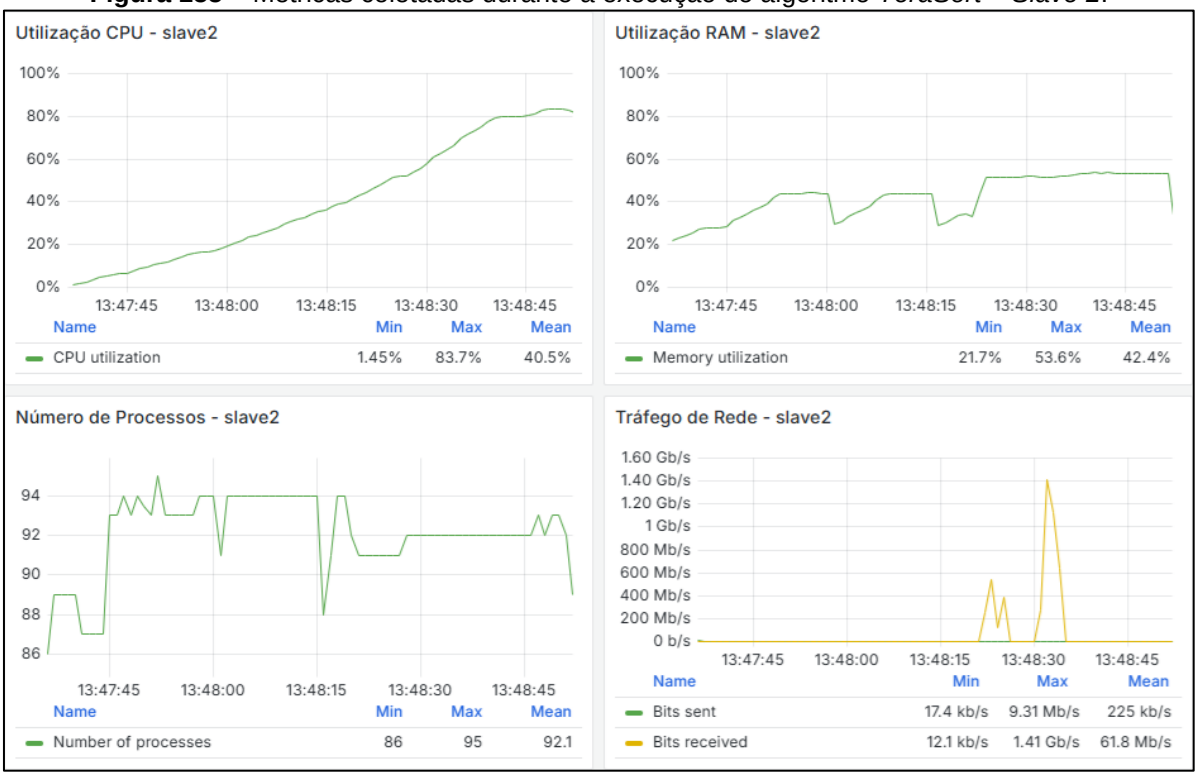
Fonte: Autores, 2025.

Figura 154 – Métricas coletadas durante a execução do algoritmo *TeraSort – Slave 1*.



Fonte: Autores, 2025.

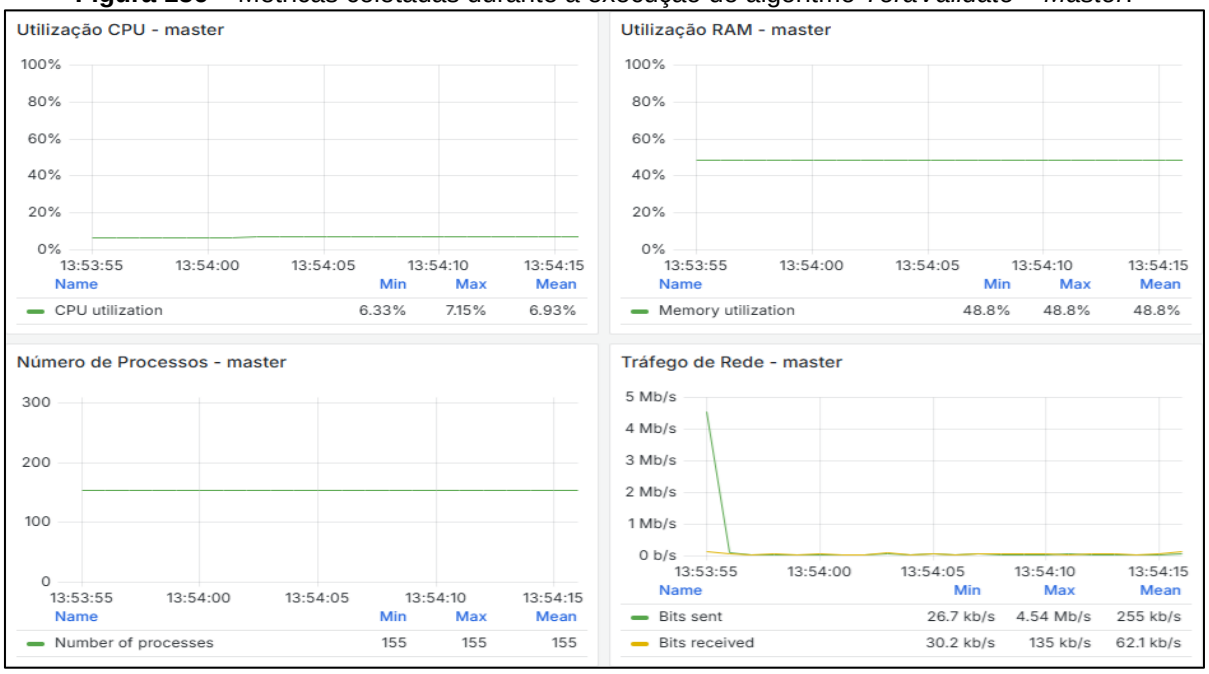
Figura 155 – Métricas coletadas durante a execução do algoritmo *TeraSort – Slave 2*.



Fonte: Autores, 2025.

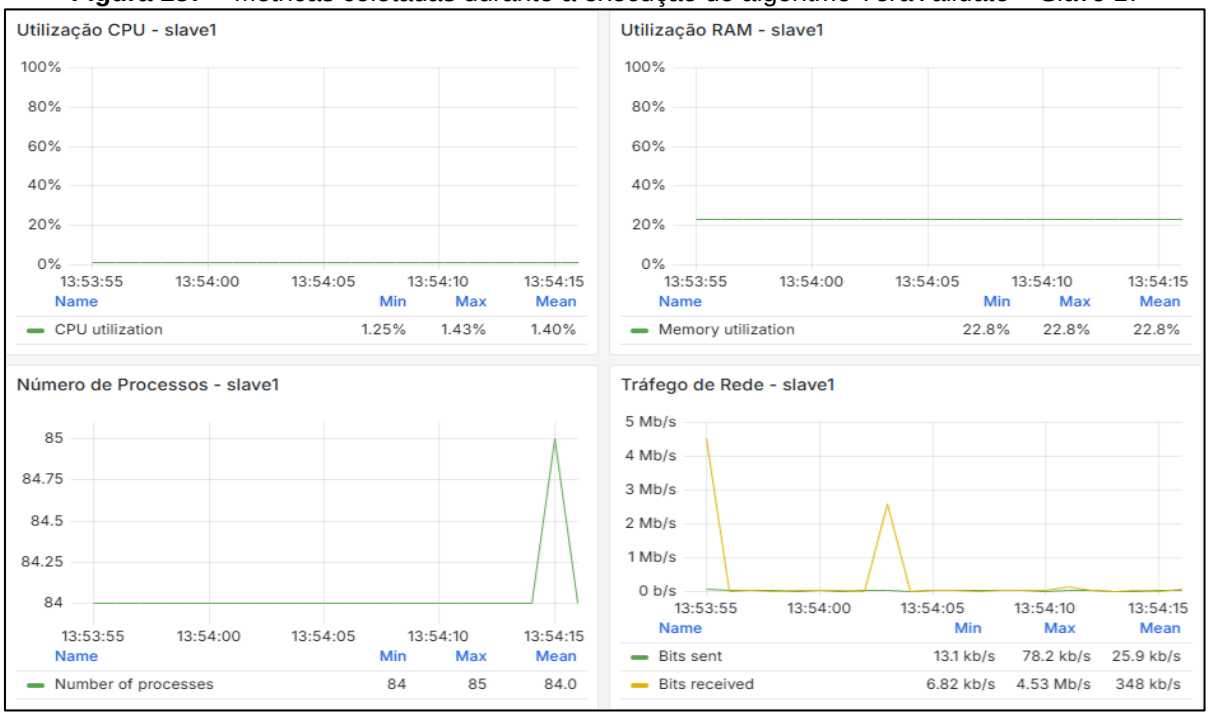
Nas Figuras 156, 157 e 158, observa-se os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução do teste envolvendo o algoritmo *TeraValidate*.

Figura 156 – Métricas coletadas durante a execução do algoritmo *TeraValidate – Master*.



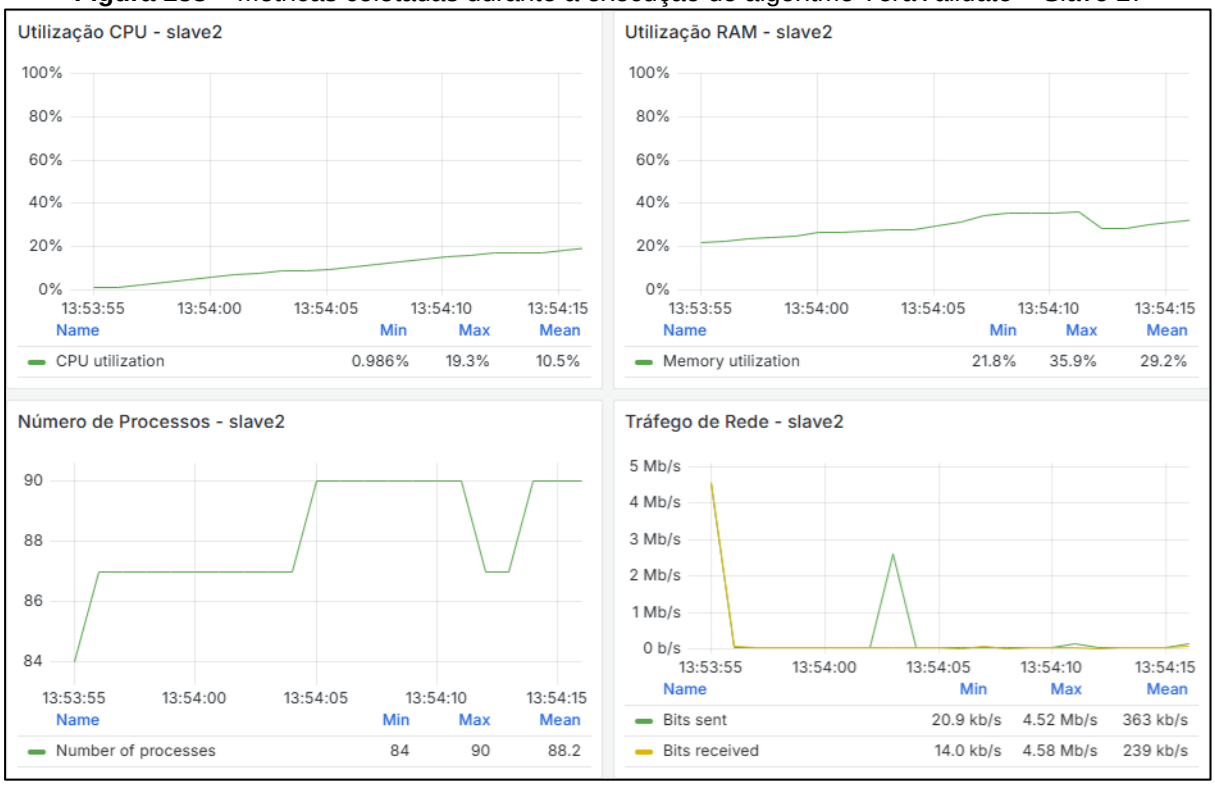
Fonte: Autores, 2025.

Figura 157 – Métricas coletadas durante a execução do algoritmo *TeraValidate – Slave 1*.



Fonte: Autores, 2025.

Figura 158 – Métricas coletadas durante a execução do algoritmo *TeraValidate – Slave 2*.



Fonte: Autores, 2025.

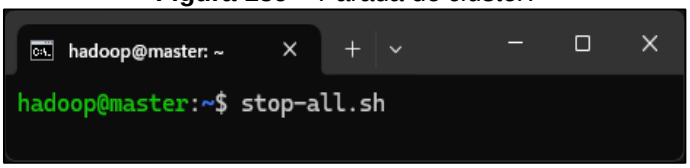
6.3 Algoritmos *TestDFSIO* (*Write / Read*)

Nesta seção, serão apresentados os procedimentos para a realização dos testes envolvendo o algoritmo *TestDFSIO*. Para os testes, foi utilizado um arquivo com tamanho de 1 *GByte*. Lembrando que os comandos descritos a seguir devem ser executados no modo terminal da máquina *master*.

1º Passo: Parar o *Cluster*

O processo de execução do algoritmo começa com a “parada” do *cluster*. Esse cenário ocorre sempre que um novo teste tiver que ser realizado no *cluster*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **stop-all.sh**. Além disso, essa ação é necessária para limpar o histórico exibido na interface *web* do *cluster*, disponível através da porta 8088, como mostra a Figura 159.

Figura 159 – Parada do *cluster*.

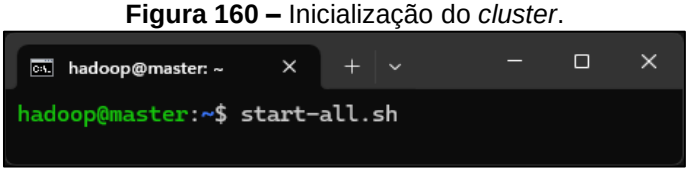


Fonte: Autores, 2025.

2º Passo: Iniciar o *Cluster*

Em seguida, será necessário reiniciar o *cluster*. Para isso, no modo terminal da máquina *master*, execute a seguinte linha de comando: **start-all.sh**, como mostra a Figura 160.

Figura 160 – Inicialização do *cluster*.

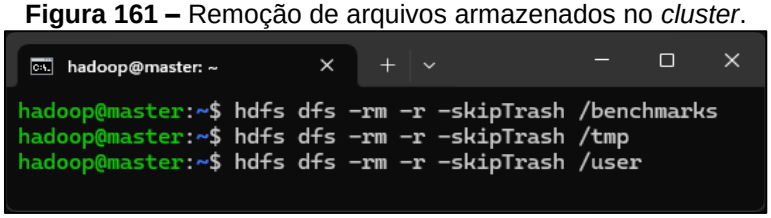


Fonte: Autores, 2025.

3º Passo: Apagar os arquivos armazenados no *Cluster*

Continuando, é preciso apagar os arquivos armazenados no *cluster*, referentes a outros testes, caso tenham sido realizados. Para isso, execute os seguintes comandos no modo terminal da máquina *master*, respectivamente: **hdfs dfs -rm -r -skipTrash /benchmarks**; **hdfs dfs -rm -r -skipTrash /tmp;** e **hdfs dfs -rm -r -skipTrash /user**, como mostra a Figura 161.

Figura 161 – Remoção de arquivos armazenados no *cluster*.

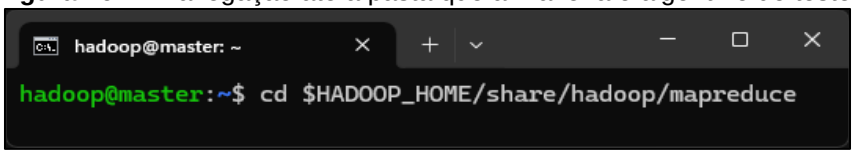


Fonte: Autores, 2025.

4º Passo: Acessar a pasta que armazena os algoritmos de testes do *Hadoop*

Em seguida, é preciso acessar a pasta onde se encontram os algoritmos de teste do *Apache Hadoop*. Para isso, execute a seguinte linha de comando no modo terminal da máquina *master*: **cd $HADOOP_HOME/share/hadoop/mapreduce**, como mostra a Figura 162.

Figura 162 – Navegação até a pasta que armazena o algoritmo do teste.

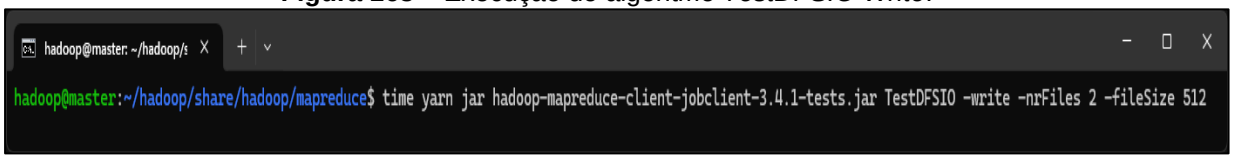


Fonte: Autores, 2025.

5º Passo: Executar o algoritmo *TestDFSIO* (*Write*)

Após entrar na pasta de testes do *Hadoop*, execute a seguinte linha de comando no modo terminal da máquina *master*: **time yarn jar hadoop-mapreduce-client-jobclient-3.4.1-tests.jar TestDFSIO -write -nrFiles 2 -fileSize 512**, como mostra a Figura 163.

Figura 163 – Execução do algoritmo *TestDFSIO Write*.

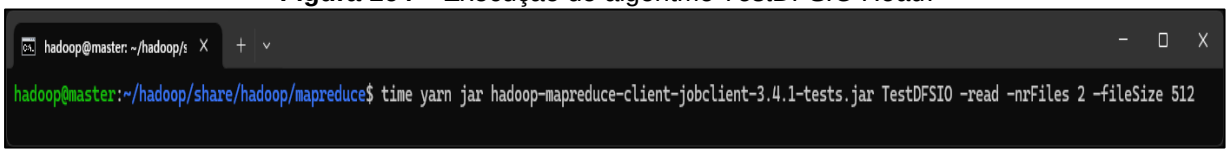


Fonte: Autores, 2025.

6º Passo: Executar o algoritmo *TestDFSIO Read*

Em seguida, para executar o teste de leitura, execute a seguinte linha de comando no modo terminal da máquina *master*: **time yarn jar hadoop-mapreduce-client-jobclient-3.4.1-tests.jar TestDFSIO -read -nrFiles 2 -fileSize 512**, como mostra a Figura 164.

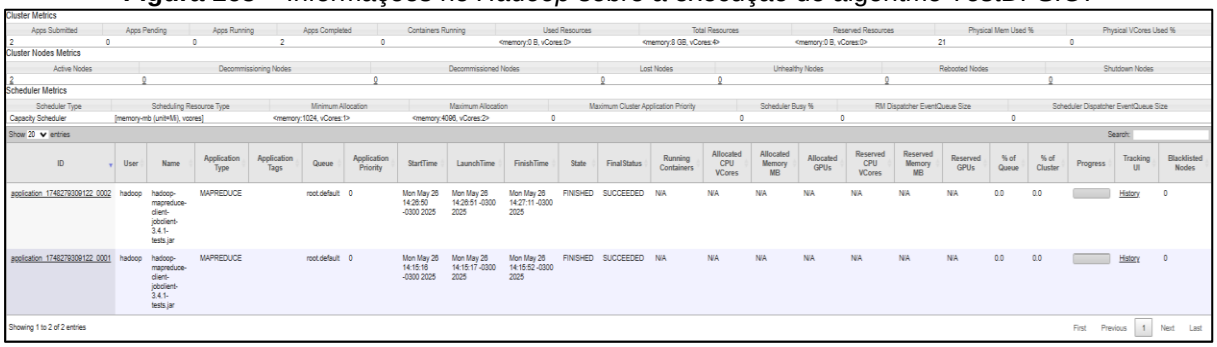
Figura 164 – Execução do algoritmo *TestDFSIO Read*.



Fonte: Autores, 2025.

Na Figura 165, observa-se todas as informações referentes a execução dos algoritmos "*TestDFSIO*" diretamente da página de métricas do *Apache Hadoop*, bem como os recursos de *hardware* utilizados durante a execução dos testes. Nela, é possível verificar a quantidade de nós (*nodes*) ativos, no campo **Active Nodes**, além da data e hora de início e término da execução dos testes, indicadas nos campos **StartTime** e **FinishTime**. Além disso, observa-se se o teste foi concluído com sucesso ou não, como pode ser verificado no campo **FinalStatus**. Por fim, no campo **Total Resources**, é possível verificar os recursos de *hardware* disponíveis durante os testes no *cluster*, incluindo os núcleos de processamento (4 núcleos) e a memória *RAM* (8 *GByte*).

Figura 165 – Informações no *Hadoop* sobre a execução do algoritmo *TestDFSIO*.



Fonte: Autores, 2025.

A Figura 166 mostra a utilização do espaço em disco nos nós (*nodes*) do *cluster*, após a execução dos testes envolvendo o algoritmo *TestDFSIO*.

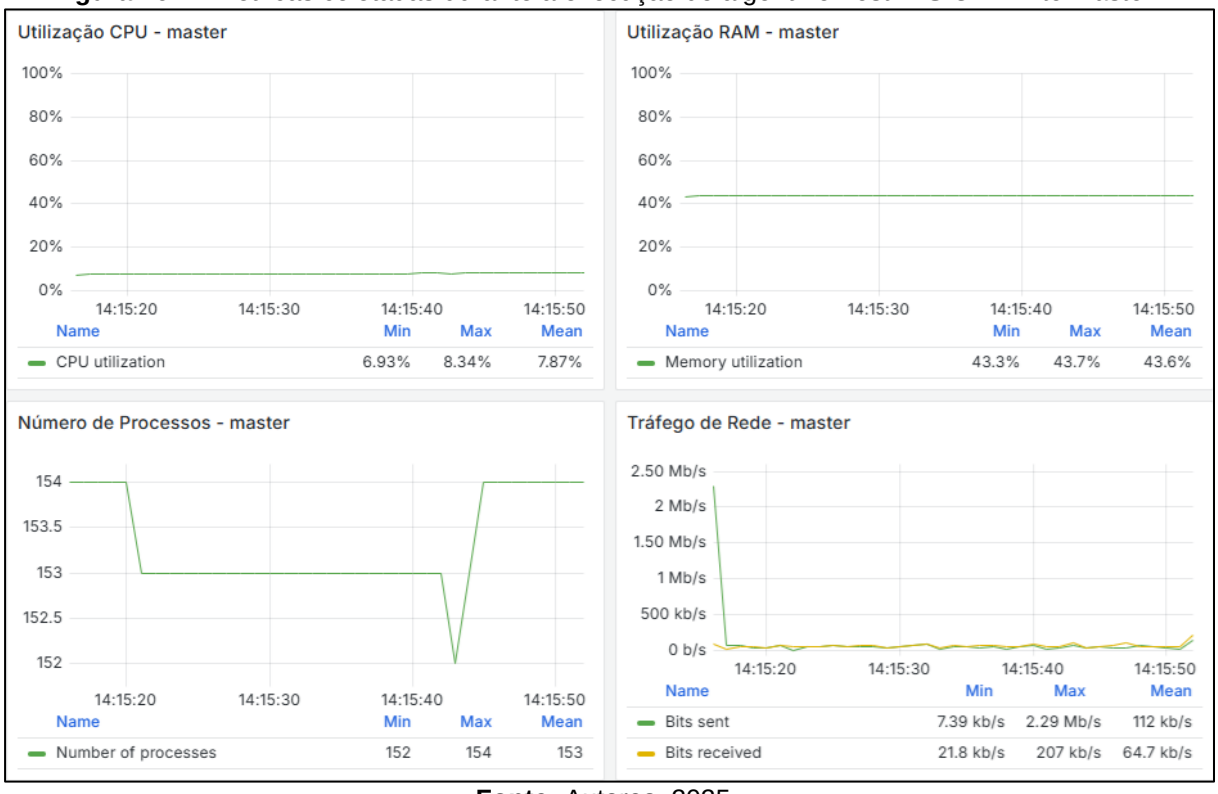
Figura 166 – Capacidade do *cluster* após a execução do algoritmo *TestDFSIO*.

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Block pool usage StdDev	Version
✔/default-rack/slave1:9866 (192.168.1.101:9866)	http://slave1:9864	0s	21m	1.01 GB	4.41 GB	18.58 GB	18	1.01 GB (5.43%)	0%	3.4.1
✔/default-rack/slave2:9866 (192.168.1.102:9866)	http://slave2:9864	1s	21m	1.01 GB	4.39 GB	18.58 GB	18	1.01 GB (5.43%)	0%	3.4.1

Fonte: Autores, 2025.

Já as Figuras 167, 168 e 169, apresentam os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução do teste envolvendo o algoritmo *TestDFSIO Write*.

Figura 167 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Write Master*.

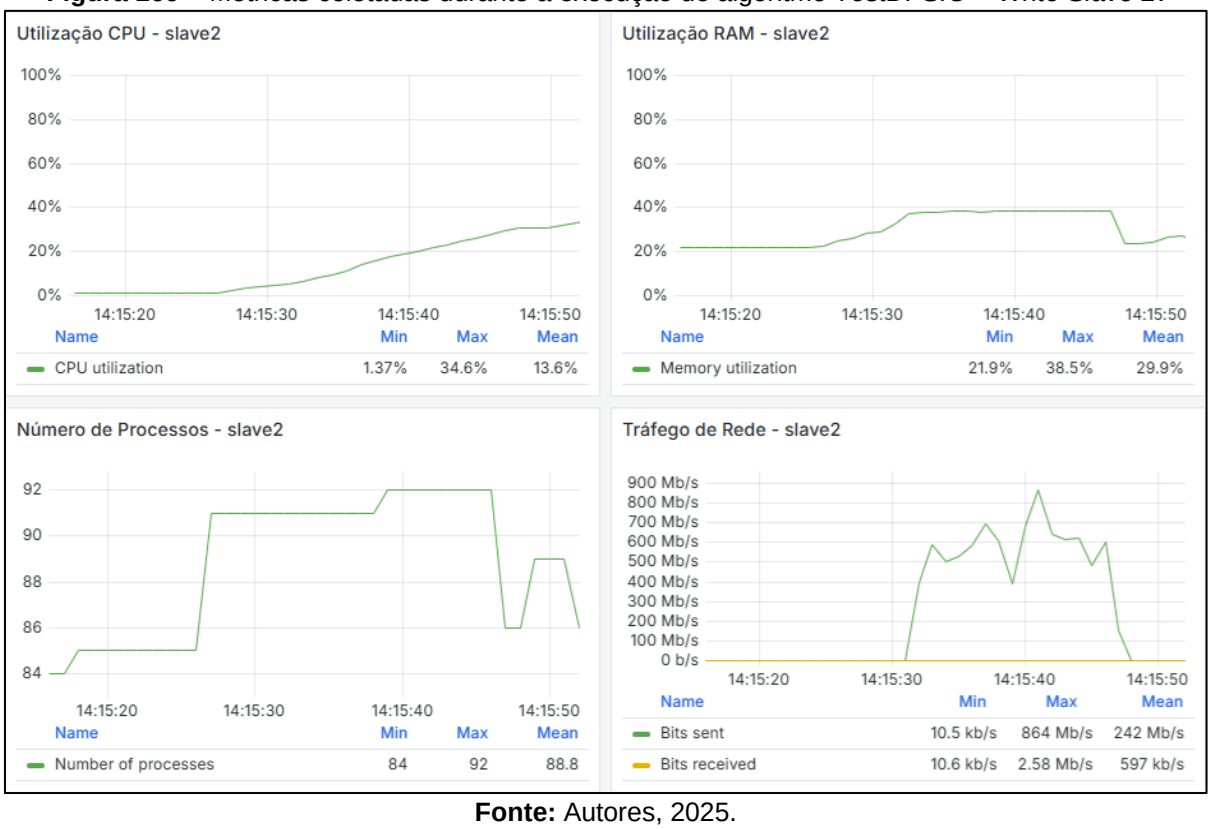


Fonte: Autores, 2025.

Figura 168 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Write Slave 1*.

Fonte: Autores, 2025.

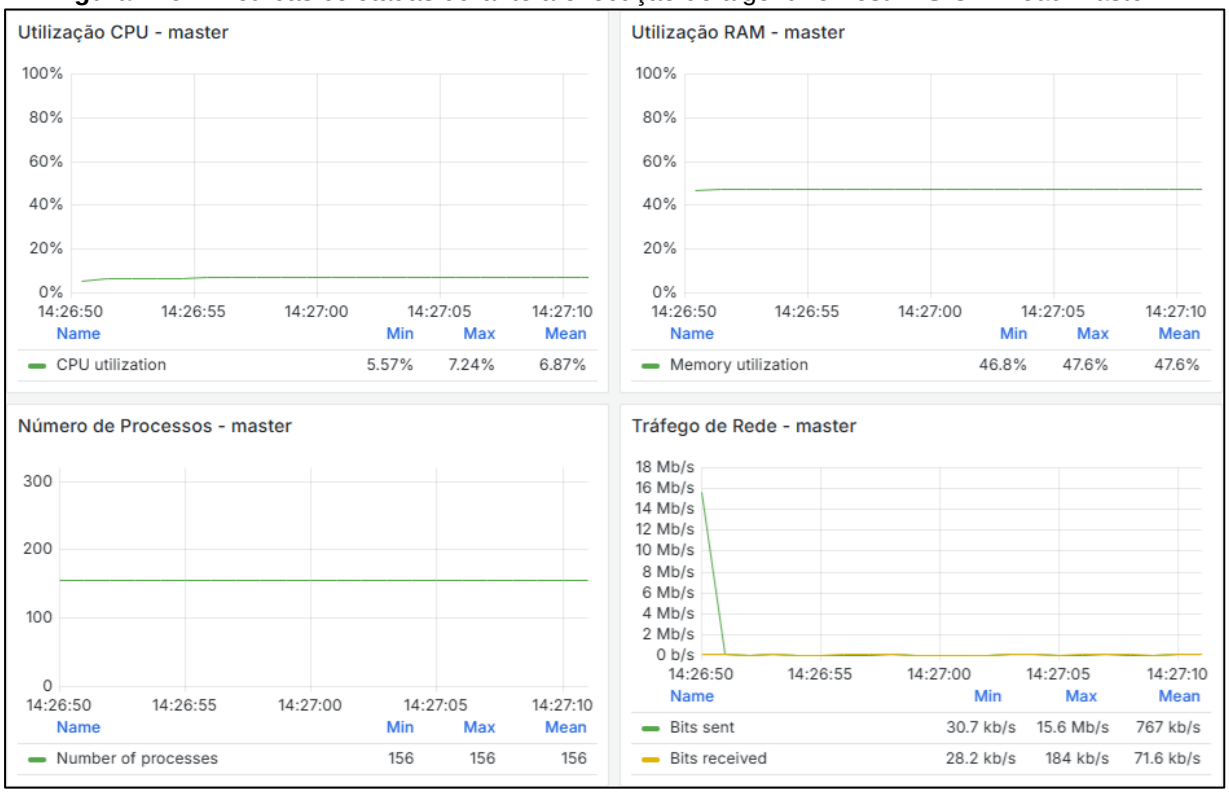
Figura 169 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Write Slave 2*.



Fonte: Autores, 2025.

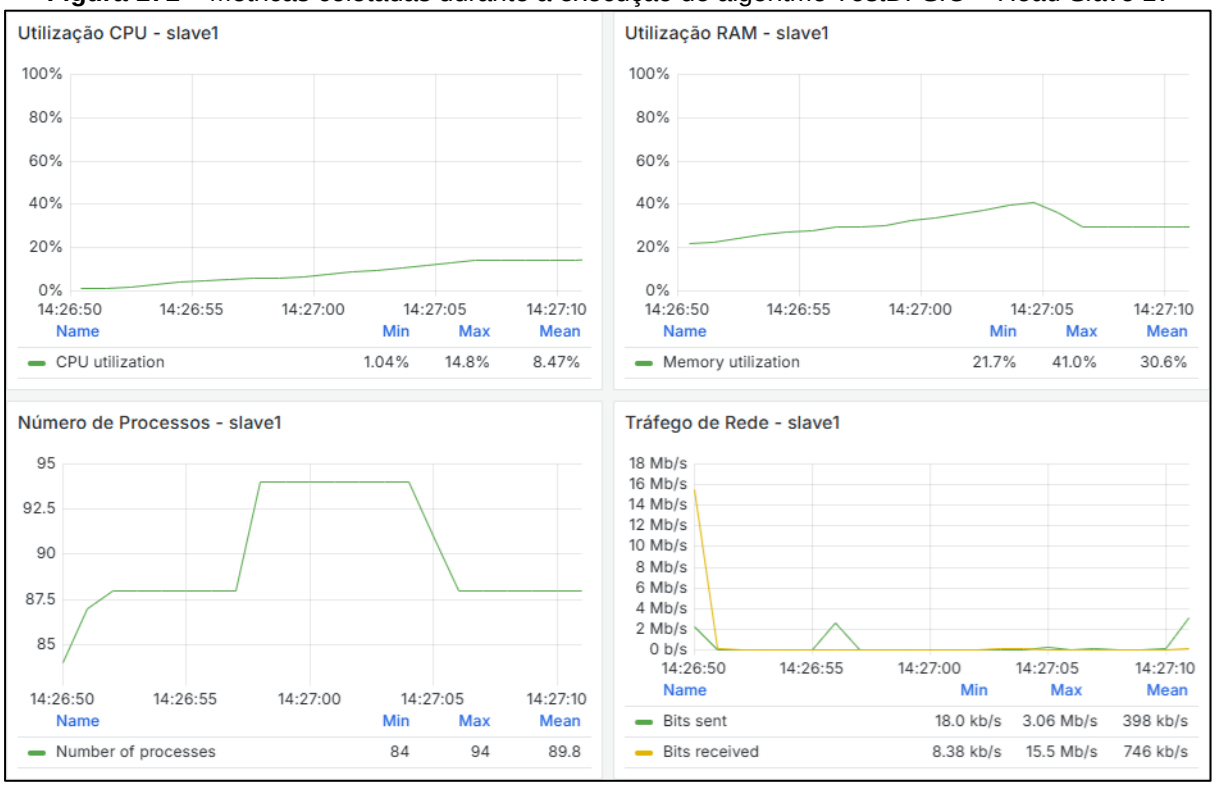
Nas Figuras 170, 171 e 172, observa-se os gráficos com as métricas de desempenho do *cluster* coletadas durante a execução do teste envolvendo o algoritmo *TestDFSIO Read*.

Figura 170 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Read Master*.



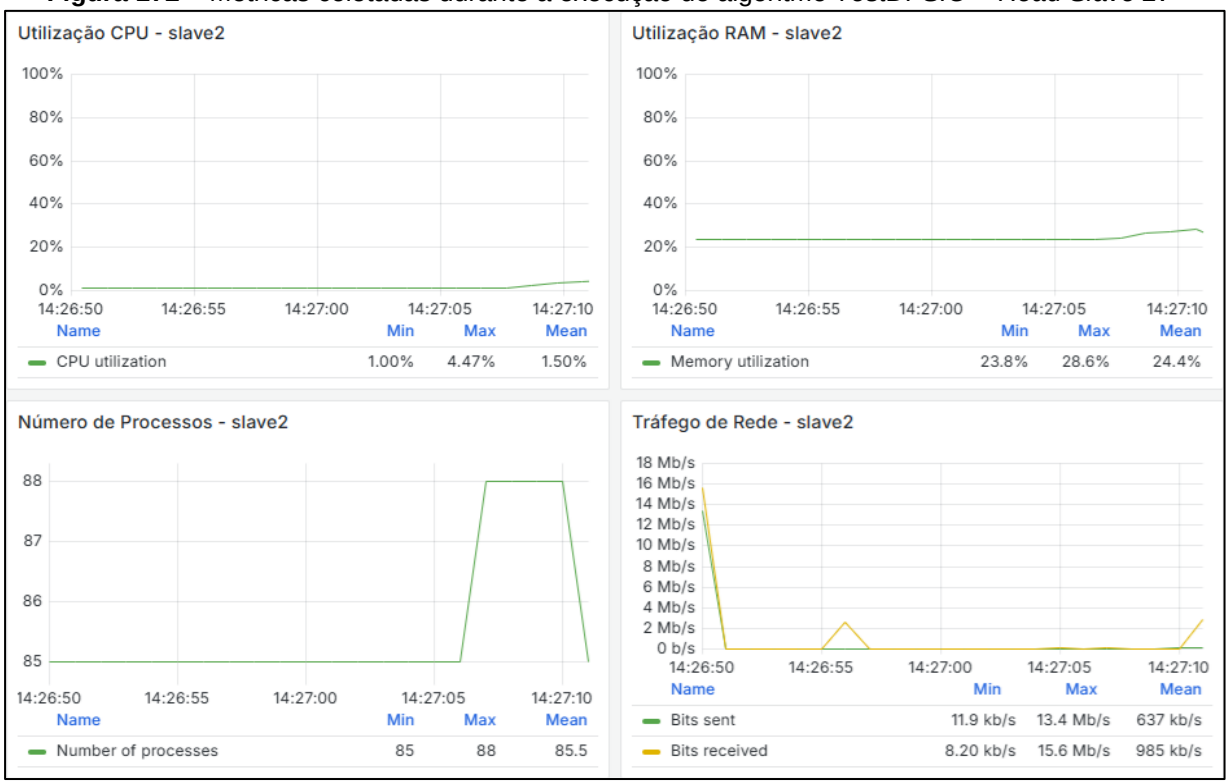
Fonte: Autores, 2025.

Figura 171 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Read Slave 1*.



Fonte: Autores, 2025.

Figura 172 – Métricas coletadas durante a execução do algoritmo *TestDFSIO – Read Slave 2*.



Fonte: Autores, 2025.

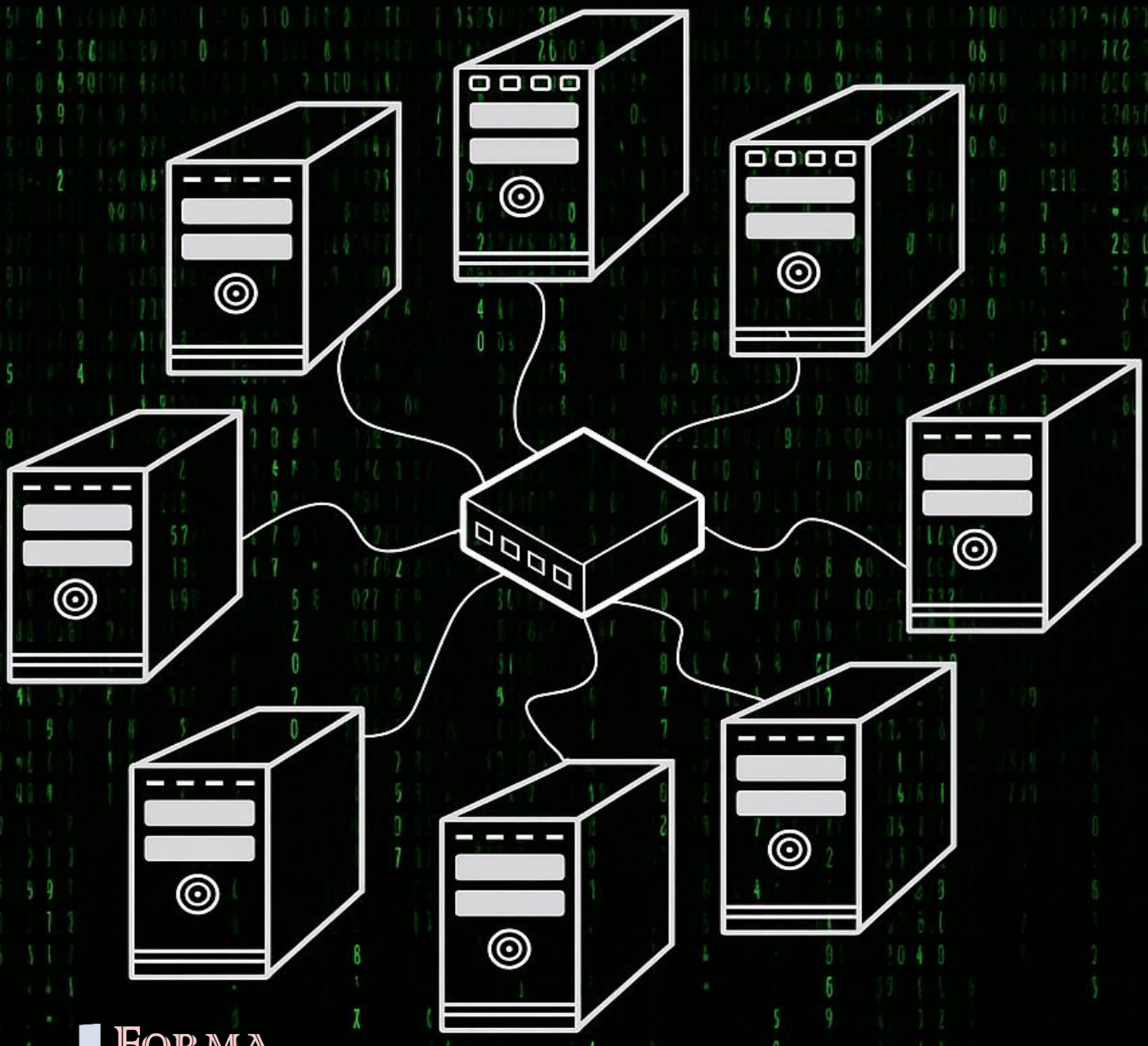
SOBRE OS AUTORES

João Fernandes Santos Filho

Bacharel em Ciência da Computação pelo Instituto Federal de Sergipe – IFS (2024). Técnico em Manutenção e Suporte em Informática pelo Instituto Federal de Sergipe – IFS (2020).
E-mail: joaofilho8274@gmail.com
Lattes: https://lattes.cnpq.br/0063920239671194

José Aprígio Carneiro Neto

Pós-Doutor em Engenharia e Computação Inteligente pelo Instituto Politécnico do Porto ISEP/IPP Porto Portugal. Pós-Doutor em Engenharia de Produção e Sistemas pela Universidade do Minho (UNIMINHO) - Braga - Portugal. Pós-Doutor em Ciência da Computação pela Universidade Federal de Sergipe (UFS). Doutor em Ciência da Propriedade Intelectual pela Universidade Federal de Sergipe (UFS). Mestre em Engenharia de Software pelo Centro de Estudos e Sistemas Avançados do Recife - C.E.S.A.R. EDU. Especialista em Tecnologias da Informação pela Universidade Federal do Ceará (UFC). Graduado em Processamento de Dados pela Universidade Estadual do Piauí (UESPI). Graduado em Formação Pedagógica em Informática pelo Centro Universitário Leonardo Da Vinci (UNIASSELVI). Professor efetivo do Instituto Federal de Sergipe (IFS) na área de Ciência da Computação.
E-mail: aprigio.carneiro.ac@gmail.com
Lattes: http://lattes.cnpq.br/6225402292538909



FORMA EDUCACIONAL

ISBN 978-658517547-0