

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS - GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

EMISVALDO PEREIRA DA SILVA

ABORDAGEM PARA PREVISÃO DA  
DISPONIBILIDADE DE RECURSOS  
COMPUTACIONAIS UTILIZANDO CARTA DE  
CONTROLE I-AM

EXAME DE QUALIFICAÇÃO

PONTA GROSSA

2023

**EMISVALDO PEREIRA DA SILVA**

**ABORDAGEM PARA PREVISÃO DA  
DISPONIBILIDADE DE RECURSOS  
COMPUTACIONAIS UTILIZANDO CARTA DE  
CONTROLE I-AM**

Exame de Qualificação apresentado ao Programa  
de Pós-Graduação em Ciência da Computação  
da Universidade Tecnológica Federal do Paraná  
– Campus Ponta Grossa. Área de Concentração:  
Sistemas e Métodos De Computação

Orientador(a): Prof. Dr. Lourival Aparecido De  
Góis

**PONTA GROSSA**

**2023**

## RESUMO

Da Silva, Emisvaldo Pereira. **Abordagem para previsão da disponibilidade de recursos computacionais utilizando carta de controle I-AM.** 2023. Número total de XX páginas. Exame de Qualificação (Mestrado em Ciências da Computação) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2023.

Nesta pesquisa, é explorada a aplicação de cartas de controle para monitorar a disponibilidade da CPU em sistemas distribuídos, onde a variação no uso de CPU é um fator intrínseco. As cartas de controle são gráficos que comparam o desempenho de um processo com limites de controle. Este estudo apresenta uma abordagem específica para monitorar os índices de utilização da CPU em máquinas de um sistema distribuído, visando prever sua disponibilidade.

**Palavras-chave:** CPU. Sistemas distribuídos. Controle de processos. Carta de controle.

## ABSTRACT

Da Silva, Emisvaldo Pereira. **CPU availability prediction in distributed systems using process control.** 2022. Total number of XX pages. Partial Exam (Master in Computer Science) – Federal University of Technology - Paraná. Ponta Grossa, 2023.

In this research, the application of control charts to monitor CPU availability in distributed systems is explored, where variation in CPU usage is an intrinsic factor. Control charts are graphs that compare the performance of a process with control limits. This study presents a specific approach to monitor CPU utilization rates on machines in a distributed system, aiming to predict their availability.

**Keywords:** CPU. Distributed systems. Process control. Control chart.

## Lista de Figuras

1	Fluxograma do processo de seleção dos estudos incluídos na revisão sistemática. <sup>1</sup>	25
2	Categorização das abordagens de previsão de disponibilidade de recursos, mostrando as técnicas, algoritmos ou modelos utilizados em cada categoria. <sup>2</sup>	28
3	Sistema distribuído abordado. <sup>3</sup> . . . . .	29

## Lista de Tabelas

1	Estudos incluídos, incluindo o título do estudo, autores, ano de publicação. <sup>4</sup>	27
---	---	----

# Conteúdo

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>INTRODUÇÃO</b>	<b>8</b>
1.1	MOTIVAÇÃO . . . . .	10
1.2	OBJETIVO GERAL . . . . .	12
1.3	OBJETIVOS ESPECÍFICOS . . . . .	12
1.4	HIPÓTESES . . . . .	13
1.5	JUSTIFICATIVA . . . . .	14
1.6	METODOLOGIA . . . . .	15
1.6.1	Plano de Trabalho . . . . .	16
<b>2</b>	<b>REVISÃO SISTEMÁTICA</b>	<b>18</b>
2.1	DEFINIÇÃO DOS CRITÉRIOS DE BUSCA . . . . .	19
2.1.1	Termos de Pesquisa: . . . . .	19
2.1.2	Sinônimos e Variações: . . . . .	19
2.1.3	Objetivos da Revisão Sistemática . . . . .	20
2.1.4	Perguntas de Pesquisa . . . . .	20
2.2	FONTES DE INFORMAÇÃO . . . . .	20
2.3	ESTRATÉGIAS DE BUSCA . . . . .	21
2.3.1	Strings de Busca: . . . . .	21
2.3.2	String de Busca em Inglês: . . . . .	21
2.4	PROCESSO DE SELEÇÃO DOS ESTUDOS . . . . .	22
2.4.1	Critérios de Inclusão: . . . . .	22
2.4.2	Critérios de Exclusão: . . . . .	23
2.5	EXTRAÇÃO E ANÁLISE DOS DADOS . . . . .	23
2.5.1	Perguntas de Pesquisa: . . . . .	23
2.6	RESULTADOS . . . . .	25
2.6.1	Descrição dos Estudos Utilizados . . . . .	25

2.6.2	Abordagens de Previsão de Disponibilidade de Recursos . . . . .	28
2.6.3	Limitações e Lacunas . . . . .	29
2.6.4	Tendências e Direções Futuras . . . . .	29
2.6.5	Análise dos Estudos . . . . .	30
2.6.6	Síntese dos Resultados . . . . .	30
2.7	CONCLUSÃO . . . . .	35
<b>3</b>	<b>METODOLOGIA</b>	<b>37</b>
3.1	COLETA DE DADOS . . . . .	39
3.2	ANÁLISE EXPLORATÓRIA DE DADOS . . . . .	42
3.3	AVALIAÇÃO DA CARTA DE CONTROLE I-AM . . . . .	45
3.4	DESENVOLVIMENTO DO MODELO DE PREVISÃO . . . . .	46
3.5	AVALIAÇÃO E COMPARAÇÃO DO MODELO . . . . .	48
3.6	CONSIDERAÇÕES FINAIS DA METODOLOGIA . . . . .	49



# 1 INTRODUÇÃO

A disponibilidade de recursos computacionais é fundamental para garantir a eficiência e a escalabilidade em sistemas distribuídos, Santos (2016) demonstrou que a alta variação no uso de recursos computacionais tem sido um obstáculo para a previsão precisa da disponibilidade desses recursos.

Os sistemas com arquitetura paralela e distribuída se tornaram cada vez mais populares e resolvem uma quantidade crescente de problemas de computação. Mas o problema de análise comportamental para tais sistemas, na sua grande maioria, compostos por máquinas das mais variadas arquiteturas continua a ser um fator que precisa ser aprimorado. Por tanto, Coulouris et al. (2013) destacam a importância do desenvolvimento de métodos de previsão de disponibilidade para sistemas compostos por máquinas de arquiteturas variadas.

Prever a disponibilidade de recurso é uma tarefa complexa e essa complexidade é ainda mais acentuada em sistemas distribuídos. Diversas são as metodologias e abordagens já propostas no intuito de predizer esta disponibilidade. Tanenbaum (2007) aponta que a heterogeneidade é um dos principais dificultadores do processo de predição de disponibilidade em sistemas distribuídos.

Isto pode ser constatado considerando que cada sistema pode ter características únicas que afetam a utilização dos recursos. Por exemplo, sistemas com arquiteturas diferentes podem ter taxas de utilização de CPU diferentes, o que pode afetar a precisão das previsões. Branco (2004) destaca ainda que a heterogeneidade dos sistemas distribuídos e a configuração de cada sistema podem afetar a utilização de recursos computacionais e, portanto, a precisão das previsões de disponibilidade.

Silva et al. (2020) apontaram que outro ponto dificultador da previsão se dá em razão do caráter estocástico envolvido. Isso se deve ao fato de que a utilização de recursos pode variar de forma imprevisível, dependendo de vários fatores, como a carga de trabalho e o uso de recursos por parte de outros sistemas. Isso pode tornar difícil prever a disponibilidade com precisão, especialmente em sistemas distribuídos onde a utilização de recursos pode variar de forma dinâmica.

Disponibilidade de CPU, uso de memória e de disco são destacados como os principais aspectos a serem monitorados. Paulino (2016) destaca que a disponibilidade de CPU é um aspecto importante a ser monitorado no mapeamento de disponibilidade de recursos computacionais. A partir de Pinto (2021), constata-se que uma série temporal de utilização de CPU é fortemente correlacionada, o que torna propício à utilização de métodos como média móvel, suavização exponencial e média móvel auto regressiva integrada para a tarefa de predição de disponibilidade de recursos. Outros estudos, como Oliveira, Barbar, and Soares (2015) que explora a aplicação de algoritmos de aprendizado de máquina e redes neurais para tratar problemas não lineares na predição de disponibilidade de recursos. A maioria desses preditores faz treinamento na série histórica antes de serem utilizados, baseando-se na premissa de que o futuro repetirá os padrões dos dados passados.

Silva et al. (2020) destacaram que as altas variações no uso da CPU criam novos padrões a cada instante, o que dificulta a adaptação de modelos que se baseiam em conjuntos fixos de dados a novos conjuntos. Seguindo uma linha diferente dos trabalhos anteriores, neste estudo foi utilizado o método de carta de controle ou gráfico de controle, ou ainda carta de controle remoto de qualidade, que é uma categoria de carta cujo objetivo é acompanhar, por exemplo, um determinado processo em uma linha de produção e é considerado uma das sete ferramentas de qualidade, grupo ao qual também estão inseridos: fluxograma, diagrama de causa e efeito, folhas de verificação, histogramas, cartas de dispersão e diagrama de Pareto. Essa ferramenta foi escolhida com base na revisão bibliográfica realizada, que é apresentada na seção de estado da arte.

Uma carta de controle I-AM é usada para monitorar a média e a variação do seu processo quando se tem dados contínuos que são observações individuais e não em subgrupos. Ferreira et al. (2018) afirma que utiliza-se esta carta de controle para monitorar a estabilidade do processo ao longo do tempo para que seja possível identificar e corrigir as instabilidades em um processo.

Segundo Minitab (2020), para interpretar uma carta I-AM, deve-se examinar a carta de valores individuais (carta I) e a carta de amplitude móvel (carta MR) para determinar se a média e a variação do processo estão sob controle.

Para a utilização desta abordagem, serão necessários vários testes a serem realizados para o acompanhamento de disponibilidade de diversas máquinas de usuários comuns e, com isto, demonstrar a viabilidade da utilização do método, apurando o nível de acurácia e custo computacional envolvido.

A proposta com esta abordagem, busca subsidiar o gerenciamento de recursos e escalonadores utilizando a Carta de Controle I-AM. Dessa forma, espera-se que, prever a disponibilidade futura de recursos seja mais preciso do que usar abordagens tradicionais baseadas em médias.

Desse modo, ao final deste trabalho espera-se ter desenvolvido um padrão que possa ser utilizado para monitorar de forma contínua e precisa a disponibilidade de CPU em sistemas distribuídos, proporcionando um maior gerenciamento de recursos computacionais e uma utilização mais eficiente desses recursos. Além disso, espera-se ter comparado o desempenho do padrão proposto com outras abordagens de previsão de disponibilidade de CPU e contribuir para a literatura científica sobre este tema.

## 1.1 MOTIVAÇÃO

Com a utilização de sistemas distribuídos, surge a necessidade de gerenciar os recursos computacionais compartilhados por esses sistemas, o que pode se tornar um desafio para garantir o gerenciamento dessas máquinas (Tanenbaum (2007)).

A motivação principal para o uso de sistemas distribuídos é somar o poder computacional de vários equipamentos, estando esses equipamentos próximos ou não, localizados de grandes instituições a domicílios. A soma desse poder culmina na criação de supercomputadores virtuais, o que torna possível a execução de uma grande quantidade de tarefas em um tempo reduzido. Existem empresas que comercializam esse poder computacional por meio de serviços em nuvem e há instituições que desenvolvem esses sistemas sem fins lucrativos, tendo como foco principal a contribuição para a ciência. A principal vantagem de tais sistemas é a possibilidade da utilização de alto poder computacional para a execução de tarefas em um espaço de tempo reduzido e minimização dos custos envolvidos (Varghese and Buyya (2018)).

Em ambos os casos, e principalmente nos que envolvem computadores localizados em domicílios

para a formação desse sistema, existe a grande dificuldade que é gerenciar os recursos computacionais possíveis de serem compartilhados.

É importante notar que, apesar das vantagens inerentes a estes sistemas, a administração dos mesmos pode ser um desafio, pois é preciso garantir que os recursos compartilhados estejam sendo utilizados de forma eficiente e que o sistema como um todo esteja funcionando corretamente. Isso pode envolver a utilização de técnicas de gerenciamento de recursos, bem como a solução de problemas de comunicação entre os equipamentos.

Visando mitigar a dificuldade de gerenciamento de tais recursos, essa pesquisa busca demonstrar um método que foca na previsão de disponibilidade de CPU, buscando dar subsídio aos escalonadores de tarefas. Muitos são os trabalhos focados nesses escalonadores e nos parâmetros de entrada dos mesmos, e ao encontro desses esforços está sendo apresentado esse para que possa ser um subsídio a mais.

A dificuldade de gerenciamento de recursos computacionais em sistemas distribuídos pode afetar a eficiência e o desempenho desses sistemas de várias maneiras. Por exemplo, se os recursos computacionais não forem distribuídos de forma equilibrada entre as máquinas do sistema, algumas podem ficar sobrecarregadas enquanto outras ficam ociosas. Isso pode levar a um desperdício de recursos e a uma diminuição da eficiência do sistema. Além disso, se as tarefas não forem escalonadas de forma adequada, pode ocorrer um atraso na execução das tarefas, uma diminuição no desempenho do sistema ou em casos extremos, escalonamentos indevidos. Para tanto, é necessário utilizar técnicas de gerenciamento de recursos que permitam alocar as tarefas de forma eficiente entre os equipamentos e garantir que todas as tarefas sejam concluídas de forma otimizada. De acordo com Sharma (2009), alguns exemplos de técnicas de gerenciamento de recursos incluem:

- Escalonamento de tarefas: é o processo de alocar tarefas para serem executadas em um ou mais equipamentos. Existem vários algoritmos de escalonamento, cada um com suas próprias características e aplicações;
- Alocação de recursos: é o processo de atribuir recursos computacionais, como CPU, memória e disco, para as tarefas de acordo com suas necessidades. A alocação de

recursos deve ser feita de forma a maximizar a eficiência do sistema e minimizar o tempo de espera das tarefas;

- Balanceamento de carga: é o processo de distribuir a carga de trabalho entre os equipamentos de forma a evitar sobrecarga em alguns deles e ociosidade em outros;
- Gerenciamento de falhas: é o processo de detectar e corrigir falhas em um sistema distribuído. Isso inclui a identificação de falhas em equipamentos, a remoção de tarefas que estejam sendo executadas nesses equipamentos e a reatribuição dessas tarefas para outros equipamentos.

A previsão de disponibilidade de CPU pode ser uma ferramenta útil para os escalonadores de tarefas, pois permite antecipar o uso futuro da CPU e tomar decisões informadas sobre a alocação de tarefas (Wen et al. (2020)). Isso pode ajudar a garantir que os recursos computacionais sejam utilizados de forma eficiente e que as tarefas sejam concluídas de forma otimizada.

## **1.2 OBJETIVO GERAL**

O objetivo geral desta pesquisa é desenvolver um modelo preciso e eficiente para monitorar de forma contínua a disponibilidade de recursos computacionais em um sistema distribuído, visando melhorar a qualidade do serviço oferecido aos usuários finais. Este objetivo está alinhado com os achados da revisão sistemática realizada, que destacou a importância e os desafios de técnicas eficientes de previsão de disponibilidade de recursos em sistemas distribuídos.

## **1.3 OBJETIVOS ESPECÍFICOS**

Os objetivos específicos deste trabalho são:

- Investigar a eficácia da carta de controle I-AM para monitorar a disponibilidade de recursos computacionais em sistemas distribuídos, conforme é sugerido pela revisão sistemática;
- Propor um modelo de previsão de disponibilidade de recursos computacionais que

seja aplicável em sistemas distribuídos, com alta precisão e confiabilidade, buscando aprimorar as técnicas identificadas na revisão;

- Determinar o intervalo ideal para a atualização dos parâmetros de disponibilidade, visando reduzir o tempo de resposta e garantir a eficiência do modelo, um ponto crítico levantado na revisão;
- Avaliar e comparar a eficácia do modelo proposto com abordagens existentes, complementando as análises já realizadas pela revisão sistemática;
- Realizar experimentos para medir o desempenho do modelo proposto em diferentes cenários, tais como variações de carga, falhas de equipamentos, entre outros;
- Contribuir para o avanço da literatura científica em previsão de disponibilidade de recursos computacionais em sistemas distribuídos, seguindo as recomendações e oportunidades levantadas na revisão.

## 1.4 HIPÓTESES

As hipóteses deste trabalho são:

- A utilização de ferramentas e métodos de controle estatístico de processo é viável para monitorar a disponibilidade de CPU de equipamentos pertencentes a um sistema distribuído;
- O padrão proposto terá um nível de assertividade alto ao prever o uso de CPU em sistemas distribuídos;
- O intervalo de tempo ideal para atualização dos parâmetros de disponibilidade dependerá das características do sistema distribuído e da carga de trabalho;
- O padrão proposto terá desempenho superior a outras abordagens de previsão de disponibilidade de CPU, como médias móveis, suavização exponencial e médias móveis auto regressivas integradas (ARIMA), algoritmos de aprendizado de máquina e redes neurais.

## 1.5 JUSTIFICATIVA

A garantia de disponibilidade de recursos computacionais é fundamental para garantir a eficiência e escalabilidade de sistemas distribuídos. A alta variação no uso de recursos computacionais tem sido um obstáculo para a previsão precisa da disponibilidade desses recursos (Santana (2006)).

Sistemas com arquitetura paralela e distribuída se tornaram cada vez mais populares e resolvem uma quantidade crescente de problemas de computação. No entanto, o problema de análise comportamental para tais sistemas, que são na sua grande maioria compostos por máquinas de arquiteturas variadas, continua a ser um fator que precisa ser aprimorado (Coulouris et al. (2013))

A previsão de disponibilidade de recurso é uma tarefa complexa e essa complexidade é ainda mais atenuada em sistemas distribuídos. A heterogeneidade dos sistemas distribuídos pode ser um fator dificultador da previsão de disponibilidade de recursos computacionais, pois cada sistema pode ter características únicas que afetam a utilização dos recursos. Além disso, o caráter estocástico envolvido na utilização de recursos computacionais em sistemas distribuídos pode tornar difícil a previsão precisa da disponibilidade desses recursos (Fontoura et al. (n.d.)) .

O monitoramento da disponibilidade de recursos computacionais é um aspecto importante para garantir a eficiência e o desempenho dos sistemas distribuídos. Disponibilidade de CPU, uso de memória e de disco são destacados como os principais aspectos a serem monitorados. A disponibilidade de CPU é especialmente importante porque é um recurso crítico para a execução de tarefas em sistemas distribuídos (Tesser (2011)).

Essas observações são de relevância para o desenvolvimento de uma abordagem de previsão de disponibilidade de CPU por meio da utilização de ferramentas de controle estatístico de processos para monitoramento de disponibilidade de recursos computacionais em sistemas distribuídos. A utilização dessas ferramentas busca permitir a identificação de padrões no uso dos recursos e auxiliar na tomada de decisões sobre como distribuir esses recursos de forma mais eficiente.

## 1.6 METODOLOGIA

A metodologia proposta para o presente estudo será baseada na utilização da ferramenta Carta de Controle I-AM para monitorar a disponibilidade de CPU em equipamentos pertencentes a sistemas distribuídos de maneira contínua. Para isso, serão coletados dados históricos de uso de CPU das máquinas envolvidas e analisados utilizando a Carta de Controle I-AM.

Em seguida, será desenvolvido um modelo que possa ser aplicado a sistemas distribuídos, garantindo um alto nível de assertividade nas previsões. O intervalo de tempo ideal para atualização dos parâmetros de disponibilidade será definido a partir da análise dos dados coletados. Além disso, será avaliado o nível de assertividade das previsões geradas pelo modelo proposto e o custo computacional envolvido no processo.

Existe também a proposta de se incluir uma etapa de otimização do padrão de disponibilidade de CPU, utilizando técnicas de otimização de hiperparâmetros. Isso permitirá ajustar os parâmetros do modelo de previsão de forma a maximizar sua precisão, garantindo que ele seja capaz de prever a disponibilidade de CPU de forma o mais precisa possível, sem gerar um excessivo custo computacional. Para validar o método proposto, serão realizados testes em um ambiente com dados reais da utilização de CPUs de sistemas distribuídos, comparando o desempenho do modelo otimizado com outras abordagens de previsão de disponibilidade, como médias móveis, suavização exponencial e médias móveis auto regressivas integradas (ARIMA). Espera-se que o resultado deste estudo possa contribuir para a literatura científica sobre métodos de previsão de disponibilidade de recursos computacionais em sistemas distribuídos e fornecer uma alternativa eficiente para o gerenciamento de recursos em sistemas distribuídos.

Além disso, seria importante a inclusão de uma avaliação do custo-benefício da utilização do modelo proposto em sistemas distribuídos. Isso pode ser feito considerando o nível de precisão das previsões geradas pelo padrão em relação ao custo computacional envolvido no processo de otimização e treinamento do modelo.

Também seria interessante avaliar o impacto da utilização do padrão de disponibilidade de CPU na eficiência e desempenho dos sistemas distribuídos. Isso pode ser feito comparando o tempo de execução de tarefas em sistemas que utilizam o modelo proposto com sistemas que



utilizam outras abordagens de previsão de disponibilidade de CPU, como médias móveis ou redes neurais.

Seria importante realizar uma análise de viabilidade econômica da utilização do modelo de disponibilidade de CPU em sistemas distribuídos, avaliando o potencial de retorno sobre o investimento em relação ao benefício obtido com a utilização do padrão. Isso pode incluir uma análise do custo de implementação e manutenção do padrão, bem como um cálculo do ganho de eficiência obtido com sua utilização. Além disso, considerar o impacto do padrão na satisfação dos usuários finais, avaliando como ele afeta a qualidade do serviço oferecido pelos sistemas distribuídos.

Para isso, seria necessário desenvolver uma interface para integrar o modelo de disponibilidade de CPU ao sistema de gerenciamento de recursos e realizar testes em diferentes cenários de uso, avaliando o impacto desta abordagem no desempenho e na eficiência do sistema de gerenciamento de recursos, comparando-o com sistemas que utilizam outras sistemáticas para a tomada de decisões sobre a distribuição de recursos.

Ao final da dissertação, espera-se ter desenvolvido um modelo de previsão de disponibilidade de CPU para sistemas distribuídos que seja eficiente e preciso, além de ter avaliado o seu impacto na eficiência e desempenho de sistemas distribuídos e o seu potencial de integração a um sistema de gerenciamento de recursos. Isso poderá contribuir para a literatura científica sobre métodos de previsão de disponibilidade de recursos computacionais em sistemas distribuídos e fornecer uma alternativa viável para o gerenciamento de recursos em sistemas distribuídos.

Por fim, o desempenho do padrão proposto será comparado com outras abordagens de previsão de disponibilidade de CPU, como médias móveis, suavização exponencial e médias móveis auto regressivas integradas (ARIMA), algoritmos de aprendizado de máquina e redes neurais. O objetivo é verificar se a utilização da carta de controle pode ser uma alternativa viável para o monitoramento de disponibilidade de recursos computacionais em sistemas distribuídos.

### **1.6.1 Plano de Trabalho**

O plano de trabalho inclui a coleta de dados de uso de CPU de diversas máquinas em sistemas distribuídos, a análise desses dados usando técnicas de carta de controle I-AM e a comparação

do desempenho do método proposto com outras abordagens de previsão de disponibilidade de CPU.

As etapas do plano de trabalho incluem:

- Coleta de dados de uso de CPU de diversas máquinas em sistemas distribuídos;
- Análise dos dados usando técnicas de carta de controle I-AM;
- Definição do intervalo de tempo ideal para atualização dos parâmetros de disponibilidade;
- Avaliação do nível de assertividade das previsões de uso de CPU geradas pelo método proposto, bem como o custo computacional envolvido no processo;
- Comparação do desempenho do método proposto com outras abordagens de previsão de disponibilidade de CPU, como médias móveis, suavização exponencial e médias móveis auto regressivas integradas (ARIMA), algoritmos de aprendizado de máquina e redes neurais;
- Redação e submissão do artigo científico descrevendo os resultados obtidos.

## 2 REVISÃO SISTEMÁTICA

A disponibilidade de recursos computacionais é um aspecto crucial em sistemas distribuídos, uma vez que afeta diretamente o desempenho, a confiabilidade e a eficiência desses sistemas. A capacidade de prever a disponibilidade desses recursos de forma precisa e confiável desempenha um papel fundamental no gerenciamento eficaz desses ambientes computacionais. A previsão da disponibilidade de recursos possibilita a tomada de decisões informadas sobre a alocação, escalabilidade e otimização dos recursos, além de permitir a detecção precoce de possíveis falhas ou gargalos.

Nesse contexto, a aplicação de técnicas de controle estatístico de processos tem sido amplamente explorada como uma abordagem promissora para o monitoramento de recursos dos mais diversos tipos. Carta de controle é uma ferramenta estatística que permite monitorar e controlar a variabilidade dos dados coletados ao longo do tempo, facilitando a identificação de desvios e anomalias nos recursos computacionais. O controle estatístico de processos, por sua vez, abrange uma série de técnicas e métodos estatísticos, incluindo-se as cartas de controle, para analisar e melhorar a estabilidade e o desempenho dos processos.

Diante desse cenário, esta revisão sistemática tem como objetivo identificar as abordagens mais eficazes e inovadoras para a previsão da disponibilidade de recursos computacionais em sistemas distribuídos, com foco no uso da carta de controle I-AM e técnicas de controle estatístico de processos. Além disso, pretende-se avaliar a qualidade e a confiabilidade dos estudos que investigaram o uso dessas abordagens, identificar as técnicas de controle estatístico de processos mais aplicadas e investigar as práticas e desafios atuais relacionados à previsão da disponibilidade de recursos computacionais nesse contexto.

Por meio dessa revisão sistemática, busca-se fornecer uma análise abrangente e atualizada sobre as abordagens de previsão da disponibilidade de recursos computacionais em sistemas distribuídos, bem como propor recomendações e diretrizes para o aprimoramento do gerenciamento de recursos nesses ambientes. Com base na análise crítica dos estudos selecionados, espera-se contribuir para o avanço do conhecimento nessa área e para o aprimoramento das práticas de previsão da disponibilidade de recursos computacionais.

## 2.1 DEFINIÇÃO DOS CRITÉRIOS DE BUSCA

Para realizar a busca dos estudos relevantes, foram definidos critérios específicos para garantir a inclusão de estudos pertinentes à temática da previsão da disponibilidade de recursos computacionais em sistemas distribuídos. Os critérios de busca incluíram palavras-chave relacionadas ao tema, bem como filtros de idioma e período de publicação. Aqui é definida uma lista de palavras-chave relevantes para o estudo, incluindo sinônimos e variações. Em seguida, são selecionadas as bases de dados que serão utilizadas para fazer a pesquisa bibliográfica.

### 2.1.1 Termos de Pesquisa:

- Disponibilidade de recursos computacionais
- Previsão de recursos
- Carta de controle I-AM (Individual com Amplitudes Móveis) e I-MR(Individual com Moving Range)
- Controle estatístico de processos
- Gerenciamento de recursos computacionais
- Análise de desempenho
- Sistemas distribuídos

### 2.1.2 Sinônimos e Variações:

- Recursos de computação
- Capacidade computacional
- Controle estatístico de qualidade
- Gráfico de controle I-AM (Individual com Amplitudes Móveis)
- Gerenciamento de desempenho
- Sistema distribuído em nuvem

- Cloud computing

### **2.1.3 Objetivos da Revisão Sistemática**

- Identificar as abordagens mais eficazes e inovadoras para previsão da disponibilidade de recursos computacionais em sistemas distribuídos;
- Avaliar a qualidade e a confiabilidade dos estudos que investigaram o uso da carta de controle I-AM para previsão da disponibilidade de recursos computacionais em sistemas distribuídos;
- Identificar as técnicas de controle estatístico de processos mais aplicadas em sistemas distribuídos;
- Investigar as práticas e desafios atuais relacionados à previsão da disponibilidade de recursos computacionais em sistemas distribuídos;
- Propor recomendações e diretrizes para a melhoria do gerenciamento de recursos computacionais em sistemas distribuídos por meio da previsão da disponibilidade de recursos com o uso da carta de controle I-AM e técnicas de controle estatístico de processos.

### **2.1.4 Perguntas de Pesquisa**

- Quais são os métodos mais utilizados para previsão da disponibilidade de recursos?
- Quais são os principais modelos estatísticos de processos aplicados?

## **2.2 FONTES DE INFORMAÇÃO**

Foram selecionadas diversas bases de dados acadêmicas e industriais para realizar a busca dos estudos relevantes. As fontes de informação incluem a IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Scopus, Web of Science e Wiley. Além disso, foram consultadas conferências e periódicos relevantes na área de sistemas distribuídos e computação.

## 2.3 ESTRATÉGIAS DE BUSCA

A estratégia de busca envolveu a combinação de termos de busca relacionados à previsão da disponibilidade de recursos computacionais, sistemas distribuídos, computação em nuvem e técnicas de controle estatístico de processos. As palavras-chave foram adaptadas conforme a sintaxe e os recursos de cada base de dados.

### 2.3.1 Strings de Busca:

- “Disponibilidade de recursos computacionais” AND “Carta de controle I-AM”
- “Previsão de recursos” AND “Controle estatístico de processos”
- “Gerenciamento de recursos computacionais” AND “Análise de desempenho”
- “Sistemas distribuídos” AND “Carta de controle I-AM”
- “Computação em nuvem” AND “Controle estatístico de processos”
- (“Disponibilidade de recursos computacionais” OR “Recursos de computação” OR “Capacidade computacional”) AND (“Carta de controle I-AM” OR “Gráfico de controle I-AM” OR “Controle estatístico de qualidade”)
- (“Previsão de recursos” OR “Gerenciamento de recursos computacionais” OR “Gerenciamento de desempenho”) AND (“Controle estatístico de processos” OR “Carta de controle I-AM” OR “Gráfico de controle I-AM”)
- (“Análise de desempenho” OR “Sistemas distribuídos” OR “Sistema distribuído em nuvem” OR “Cloud computing”) AND (“Carta de controle I-AM” OR “Gráfico de controle I-AM” OR “Controle estatístico de qualidade”)

### 2.3.2 String de Busca em Inglês:

- “Availability of computational resources” AND “I-MR control chart”
- “Resource Forecasting” AND “Statistical Process Control”
- “Computing resource management” AND “Performance analysis”

- “Distributed Systems” AND “I-MR Control Chart”
- “Cloud Computing” AND “Statistical Process Control”
- (“Availability of Computing Resources” OR “Computing Resources” OR “Computing Capacity”) AND (“I-MR Control Chart” OR “I-MR Control Chart” OR “Statistical Quality Control”)
- (“Resource Forecasting” OR “Computing Resource Management” OR “Performance Management”) AND (“Statistical Process Control” OR “I-MR Control Chart” OR “I-MR Control Chart”)
- (“Performance analysis” OR “Distributed systems” OR “Distributed cloud system” OR “Cloud computing”) AND (“I-MR control chart” OR “I-MR control chart” OR “Statistical quality control” )

## 2.4 PROCESSO DE SELEÇÃO DOS ESTUDOS

Após a realização da busca inicial, os estudos foram selecionados em duas etapas: triagem com base nos títulos e resumos, seguida pela análise completa dos artigos selecionados. Durante a triagem, os critérios de inclusão foram aplicados para identificar os estudos potencialmente relevantes. Na análise completa, os estudos foram avaliados em relação aos critérios de inclusão e exclusão para garantir a seleção dos estudos mais adequados ao escopo da revisão.

### 2.4.1 Critérios de Inclusão:

- Tema: O trabalho deve abordar o tema de monitoramento ou previsão de recursos computacionais.
- Método: O trabalho deve utilizar a carta de controle I-MR ou alguma variação de carta de controle ou alguma ferramenta ou método de controle estatístico de processo.;
- Ano: Artigos publicados entre 2018 e 2023;
- Idioma: Artigos publicados em português ou inglês;

### 2.4.2 Critérios de Exclusão:

- Tema: Estudos que não abordem a temática central da pesquisa;
- Disponibilidade: o trabalho deve estar disponível integralmente;
- Duplicidade: Artigos duplicados.

Os estudos selecionados foram submetidos a uma avaliação da qualidade e confiabilidade. Foram considerados critérios como a rigorosidade metodológica, a validade dos resultados, a clareza na apresentação dos métodos utilizados e a contribuição para o campo de estudo. Estudos de alta qualidade foram priorizados para análise e discussão na revisão sistemática.

## 2.5 EXTRAÇÃO E ANÁLISE DOS DADOS

Após a seleção dos estudos, os dados relevantes foram extraídos e organizados em uma planilha para permitir uma análise sistemática. As informações extraídas incluíram detalhes sobre os estudos (autor, ano da publicação, contexto de aplicação, métodos utilizados), as abordagens de previsão de disponibilidade de recursos, as técnicas de controle estatístico de processos empregadas, as métricas avaliadas e os principais resultados obtidos.

### 2.5.1 Perguntas de Pesquisa:

- Quais são os métodos mais utilizados para previsão da disponibilidade de recursos?
  - Aprendizado de máquina: é uma subárea da inteligência artificial que utiliza algoritmos e técnicas para permitir que os computadores aprendam a partir de dados e realizem tarefas sem serem explicitamente programados. O aprendizado de máquina abrange diversos métodos, incluindo algoritmos de aprendizado supervisionado, não supervisionado e por reforço. Alguns dos estudos relevantes que utilizam ou propõem modelos de aprendizado de máquina para prever a disponibilidade de recursos incluem [A02], [A09], [A13], [A15], [A25] e [A31].
  - Mecanismos de leilão: são modelos baseados em teoria dos jogos e microeconomia que permitem a alocação eficiente de recursos escassos entre agentes racionais que competem por eles. Alguns dos artigos que usam ou propõem mecanismos de leilão



para previsão da disponibilidade de recursos são [A3], [A04] [A05], [A06], [A08], [A11], e [A12].

- Modelos estatísticos: são uma representação matemática de um fenômeno real que usa conceitos e métodos estatísticos para descrever e explicar a relação entre variáveis. Esses modelos são baseados em dados observados e tentam capturar as características essenciais do processo subjacente ao fenômeno estudado. Alguns dos artigos que usam ou propõem esses modelos para previsão da disponibilidade de recursos são [A14], [A16]. [A19], [A20], [A22], [A23], [A24], [A26], [A27], [A29], [A30], [A32] e [A33].
- Quais são os principais modelos estatísticos de processos aplicados?
  - Programação linear: envolve maximizar ou minimizar uma função linear sujeita a restrições lineares, sendo amplamente aplicada na otimização de recursos e tomada de decisões eficientes.. Alguns dos artigos que usam ou propõem Programação linear para previsão da disponibilidade de recursos são [24] e [26].
  - Cartas de controle: são ferramentas que permitem monitorar a qualidade de um processo por meio da análise de variáveis que indicam o seu desempenho. Eles consistem em gráficos que mostram os limites superior e inferior de controle, que representam os valores aceitáveis para a variável monitorada, e os pontos que correspondem às observações da variável ao longo do tempo. Alguns dos artigos que usam ou propõem gráficos de controle para previsão da disponibilidade de recursos são [A14], [A16], [A19], [A20], [A27], [A32] e [A33].
  - Testes de hipóteses: são métodos que permitem testar se uma afirmação sobre uma população é verdadeira ou falsa, com base em uma amostra. Eles envolvem a formulação de duas hipóteses opostas, a nula e a alternativa, e a escolha de um nível de significância que determina a probabilidade máxima de rejeitar a hipótese nula quando ela é verdadeira. Alguns dos artigos que usam ou propõem testes de hipóteses para previsão da disponibilidade de recursos é [A18].

## 2.6 RESULTADOS

Nesta seção, são apresentados os resultados obtidos por meio da pesquisa sobre sistemas distribuídos e técnicas de gerenciamento de recursos. São descritos os principais achados e contribuições do estudo, são destacadas as descobertas mais relevantes e os insights obtidos. Os resultados são apresentados de acordo com os objetivos estabelecidos na seção de metodologia.

### 2.6.1 Descrição dos Estudos Utilizados

Inicialmente, são descritos os estudos utilizados na revisão, incluindo sua origem, características e escala. Também são discutidos os critérios de seleção, levando em consideração a representatividade e a relevância para os objetivos da pesquisa.

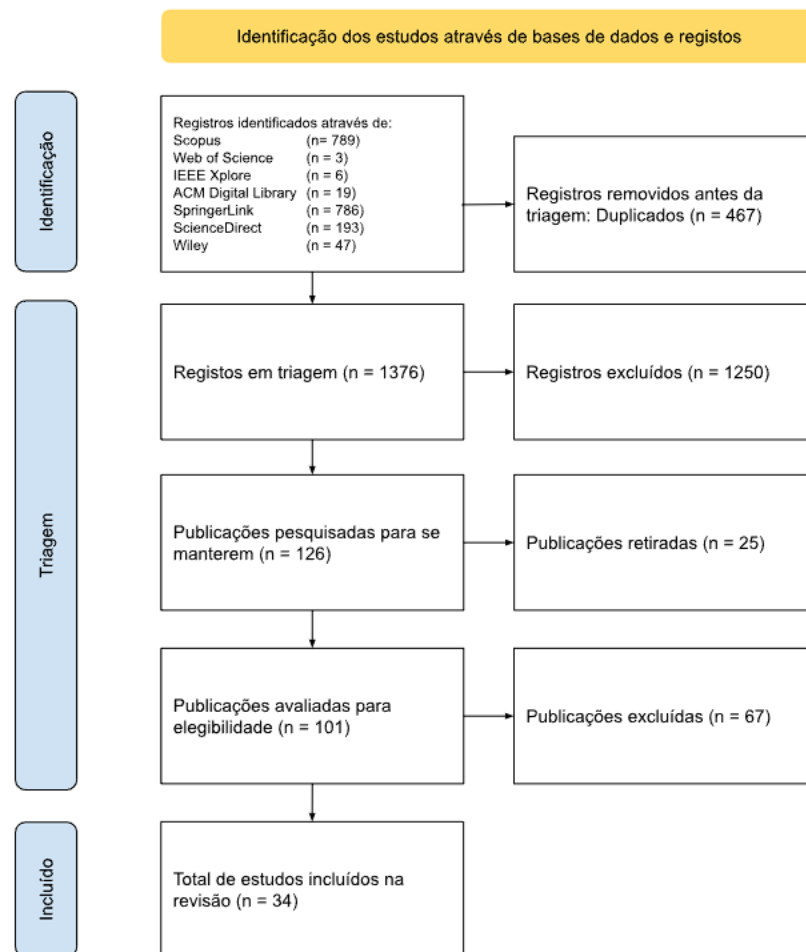


Figura 1: Fluxograma do processo de seleção dos estudos incluídos na revisão sistemática.<sup>5</sup>

<sup>5</sup>Fonte: Adaptação do autor, 2023. <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram.aspx>

Foram selecionados um total de 34 estudos primários para a revisão sistemática, publicados entre os anos 2018 e 2023. Os estudos foram identificados a partir de fontes de dados como bibliotecas digitais, bases de dados científicas e conferências relevantes. Os critérios de inclusão foram estabelecidos para garantir a relevância dos estudos aos objetivos da revisão.

Os estudos incluídos abordaram uma variedade de sistemas distribuídos, como nuvens computacionais, redes de sensores sem fio e ambientes de computação em grade. A maioria dos estudos estava focada em prever a disponibilidade de recursos como CPU, memória e largura de banda.

Identificador	Título	Autores	Ano de publicação
A01	Classification of resource management approaches in fog/edge paradigm and future research prospects: a systematic review	Kansal, P. and Kumar, M. and Verma, O.P.	2022
A02	Machine and Deep Learning for Resource Allocation in Multi-Access Edge Computing: A Survey	Djigal, H. and Xu, J. and Liu, L. and Zhang, Y.	2022
A03	Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks	Jiao, Y. and Wang, P. and Niyato, D. and Suankaewmanee, K.	2019
A04	Edge Network Resource Synergy for Mobile Blockchain in Smart City	Guo, S. and Qi, Y. and Yu, P. and Shao, S. and Qiu, X.	2020
A05	A dynamic resource allocation algorithm based on auction model in mobile blockchain network	Gao, Z. and Lin, B. and Xiao, K. and Wang, Q. and Mo, Z. and Yang, Y.	2019
A06	Hierarchical Combinatorial Auction in Computing Resource Allocation for Mobile Blockchain	Xu, Y. and Zhu, K. and Li, S.	2020
A07	Robust Task Offloading in Dynamic Edge Computing	Wang, H. and Xu, H. and Huang, H. and Chen, M. and Chen, S.	2023
A08	Three-stage Stackelberg game based edge computing resource management for mobile blockchain	Fan, Y. and Jin, Z. and Shen, G. and Hu, D. and Shi, L. and Yuan, X.	2021
A09	Self-learning and self-adaptive resource allocation for cloud-based software services	Chen, X. and Lin, J. and Lin, B. and Xiang, T. and Zhang, Y. and Huang, G.	2019
A10	Graph-Based Computing Resource Allocation for Mobile Blockchain	Abdellatif, K. and Abdelmouttalib, C. and EL-Azouzi R. and Sadik M. and Sabir E. and Tmirt A. and Altman E.	2019
A11	Optimal pricing-based edge computing resource management in mobile blockchain	Xiong, Z. and Feng, S. and Niyato, D. and Wang, P. and Han, Z.	2018
A12	Social Welfare Maximization Auction in Edge Computing Resource Allocation for Mobile Blockchain	Jiao, Y. and Wang, P. and Niyato, D. and Xiong, Z.	2018
A13	Edge-Computing and Machine-Learning-Based Framework for Software Sensor Development	Hanzelik, P.P. and Kummer, A. and Abonyi, J.	2022
A14	Monitoring Runtime Metrics of Fog Manufacturing via a Qualitative and Quantitative (QQ) Control Chart	Li, Y. and Wang, L. and Lee, D. and Jin, R.	2022
A15	Adaptive anomaly detection in performance metric streams	Ibidunmoye, O. and Rezaie, A.-R. and Elnroth, E.	2018
A16	Monitoring processes with multiple dependent production lines using time between events control charts	Ahmad, H. and Ahmadi Nadi, A. and Amini, M. and Sadeghpour Gildeh, B.	2023
A17	Gradient analysis of Markov-type control schemes and its applications	Yashchin, E.	2022
A18	Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions	Belay, M.A. and Blakseth, S.S. and Rasheed, A. and Salvo Rossi, P.	2023
A19	Process capability monitoring and change-point analysis for S-type quality characteristic	Liao, M.-Y. and Wu, C.-W.	2023
A20	Nonparametric Control Charts for Monitoring Serial Dependence based on Ordinal Patterns	Wei, C.H. and Testik, M.C.	2023
A21	On challenges of cloud monitoring	Fournajidi, W. and Erwin, T. and Steinbacher, J. and Miransky, A. and Featherstone J. and Lyons K. and Wigglesworth J. and Saraf D.	2020
A22	Adaptive Markov-based approach for dynamic virtual machine consolidation in cloud data centers with quality-of-service constraints	Monshizadeh Naeen, H. and Zeinali, E. and Toroghi Haghighat, A.	2020
A23	A stochastic process-based server consolidation approach for dynamic workloads in cloud data centers	Monshizadeh Naeen, H. and Zeinali, E. and Toroghi Haghighat, A.	2020
A24	A framework for tunable anomaly detection	Alam, M.R. and Gerostathopoulos, I. and Prehofer, C. and Attanasi, A. and Bures, T.	2019
A25	Optimal Resource Estimation Policy Selection for Ecommerce Applications in Cloud	P. A. Bommanavar and R. Krishnan and S. Shadedha	2019
A26	Robust Service Provisioning With Service Function Chain Requirements in Mobile Edge Computing	J. Li and W. Liang and Y. Ma	2021
A27	A Monitoring System for Anomaly Detection in Fog Manufacturing	L. Wang and Y. Zhang and R. Jin	2020
A28	Efficient Computing Resource Sharing for Mobile Edge-Cloud Computing Networks	Mishra, Shashank and Gupta, Aman and Jairam Naik, K. and Mandal, Jyotsna Kumar and De, Debashis	2023
A29	Abnormal behavior detection using resource level to service level metrics mapping in virtualized systems	Benmakrelouf, Souhila and St-Onge, Cédric and Kara, Nadja and Tout, Hanine and Edstrom, Claes and Lemieux, Yves	2020
A30	Resource needs prediction in virtualized systems: Generic proactive and self-adaptive solution	Benmakrelouf, Souhila and Kara, Nadja and Tout, Hanine and Rabipour, Rafi and Edstrom, Claes	2019
A31	RTSLPS: Real time server load prediction system for the ever-changing cloud computing environment	Toumi, Hajer and Brahmi, Zaki and Gammoudi, Mohammed Mohsen	2022
A32	On designing an optimal SPRT control chart with estimated process parameters under guaranteed in-control performance	Teoh, J.W. and Teoh, W.L. and Khoo, Michael B.C. and Castagliola, Philippe and Moy, W.H.	2022
A33	Optimal estimated process parameters side sensitive group runs chart based on expected average run length	You, Huay Woon	2018
A34	Stability control in virtual machine: Resource allocation for cloud computing	Ismail, M.B., Hashim, H., Yusof, Y.M.	2019

Tabela 1: Estudos incluídos, incluindo o título do estudo, autores, ano de publicação.<sup>6</sup>

<sup>5</sup>Fonte: Autoria Própria, 2023.

<sup>6</sup>Fonte: Autoria Própria, 2023.

### 2.6.2 Abordagens de Previsão de Disponibilidade de Recursos

Após a análise dos estudos incluídos, foram identificadas diversas abordagens utilizadas para prever a disponibilidade de recursos computacionais em sistemas distribuídos. As abordagens foram agrupadas em categorias com base nas técnicas e algoritmos empregados.

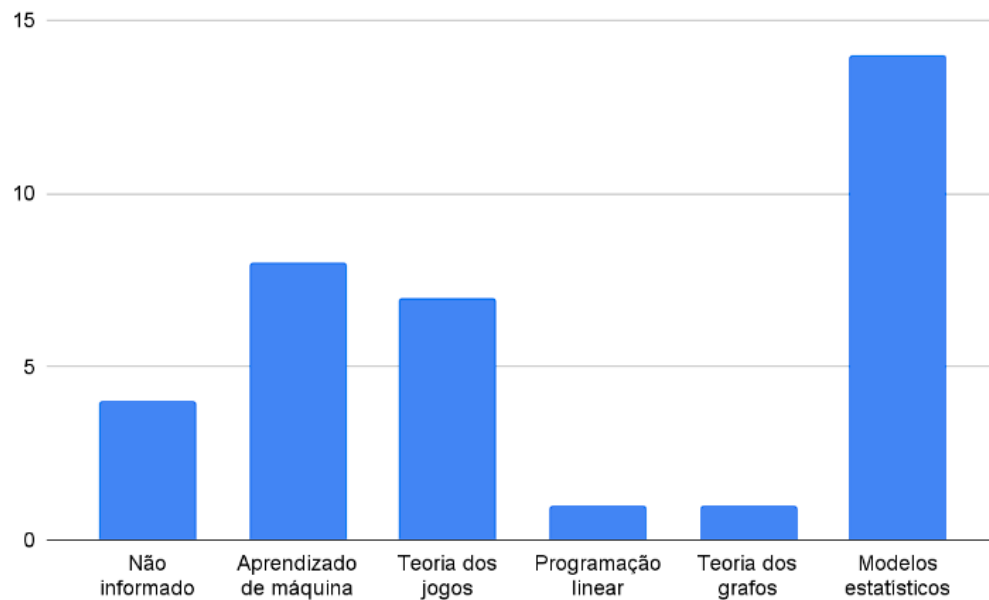


Figura 2: Categorização das abordagens de previsão de disponibilidade de recursos, mostrando as técnicas, algoritmos ou modelos utilizados em cada categoria. <sup>8</sup>

As categorias identificadas foram: modelos estatísticos, aprendizado de máquina, teoria dos jogos, teoria dos grafos e programação linear. Cada categoria foi caracterizada por suas principais características e métodos utilizados para a previsão da disponibilidade de recursos.

O gráfico 2 mostra as técnicas, algoritmos ou modelos utilizados em cada categoria. Enquanto o gráfico 3 expõe quais os tipos de sistemas distribuídos foram aplicados às técnicas de previsão de disponibilidade. Categorização das abordagens de previsão de disponibilidade de recursos, mostrando as técnicas, algoritmos ou modelos utilizados em cada categoria.

---

<sup>8</sup>Fonte: Autoria Própria, 2023.

<sup>9</sup>Fonte: Autoria Própria, 2023.

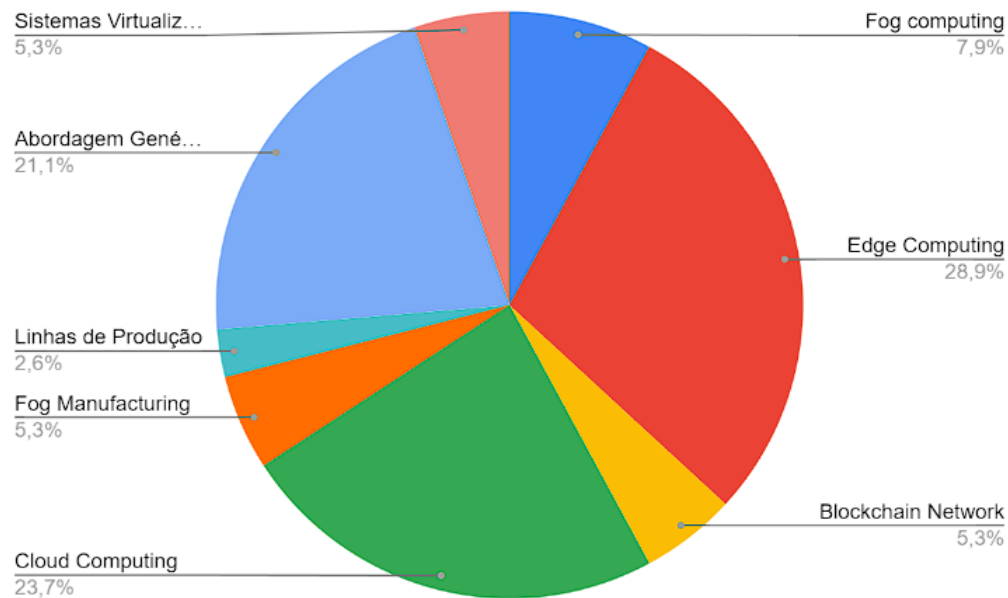


Figura 3: Sistema distribuído abordado. <sup>9</sup>

### 2.6.3 Limitações e Lacunas

Algumas das limitações identificadas são: falta de dados de disponibilidade de recursos em cenários realistas ou dependência de conjuntos de dados específicos. Além disso, foi identificado um número limitado de estudos que investigaram a previsão da disponibilidade de recursos em sistemas distribuídos de grande escala ou em ambientes dinâmicos.

Outra lacuna identificada foi a falta de padronização nas métricas de avaliação utilizadas nos estudos, o que dificultou a comparação direta entre as abordagens.

### 2.6.4 Tendências e Direções Futuras

Com base nos resultados encontrados e nas lacunas identificadas, há várias tendências e direções futuras promissoras para a pesquisa em previsão de disponibilidade de recursos em sistemas distribuídos. Algumas sugestões incluem: explorar abordagens híbridas que combinam modelos de séries temporais com técnicas de aprendizado de máquina, investigar a aplicação de algoritmos genéticos em cenários dinâmicos e ampliar os estudos para incluir sistemas distribuídos de grande escala, como infraestruturas de nuvem.

### 2.6.5 Análise dos Estudos

Os estudos incluídos nesta revisão sistemática foram descritos e avaliados quanto à qualidade metodológica. As análises estatísticas apresentadas foram comparadas e os resultados foram sintetizados para concluir sobre o que a literatura informa em relação à intervenção.

Nesta seção, é realizada uma análise crítica dos estudos existentes na área de sistemas distribuídos e técnicas de gerenciamento de recursos. O objetivo é fornecer uma visão panorâmica das pesquisas relevantes e identificar lacunas e oportunidades para a pesquisa.

### 2.6.6 Síntese dos Resultados

Os resultados dos estudos foram sintetizados e analisados com base nos objetivos da revisão sistemática. As principais descobertas foram agrupadas e organizadas de acordo com os tópicos relevantes para responder às perguntas de pesquisa. A análise dos resultados levou em consideração a efetividade das abordagens, as técnicas de controle estatístico de processos mais aplicadas, os desafios identificados e as tendências atuais na previsão. Aqui é apresentada a síntese narrativa dos artigos selecionados organizada pelos seguintes temas:

- T1: Métodos baseados em leilão
- T2: Métodos baseados em aprendizado de máquina
- T3: Métodos baseados em modelos estatístico

**2.6.6.1 T1: Métodos baseados em leilão** Os métodos baseados em leilão são aqueles que usam mecanismos de mercado para alocar recursos computacionais entre os provedores e os consumidores de serviços em ambientes de computação em nuvem e névoa. Nesse contexto, os recursos computacionais podem incluir capacidade de processamento, armazenamento, largura de banda de rede ou qualquer outro tipo de recurso necessário para executar tarefas computacionais. Alguns exemplos de artigos que usam esse tipo de método são:

- Em [A03] é proposto um mecanismo de leilão para a alocação de recursos de computação em nuvem/nevoeiro para redes blockchain públicas. Os autores consideram dois cenários: um em que os usuários finais podem escolher entre a nuvem e o nevoeiro, e outro em que os provedores de nevoeiro podem subcontratar recursos da nuvem. Eles formulam

os problemas de alocação de recursos como jogos de leilão e analisam as propriedades dos equilíbrios de Nash e de Bayes-Nash. Eles também propõem algoritmos distribuídos para encontrar os equilíbrios e avaliam o desempenho dos mecanismos de leilão por meio de simulações. Eles mostram que os mecanismos de leilão podem melhorar a eficiência, a justiça e a utilidade dos participantes em comparação com outros esquemas de alocação.

- Em [A12] é proposto um mecanismo de leilão de maximização de bem-estar social para a alocação de recursos de computação de borda para o blockchain móvel. Os autores consideram um cenário em que os usuários móveis podem comprar recursos de computação de borda para executar suas tarefas de mineração de blockchain. Eles modelam o problema como um jogo de leilão com externalidades alocativas, ou seja, o efeito que a alocação de recursos tem sobre a utilidade dos outros usuários. Eles mostram que o mecanismo de leilão proposto é verdadeiro, individualmente racional e computacionalmente eficiente. Eles também definem uma função de poder de hash que caracteriza a probabilidade de minerar com sucesso um bloco. Eles avaliam o desempenho do mecanismo de leilão por meio de simulações e mostram que ele pode melhorar o bem-estar social, a utilidade dos usuários e a receita do provedor de computação de borda.
- Em [A05] é proposto um algoritmo de alocação dinâmica de recursos baseado em um modelo de leilão na rede blockchain móvel. Os autores consideram um cenário em que os usuários móveis podem comprar recursos de computação de borda para executar suas tarefas de mineração de blockchain. Eles modelam o problema como um jogo de leilão baseado em Vickrey, onde os usuários enviam seus lances para os provedores de borda e o provedor escolhe o lance mais alto. Eles propõem um algoritmo distribuído para encontrar o equilíbrio de Nash do jogo e ajustar os lances dos usuários de acordo com a demanda e a oferta dos recursos. Eles avaliam o desempenho do algoritmo por meio de simulações e mostram que ele pode melhorar a utilidade dos usuários, a receita do provedor e a eficiência da alocação de recursos.

Os principais critérios e indicadores utilizados para avaliar esses métodos são:

- Bem-estar social: a soma dos utilitários ou lucros dos participantes do leilão;



- Eficiência alocativa: a razão entre o bem-estar social obtido pelo leilão e o bem-estar social ótimo;
- Justiça: a distribuição equitativa dos utilitários ou lucros entre os participantes do leilão;
- Segurança: a propriedade que garante que o leilão é resistente a ataques maliciosos ou fraudes.

Os principais desafios e oportunidades para melhorar esses métodos são:

- Os desafios são: Falta de padronização e interoperabilidade entre diferentes plataformas de blockchain móvel, a escalabilidade e o alto consumo de energia.
- Já as oportunidades incluem a descentralização e transparência, a redução de custos e intermediários, a eficiência e a segurança.

**2.6.6.2 T2: Métodos baseados em aprendizado de máquina** Os métodos baseados em aprendizado de máquina são aqueles que usam técnicas de inteligência artificial para aprender padrões, prever demandas, otimizar decisões e adaptar-se às mudanças nos ambientes de computação em nuvem e névoa. Esses métodos visam melhorar o desempenho, a eficiência e a qualidade dos serviços prestados. Alguns exemplos de artigos que usam esse tipo de método são:

- Em [A09] são discutidos os desafios enfrentados pelos engenheiros de nuvem ao alocar recursos para serviços de software baseados em nuvem. O artigo propõe uma abordagem de alocação de recursos de auto aprendizagem e auto adaptativa que pode alocar recursos apropriados para garantir boa qualidade de serviços, bem como baixo custo de recursos.
- Em [A25] é proposto método para estimar os recursos necessários para executar aplicações de comércio eletrônico na nuvem e selecionar a melhor política de alocação de recursos entre várias opções. O objetivo do método é minimizar o custo total de operação (TCO) da aplicação, considerando os custos de infraestrutura, manutenção e penalidades por violação de acordos de nível de serviço (SLA). O método usa um modelo analítico baseado em teoria das filas para estimar os recursos necessários para atender a uma

determinada demanda de usuários e um algoritmo genético para encontrar a política ótima de alocação de recursos entre diferentes provedores de nuvem. O artigo apresenta o método e sua validação em um cenário de simulação usando dados reais de uma aplicação de comércio eletrônico.

- Em [A31] é proposto um algoritmo para a detecção dinâmica e em tempo real de variações anormais e mapeamento de métricas de nível de serviço para métricas de nível de recurso em sistemas virtualizados. O algoritmo combina técnicas de divergência e detecção de outliers para uma análise de carga de trabalho em tempo real e uma detecção de comportamento anormal do sistema, bem como uma detecção de periodicidade. O algoritmo é capaz de detectar e mapear mudanças incomuns em ambas as métricas de nível de serviço e de recurso (por exemplo, vazão e uso de CPU) em um sistema virtualizado, sem nenhum conhecimento prévio ou suposições sobre seu comportamento interno. O algoritmo também fornece uma análise efetiva do estado do sistema e gera notificações ou alertas para adaptação de recursos. Os autores avaliam o algoritmo usando vários tipos de cargas de trabalho e mostram que ele pode detectar variações anormais com uma precisão que varia de 92% a 100%.

Os principais critérios e indicadores utilizados para avaliar esses métodos são:

- Desempenho: a capacidade dos métodos de atingir os objetivos desejados, como precisão, velocidade, confiabilidade ou segurança;
- Eficiência: a razão entre os benefícios obtidos pelos métodos e os custos incorridos pelos mesmos, como consumo energético, uso da rede ou complexidade computacional;
- Qualidade: a satisfação dos usuários finais com os serviços prestados pelos métodos, como disponibilidade, latência ou experiência do usuário.

Os principais desafios e oportunidades para melhorar esses métodos são:

- Lidar com a grande quantidade, variedade e veracidade dos dados gerados pelos ambientes de computação em nuvem e névoa;
- Projetar técnicas de aprendizado de máquina que sejam adaptáveis, robustas e transpa-

rentes aos cenários dinâmicos e incertos;

- Considerar aspectos éticos, legais e sociais na aplicação do aprendizado de máquina, como privacidade, segurança ou responsabilidade.

**2.6.6.3 T3: Métodos baseados em métodos estatístico** Os métodos baseados em otimização e controle estatístico são aqueles que usam técnicas estatísticas ou computacionais para coletar, analisar e identificar comportamentos anormais nos ambientes de computação em nuvem e névoa. Esses métodos visam garantir a qualidade, a confiabilidade e a segurança dos serviços prestados. Alguns exemplos de artigos que usam esse tipo de método são:

- [A20] apresenta uma solução para o problema de monitoramento da capacidade do processo e análise de mudanças em características de qualidade do tipo S. O artigo propõe um algoritmo de monitoramento baseado em gráficos de controle de capacidade do processo para monitorar a capacidade do processo em tempo real. O algoritmo é avaliado por meio de simulações e os resultados mostram que ele é capaz de melhorar significativamente o desempenho do sistema em comparação com outros algoritmos existentes
- [A24] apresenta uma solução para o problema de monitoramento de cargas de trabalho dinâmicas em data centers em nuvem. O artigo propõe um algoritmo de consolidação de servidores baseado em processos estocásticos para reduzir o custo total dos data centers enquanto satisfaz os requisitos de qualidade de serviço. O algoritmo é avaliado por meio de simulações e os resultados mostram que ele é capaz de melhorar significativamente o desempenho do sistema em comparação com outros algoritmos existentes.
- Em [A27] apresenta uma solução para o problema de monitoramento de recursos em tempo real em aplicativos de comércio eletrônico em nuvem. O artigo propõe um algoritmo de seleção de política de estimativa de recursos para escolher a melhor política de estimativa de recursos para uma carga de trabalho específica. O algoritmo é avaliado por meio de simulações e os resultados mostram que ele é capaz de melhorar significativamente o desempenho do sistema em comparação com outros algoritmos existentes.

Os principais critérios e indicadores utilizados para avaliar esses métodos são:

- Qualidade: a capacidade dos métodos de detectar anomalias com precisão, sensibilidade e especificidade;
- Confiabilidade: a capacidade dos métodos de manter um desempenho consistente em diferentes condições e cenários;
- Segurança: a capacidade dos métodos de proteger os dados e os serviços contra ataques maliciosos ou fraudes.

Os principais desafios e oportunidades para melhorar esses métodos são:

- Lidar com a complexidade, a heterogeneidade e a incerteza dos dados provenientes dos ambientes de computação distribuída;
- Projetar técnicas de monitoramento e detecção de anomalias que sejam escaláveis, rápidas, adaptáveis e eficientes;
- Considerar os requisitos e as expectativas dos usuários finais na definição e na avaliação das anomalias.

## 2.7 CONCLUSÃO

A presente revisão sistemática teve como objetivo identificar as abordagens mais eficazes e inovadoras para a previsão da disponibilidade de recursos computacionais em sistemas distribuídos, com foco no uso da carta de controle I-AM e técnicas de controle estatístico de processos. Foram selecionados 34 artigos publicados entre 2018 e 2023, abordando diferentes aspectos da previsão, monitoramento e gerenciamento de recursos computacionais em sistemas distribuídos.

Os resultados da análise dos artigos destacam a relevância e atualidade do tema da disponibilidade de recursos computacionais em sistemas distribuídos, que envolve diversos desafios e oportunidades. Entre os principais desafios, observa-se a complexidade, heterogeneidade e incerteza dos dados provenientes dos ambientes de computação distribuída.

No que diz respeito às oportunidades, é importante mencionar que a utilização das cartas

de controle e técnicas de controle estatístico de processos mostraram-se promissoras para a previsão da disponibilidade de recursos computacionais em sistemas distribuídos. Essas abordagens permitem o monitoramento contínuo dos recursos, identificando padrões de variação e desvios significativos, o que pode contribuir para a melhoria do desempenho e eficiência dos sistemas.

Além disso, os estudos analisados sugerem a aplicação de técnicas de aprendizado de máquina e inteligência artificial na previsão da disponibilidade de recursos, explorando a capacidade dessas abordagens de lidar com grandes volumes de dados e identificar padrões complexos. Essa integração entre métodos estatísticos e técnicas avançadas de análise de dados pode proporcionar insights valiosos para a tomada de decisões e o planejamento estratégico de recursos computacionais em sistemas distribuídos.

Outra oportunidade identificada nos artigos refere-se à importância do monitoramento em tempo real e da adaptação dinâmica dos recursos com base nas demandas do sistema. A utilização de algoritmos e mecanismos de autogerenciamento pode permitir a otimização da alocação de recursos, a detecção de falhas e a recuperação eficiente, garantindo um alto nível de disponibilidade e desempenho dos sistemas distribuídos.

Apesar dos avanços apresentados, ainda existem lacunas a serem exploradas no campo da previsão da disponibilidade de recursos computacionais em sistemas distribuídos. É necessário investigar e desenvolver abordagens mais precisas e robustas, considerando a diversidade de ambientes distribuídos e a natureza dinâmica das cargas de trabalho. Além disso, é fundamental considerar questões de segurança e privacidade no contexto da coleta e análise de dados para previsão de recursos.

Em conclusão, a presente revisão sistemática ressalta a importância de abordagens eficazes e inovadoras para a previsão da disponibilidade de recursos computacionais em sistemas distribuídos. Os resultados obtidos demonstram o potencial das técnicas estatísticas, do aprendizado de máquina e do monitoramento em tempo real para otimizar a utilização dos recursos e melhorar o desempenho dos sistemas distribuídos.

### 3 METODOLOGIA

Esta seção descreve a metodologia proposta para atingir os objetivos estabelecidos nesta pesquisa, seguindo uma abordagem baseada em análise de dados históricos e técnicas estatísticas.

Objetivo:

Esta pesquisa busca desenvolver e avaliar um modelo de previsão que estime de maneira eficaz a disponibilidade de recursos computacionais em sistemas distribuídos. Tal investigação é motivada pelo crescente papel dos sistemas distribuídos em diversas aplicações e pela necessidade de garantir uma gestão eficiente de seus recursos.

Hipóteses:

Tendo em vista a revisão literária conduzida e os desafios contemporâneos da área, foram delineadas as seguintes hipóteses:

- H1: O modelo de previsão proposto superará em precisão os métodos existentes na literatura quando se trata de prever a disponibilidade de recursos em sistemas distribuídos. Esta hipótese surge da identificação de potenciais lacunas e limitações em abordagens correntes, e da proposição de técnicas inovadoras incorporadas a este modelo.
- H2: O modelo apresentado demonstrará robustez e adaptabilidade em face de diferentes cenários inerentes aos sistemas distribuídos. Espera-se que ele seja resiliente a variações de carga, falhas de equipamentos e flutuações nas demandas dos usuários. Esta hipótese é motivada pelo reconhecimento de que a dinâmica e a complexidade dos sistemas distribuídos exigem soluções que sejam tanto precisas quanto versáteis.

Estas hipóteses guiam os métodos de avaliação e testes empíricos, ajudando a moldar a análise dos resultados e a discussão das contribuições desta pesquisa no domínio da previsão em sistemas distribuídos.

Para desenvolver o modelo de previsão, inicialmente foram analisadas séries temporais do uso de CPU utilizando técnicas estatísticas como ACF, PACF e teste ADF. Essas análises permitiram identificar as características e correlações relevantes para ajustar os parâmetros do modelo do modelo proposto.

Durante a fase de análise preliminar dos dados e modelagem das séries temporais, foram empregadas as técnicas citadas para entender e modelar a dinâmica subjacente dos dados. Abaixo, é apresentada uma breve descrição de cada técnica:

- Função de Autocorrelação (ACF): A Função de Autocorrelação, comumente referida como ACF, é uma ferramenta que permite visualizar e quantificar a correlação entre observações de uma série temporal em diferentes pontos no tempo. Em termos simples, ela mede a relação entre um valor na série temporal e os valores anteriores. A ACF foi utilizada neste estudo para identificar a presença de padrões sazonais e a dependência temporal nos dados.
- Função de Autocorrelação Parcial (PACF): A PACF é semelhante à ACF, mas, enquanto a ACF mostra a correlação combinada de um valor com seus valores anteriores, a PACF isola a correlação de um valor com um valor específico no passado, excluindo as influências de outros atrasos. Essa técnica é crucial para determinar os termos de defasagem adequados em modelos autoregressivos.
- Teste de Dickey-Fuller Aumentado (ADF): O teste ADF é uma técnica usada para testar a presença de raiz unitária em uma série temporal, o que é uma indicação de não estacionariedade. Em outras palavras, é uma ferramenta para determinar se uma série temporal possui uma tendência ou um padrão sazonal que precisa ser diferenciado (transformado) antes de ser modelada. O teste ADF foi escolhido por sua robustez e ampla aceitação na literatura para testar a estacionariedade de séries temporais.

O uso dessas técnicas estatísticas, em conjunto, forneceu uma fundação sólida para a análise subsequente e modelagem dos dados. Além disso, a familiarização com essas ferramentas pode ser essencial para leitores e pesquisadores que desejam replicar ou entender em profundidade o método empregado neste estudo.

Visando maior precisão, também foi incorporado um módulo de atualização do modelo com base nas observações mais recentes do uso de CPU. Dessa forma, a previsão pode ser ajustada continuamente com os dados atuais.

Para avaliar a eficiência do modelo, foram utilizadas métricas como MAE e RMSE, adequadas

para a unidade de medida prevista. Análises comparativas dos resultados foram realizadas para verificar o nível de assertividade em relação a outros modelos.

**Erro Médio Absoluto (MAE):** Esta métrica calcula a média das diferenças absolutas entre as previsões e os valores reais. Oferece uma visão clara da magnitude do erro, sem considerar a direção.

**Erro Quadrático Médio (RMSE):** Mede a média das diferenças quadradas entre previsões e observações. Dá mais peso a grandes erros, sendo especialmente útil quando grandes desvios são particularmente indesejáveis.

Os próximos passos envolvem detalhar os processos de construção, atualização e aplicação do modelo para prever a disponibilidade futura de CPU. Serão discutidos aspectos como seleção de variáveis, tratamento de dados faltantes, otimização de parâmetros e integração com o sistema de gerenciamento de recursos.

O foco será apresentar uma abordagem quantitativa baseada em séries temporais para prever a disponibilidade de recursos computacionais de forma precisa e confiável.

### **3.1 COLETA DE DADOS**

A coleta de dados desempenha um papel fundamental nesta pesquisa, fornecendo informações históricas sobre a disponibilidade de recursos computacionais em sistemas distribuídos. A coleta de dados é realizada de forma abrangente e representativa, com o objetivo de capturar diferentes cenários e variações ao longo do tempo.

Para obter os dados necessários, são utilizados registros de tempo de atividade, tempos de resposta e capacidade do sistema. Outros indicadores relevantes também podem ser obtidos. As fontes dos dados incluem registros de monitoramento do sistema, logs de eventos e bancos de dados históricos.

Os dados a serem coletados consistem nas métricas de uso e disponibilidade de CPU em diversas máquinas pertencentes a sistemas distribuídos. Essas métricas incluem: Percentual de uso de CPU e Percentual de CPU ociosa/disponível. As métricas serão coletadas ao longo do tempo, em intervalos regulares (por exemplo, a cada 1 segundo), por um período significativo



(por exemplo, vários meses), de forma a capturar variações e padrões de uso.

As métricas são obtidas por meio da ferramenta sysstat, que permite monitorar o desempenho de sistemas Linux. O sysstat provê o sar, que coleta dados de uso de CPU e outros recursos. Os dados serão coletados via SSH nos servidores que compõem o sistema distribuído. O sysstat será executado via scripts Bash nos servidores, coletando as métricas de uso de CPU via chamadas ao sar.

A escolha do pacote sysstat, especificamente a ferramenta sar, foi feita após considerar várias outras ferramentas de monitoramento e coleta de dados disponíveis. Os motivos que levaram a esta decisão incluem:

- **Abrangência de Métricas:** O sar oferece uma ampla gama de métricas relacionadas ao sistema, abrangendo desde uso de CPU, memória, I/O de disco, atividade de rede, entre outras. Essa abrangência permite uma visão holística do desempenho e recursos do sistema.
- **Histórico e Confiabilidade:** O pacote sysstat tem um histórico longo e comprovado na comunidade Linux. Ele é amplamente utilizado em ambientes de produção em todo o mundo, garantindo confiabilidade na coleta de métricas.
- **Flexibilidade:** O sar permite a coleta de métricas em diferentes intervalos de tempo, desde segundos até dias, tornando-o flexível para diferentes necessidades de monitoramento.
- **Facilidade de Integração:** Dada sua popularidade e padrão de saída bem documentado, o sar é fácil de integrar com outras ferramentas e plataformas de análise de dados.
- **Considerações de Alternativas:** Outras ferramentas, como top, htop, iostat, e vmstat, foram consideradas. Embora cada uma tenha suas próprias vantagens, o sar foi escolhido por sua capacidade de fornecer uma combinação de todas essas métricas de maneira consolidada e pela facilidade de coletar dados ao longo do tempo para análises de séries temporais.

Ao optar pelo sysstat e sar, o objetivo foi adotar uma ferramenta que não apenas atendesse às necessidades imediatas da pesquisa, mas que também fosse reconhecida e confiável na

indústria, garantindo a validade e aplicabilidade dos resultados obtidos.

Os dados são consolidados em uma instância PostgreSQL com timescaleDB para armazenamento eficiente de séries temporais. Esta base de dados centralizada facilita a análise exploratória e serve como fonte única da verdade para treinamento dos modelos.

Na fase inicial do projeto, a escolha de uma solução robusta e confiável de gerenciamento de banco de dados era primordial. Após considerar diversas opções disponíveis no mercado, optou-se pelo PostgreSQL em conjunto com o TimescaleDB por várias razões:

- **Maturidade e Confiabilidade do PostgreSQL:** O PostgreSQL é um sistema de gerenciamento de banco de dados relacional de código aberto, em operação há mais de duas décadas. Sua estabilidade, segurança e capacidade de lidar com grandes volumes de dados o tornam uma escolha preferida para muitos projetos de pesquisa e corporativos.
- **Extensibilidade do TimescaleDB:** O TimescaleDB, enquanto extensão do PostgreSQL, oferece capacidades especializadas para séries temporais. Isso é particularmente útil para o contexto da pesquisa, lidando com dados sequenciais e cronológicos. Ele permite operações eficientes de inserção, consulta e análise de séries temporais, otimizando o desempenho sem sacrificar a familiaridade do PostgreSQL.
- **Comparação com Outras Soluções:** Outros sistemas de gerenciamento de banco de dados, tanto relacionais quanto voltados para séries temporais, foram considerados. No entanto, a combinação de uma base robusta (PostgreSQL) com as funcionalidades específicas para séries temporais (TimescaleDB) foi determinante. Além disso, a ampla comunidade e a vasta documentação disponíveis para ambas as tecnologias foram fatores adicionais que pesaram na decisão.
- **Flexibilidade e Escalabilidade:** A capacidade do TimescaleDB de escalar horizontalmente, juntamente com o suporte do PostgreSQL a diversas funções e extensões, oferece uma solução flexível e adaptável. Isso permite ajustar e expandir as operações de banco de dados conforme o projeto evolui.

Em resumo, a decisão de adotar PostgreSQL juntamente com TimescaleDB foi baseada em uma combinação de necessidades específicas do projeto, confiabilidade comprovada das

tecnologias e sua adaptabilidade a desafios futuros.

Para lidar com dados incompletos, são removidas linhas com valores ausentes. Se houver lacunas temporais significativas, técnicas de imputação como interpolação poderão ser utilizadas. Outliers serão identificados e removidos com base em desvios padrão em relação à média.

Antes da análise, os dados são normalizados para a faixa de 0 a 1. Verificações de distribuição e transformações como raiz quadrada podem ser aplicadas conforme necessário para garantir que pressupostos dos modelos sejam atendidos.

Este processo de coleta visa garantir dados completos, confiáveis e relevantes para possibilitar a construção de modelos precisos de previsão de uso de CPU. A infraestrutura e métodos utilizados permitirão obter uma base de dados de alta qualidade para treinamento e avaliação dos modelos.

Em resumo, a coleta de dados é realizada de forma cuidadosa e abrangente, buscando capturar informações relevantes sobre a disponibilidade de recursos computacionais em sistemas distribuídos. A qualidade e a representatividade dos dados são consideradas para garantir resultados sólidos e confiáveis na análise subsequente.

## **3.2 ANÁLISE EXPLORATÓRIA DE DADOS**

A análise exploratória de dados é uma etapa essencial desta pesquisa, que tem como objetivo compreender e explorar os dados coletados sobre a disponibilidade de recursos computacionais. Essa análise inicial dos dados permite identificar padrões, tendências e características importantes, além de fornecer insights valiosos para o desenvolvimento do modelo de previsão.

Inicialmente, os dados são organizados e estruturados de forma adequada para facilitar a análise. São utilizadas técnicas de pré-processamento de dados, como limpeza, normalização e transformação, para garantir a consistência e a qualidade dos dados.

São aplicadas técnicas estatísticas e visualizações gráficas para explorar os dados em diferentes perspectivas. Isso pode incluir a análise de estatísticas descritivas, como média, mediana, desvio padrão e quartis, para entender a distribuição dos dados e identificar valores atípicos. Gráficos, como histogramas, gráficos de dispersão e séries temporais, também são utilizados

para visualizar padrões e tendências ao longo do tempo.

Além disso, são realizadas análises de correlação para identificar possíveis relações entre as variáveis de interesse, como a disponibilidade de recursos e os tempos de resposta. Isso permitirá compreender a interação entre essas variáveis e fornecer insights para o desenvolvimento do modelo de previsão.

A análise exploratória de dados também envolve a identificação de possíveis padrões sazonais, variações diárias ou semanais e outros fenômenos relevantes que possam afetar a disponibilidade de recursos. Isso é realizado por meio de técnicas de decomposição de séries temporais e análise de componentes principais.

Para a realização da Análise Exploratória de Dados (AED), foi escolhida a linguagem de programação R, reconhecida por sua capacidade analítica e versatilidade em lidar com conjuntos de dados complexos. Algumas razões para a escolha do R incluem:

- **Eficácia em Análise Estatística:** O R é uma ferramenta poderosa para análise estatística, sendo amplamente utilizado em pesquisas acadêmicas e industriais por sua precisão e extensa gama de pacotes dedicados.
- **Pacote ggplot2:** Especificamente, para visualização de dados, foi utilizado o pacote ggplot2. Este pacote, que faz parte do conjunto Tidyverse no R, é uma das ferramentas mais populares e versáteis para criar gráficos estatísticos de alta qualidade. A escolha do ggplot2 é motivada por sua capacidade de produzir visualizações complexas a partir de comandos simples e pela sua ampla flexibilidade em personalizar gráficos, o que permite extrair insights mais profundos dos dados.
- **Comunidade Ativa:** O R tem uma comunidade muito ativa, o que significa que há uma vasta quantidade de recursos, tutoriais e pacotes adicionais disponíveis. Isso facilita a solução de problemas, o aprendizado de novas técnicas e a colaboração com outros pesquisadores.
- **Integração com Outras Ferramentas:** O R se integra bem com outras ferramentas e plataformas, permitindo uma análise mais fluida. Por exemplo, é possível conectar o R diretamente ao PostgreSQL, TimescaleDB, web scraping tools e outras linguagens de

programação.

- Ao usar o R juntamente com o ggplot2, busca-se garantir uma análise rigorosa e visualizações claras, que permitirão uma compreensão mais profunda dos dados e ajudarão a guiar decisões futuras no projeto.

É importante ressaltar que a análise exploratória de dados será iterativa, ou seja, novas descobertas e insights podem direcionar ajustes nos procedimentos de coleta de dados, pré-processamento e análise subsequente. Essa abordagem iterativa permitirá explorar de maneira mais completa as características dos dados e auxiliará no desenvolvimento de um modelo de previsão mais robusto e preciso.

A AED não é um processo linear, mas sim um ciclo iterativo e adaptativo. O que isso significa é que, ao mergulhar nos dados durante a AED, podem surgir insights, anomalias ou padrões não esperados que demandem uma revisitação à fase de coleta de dados ou uma reavaliação das abordagens de modelagem. Este loop de feedback é essencial para garantir que os modelos desenvolvidos sejam tanto precisos quanto relevantes.

Por exemplo, durante a AED, pode-se descobrir que certos dados são consistentemente problemáticos ou que faltam informações cruciais em determinados períodos. Esta descoberta pode levar de volta à fase de coleta para ajustar a maneira como os dados são capturados ou para buscar informações adicionais. Da mesma forma, a AED pode revelar características dos dados que sugerem que um tipo diferente de modelo estatístico ou abordagem de aprendizado de máquina pode ser mais adequado do que o originalmente previsto.

Desta forma, o loop de feedback entre a AED, a coleta de dados e a modelagem não só assegura que os modelos sejam otimizados em relação aos dados disponíveis, mas também reforça a ideia de que a pesquisa é um processo vivo. Ela evolui e se adapta à medida que se avança na exploração, garantindo uma abordagem holística que busca a máxima precisão e relevância no mundo real.

Em suma, a análise exploratória de dados terá um papel fundamental no desenvolvimento do modelo de previsão. Ela permitirá explorar a disponibilidade de recursos computacionais em sistemas distribuídos. Essa análise permitirá compreender a natureza dos dados, identificar

padrões e tendências relevantes e fornecerá subsídios importantes para o desenvolvimento do modelo de previsão.

### 3.3 AVALIAÇÃO DA CARTA DE CONTROLE I-AM

A avaliação da carta de controle I-AM é realizada como parte desta pesquisa para investigar sua eficácia na monitoração da disponibilidade de recursos computacionais em sistemas distribuídos. A carta de controle I-AM é uma ferramenta estatística que permite monitorar processos ao longo do tempo, identificando padrões, tendências e variações que possam indicar problemas ou desvios em relação aos valores esperados.

Nesta etapa, são utilizados os dados coletados previamente sobre a disponibilidade de recursos computacionais. Esses dados são analisados com base nos princípios e técnicas da carta de controle I-AM para avaliar sua aplicabilidade e eficiência na detecção de variações significativas na disponibilidade dos recursos.

A avaliação da carta de controle I-AM envolve a definição dos limites de controle, que são limites estatísticos estabelecidos com base nos dados coletados. Esses limites são utilizados como pontos de observação para desenvolvimento do método de previsão.

Esta avaliação envolve o cálculo dos limites de controle superior (LCS) e inferior (LCI) por meio das fórmulas:

$$\text{LCS} = \text{Média} + d2 * \text{Desvio Padrão} \quad \text{LCI} = \text{Média} - d2 * \text{Desvio Padrão}$$

Onde  $d2$  é um fator tabelado que depende do tamanho da amostra.

Os dados são plotados na carta de controle e pontos fora dos limites são identificados seguindo as regras de Western Electric. Padrões como tendências, agrupamentos e misturas indicam alterações na disponibilidade de recursos.

Mudanças são simuladas artificialmente nos dados para verificar a capacidade de detecção da carta de controle por meio da inserção de outliers e alterações na média e variância.

São analisadas as seguintes características da carta de controle I-AM:

Cálculo dos limites de controle: São utilizadas técnicas estatísticas para calcular os limites de

controle superiores e inferiores com base nos dados coletados. Esses limites são estabelecidos de forma a garantir que variações dentro desses limites sejam consideradas normais, enquanto variações fora desses limites indiquem alterações.

Identificação de pontos fora dos limites de controle: É verificado se a carta de controle I-AM é capaz de identificar corretamente os pontos fora dos limites de controle, indicando variações significativas na disponibilidade de recursos. São analisados tanto os pontos acima do limite superior quanto os pontos abaixo do limite inferior.

Análise de padrões e tendências: São observados padrões e tendências nos dados, como sequências de pontos consecutivos acima ou abaixo dos limites de controle, que também possam indicar mudanças recorrentes ou tendências de melhoria ou deterioração na disponibilidade de recursos.

Capacidade de detecção de mudanças: É avaliada a capacidade da carta de controle I-AM de detectar mudanças significativas na disponibilidade de recursos. São realizadas simulações de alterações artificiais nos dados e verificada a capacidade da carta de controle de identificar essas mudanças.

A avaliação da carta de controle I-AM permite verificar sua eficácia no monitoramento contínuo da disponibilidade de recursos. Isso é feito em sistemas computacionais distribuídos. Os resultados dessa avaliação são utilizados para determinar a viabilidade e utilidade da carta de controle I-AM como uma ferramenta de monitoração nesse contexto específico e para compará-la com outras abordagens de previsão de disponibilidade de recursos.

### **3.4 DESENVOLVIMENTO DO MODELO DE PREVISÃO**

Nesta etapa da pesquisa, será realizado o desenvolvimento do modelo de previsão de disponibilidade de recursos computacionais em sistemas distribuídos. O objetivo é propor um modelo que seja aplicável, preciso e confiável na previsão da disponibilidade desses recursos, auxiliando os escalonadores de tarefas.

Para o desenvolvimento do modelo de previsão, serão adotadas as seguintes etapas:

Seleção e preparação dos dados: Serão selecionados os dados relevantes para a previsão da

disponibilidade de recursos computacionais. Esses dados podem incluir informações como histórico de utilização dos recursos, métricas de desempenho, dados de monitoramento, entre outros. Os dados selecionados serão preparados, realizando-se processos de limpeza, tratamento de valores ausentes e normalização, a fim de garantir a qualidade dos dados utilizados no modelo.

Escolha e implementação do algoritmo de previsão que serão utilizados para comparação: Será realizada uma revisão dos principais algoritmos de previsão existentes na literatura, levando em consideração sua adequação ao contexto de sistemas distribuídos. Algoritmos como médias móveis, suavização exponencial, ARIMA, algoritmos de aprendizado de máquina e redes neurais serão considerados.

Treinamento e validação do modelo: O modelo selecionado será treinado utilizando os dados disponíveis, dividindo-os em conjuntos de treinamento e validação. O conjunto de treinamento será utilizado para ajustar os parâmetros do modelo, enquanto o conjunto de validação será usado para avaliar o desempenho do modelo em dados não vistos durante o treinamento. Serão utilizadas métricas de desempenho, como erro médio absoluto (MAE) e erro quadrático médio (RMSE), para avaliar a precisão do modelo.

Ajuste e otimização do modelo: Caso necessário, o modelo será ajustado e otimizado para melhorar sua precisão e desempenho. Serão exploradas diferentes configurações de parâmetros e técnicas de ajuste para encontrar a melhor combinação que maximize a qualidade das previsões.

Teste do modelo em diferentes cenários: O modelo desenvolvido será testado em diferentes cenários para avaliar sua robustez e capacidade de generalização. Serão consideradas variações de carga, falhas de equipamentos, mudanças nas demandas dos usuários, entre outros fatores que possam afetar a disponibilidade de recursos. Os resultados obtidos serão analisados e comparados com outras abordagens de previsão, incluindo a carta de controle I-AM, para avaliar a eficácia e vantagens do modelo proposto.

Ao final desta etapa, espera-se ter um modelo de previsão de disponibilidade de recursos computacionais em sistemas distribuídos que seja preciso, confiável e capaz de auxiliar na



tomada de decisões relacionadas ao gerenciamento e alocação desses recursos. O modelo desenvolvido será uma contribuição para o avanço do conhecimento na área e poderá ser aplicado em ambientes reais para melhorar a qualidade do serviço oferecido.

### 3.5 AVALIAÇÃO E COMPARAÇÃO DO MODELO

Nesta etapa da pesquisa, será realizada a avaliação e comparação do modelo de previsão proposto com outras abordagens de previsão de disponibilidade de recursos computacionais em sistemas distribuídos. O objetivo é verificar a eficácia, precisão e desempenho do modelo desenvolvido em relação às técnicas existentes, proporcionando uma compreensão clara das vantagens e contribuições do modelo proposto.

Para a avaliação e comparação do modelo, serão realizadas as seguintes atividades:

Seleção de métricas de avaliação: Serão selecionadas métricas adequadas para avaliar a precisão e desempenho do modelo de previsão. Métricas como erro médio absoluto (MAE), erro quadrático médio (RMSE), erro percentual absoluto médio (MAPE) e coeficiente de determinação ( $R^2$ ) podem ser utilizadas. Essas métricas fornecerão uma visão quantitativa dos resultados e permitirão comparar o desempenho do modelo proposto com outras abordagens de previsão.

Realização de experimentos: Serão realizados experimentos comparativos, nos quais o modelo proposto será testado em diferentes conjuntos de dados e cenários. Será analisado o desempenho do modelo em relação aos benchmarks selecionados, avaliando sua precisão na previsão da disponibilidade de recursos computacionais. Serão consideradas variações de carga, falhas de equipamentos, mudanças nas demandas dos usuários, entre outros fatores relevantes.

Análise e interpretação dos resultados: Os resultados obtidos nos experimentos serão analisados e interpretados de forma crítica. Será verificado se o modelo proposto alcança uma maior precisão e desempenho em comparação com as abordagens de benchmark. Serão identificadas as vantagens e limitações do modelo proposto, bem como os cenários em que ele se destaca em relação às outras técnicas de previsão.

Os resultados serão analisados considerando a consistência interna e a validade em relação

à literatura existente sobre previsão de recursos em sistemas distribuídos. As limitações do estudo também serão discutidas, como a generalização para outros contextos e tipos de sistemas.

Será realizada uma comparação crítica entre os achados desta pesquisa e de estudos prévios relacionados, identificando convergências e divergências. As implicações teóricas e práticas dos resultados serão exploradas, discutindo como a pesquisa avança o conhecimento e suas aplicações.

Também serão propostas oportunidades de estudos futuros que surgiram com base nos resultados, delineando novas questões e problemas em aberto. Os aspectos éticos envolvidos no uso de dados serão discutidos.

Será feita uma análise de como o rigor metodológico empregado influenciou na validação dos resultados. As principais descobertas serão interpretadas sob múltiplas perspectivas, explorando seus diversos significados e impactos.

A discussão buscará explorar os resultados em profundidade, situando-os no contexto da literatura existente e extraíndo implicações teóricas e práticas. As limitações e aspectos éticos também serão abordados criticamente. Dessa forma, espera-se realizar uma análise ampla e multifacetada sobre esta pesquisa.

Ao final desta etapa, espera-se obter uma avaliação clara e objetiva do modelo proposto em comparação com outras abordagens de previsão. A análise e comparação permitirão identificar as vantagens e contribuições do modelo desenvolvido, fornecendo insights importantes para o avanço da pesquisa na área de previsão da disponibilidade de recursos computacionais em sistemas distribuídos.

### **3.6 CONSIDERAÇÕES FINAIS DA METODOLOGIA**

Ao refletir sobre a metodologia adotada nesta pesquisa, algumas considerações cruciais vêm à tona:

- **Complexidade dos Sistemas Distribuídos:** Sistemas distribuídos são inerentemente complexos e variáveis. A abordagem adotada procurou capturar essa complexidade,

mas é vital lembrar que modelos são simplificações da realidade e podem não abranger todas as nuances desses sistemas.

- **Escolhas Metodológicas:** As ferramentas e técnicas selecionadas para esta pesquisa foram baseadas na revisão da literatura e nas melhores práticas conhecidas. No entanto, o campo de previsão de disponibilidade de recursos em sistemas distribuídos é vasto, e outras abordagens podem oferecer insights diferentes.
- **Iteratividade e Adaptação:** A metodologia reconhece a importância da natureza iterativa da Análise Exploratória de Dados (AED) e sua interação com outras fases da pesquisa. À medida que novos insights emergem, é vital adaptar e refinar a abordagem.
- **Reconhecimento de Limitações:** Cada metodologia tem seus desafios e limitações. Ao identificá-los explicitamente, busca-se uma pesquisa transparente e honesta, abrindo espaço para aprimoramentos e futuras investigações no campo.
- **Colaboração Interdisciplinar:** A intersecção de estatísticas, ciência da computação e engenharia nesta pesquisa destaca a necessidade de uma abordagem interdisciplinar. Trabalhar com expertise diversificada pode enriquecer os resultados e aprimorar a aplicabilidade prática do modelo.

Em suma, a metodologia escolhida representa um esforço consciente para abordar a previsão da disponibilidade de recursos em sistemas distribuídos de forma rigorosa e sistemática. As considerações apresentadas visam não apenas contextualizar as escolhas feitas, mas também guiar pesquisadores e profissionais que desejam adotar ou adaptar esta abordagem no futuro.

- Branco, K. R. L. J. C. 2004. “Índices de Carga e Desempenho Em Ambientes Paralelos/Distribuídos-Modelagem e Métricas.” PhD thesis, Universidade de São Paulo.
- Coulouris, G., J. Dollimore, T. Kindberg, and G. Blair. 2013. *Sistemas Distribuídos: Conceitos e Projeto*. Bookman Editora.
- Ferreira, V. et al. 2018. “APLICAÇÃO DA CARTA DE CONTROLE i-AM PARA ESTUDO DE CAPABILIDADE EM UMA ORGANIZAÇÃO DO SETOR DE CATERING AÉREO.” In *Anais Do XXXVIII Encontro Nacional de Engenharia de Produção*. [https://abepro.org.br/biblioteca/TN\\_WIC\\_259\\_487\\_36148.pdf](https://abepro.org.br/biblioteca/TN_WIC_259_487_36148.pdf).
- Fontoura, E. S. et al. n.d. “AvSchedP: Previsao de Disponibilidade Para Grades Oportunistas.”
- Minitab. 2020. “Visão Geral - Carta i-MR - Minitab.” 2020. <https://support.minitab.com/pt-br/minitab/20/>.
- Oliveira, T. P., J. S. Barbar, and A. S. Soares. 2015. “Predição Do Tráfego de Rede de Computadores Usando Redes Neurais Tradicionais e de Aprendizagem Profunda.” *Revista de Informática Teórica e Aplicada* 22 (1): 10–28.
- Paulino, B. C. G. 2016. “Software de Análise e Monitoramento de Hardware.”
- Pinto, P. V. R. 2021. “Previsão Da Capacidade de Processamento Em Computadores Pessoais Utilizando ARIMA e Assinatura Comportamental.” Master’s thesis, Universidade Tecnológica Federal do Paraná.
- Santana, M. N. P. 2006. “Alocacao de Tarefas Paralelas Comunicantes Em Ambientes Distribuídos Heterogeneos.”
- Santos, E. 2016. “Alta Disponibilidade Em Sistemas Distribuídos Baseados Em Arquitetura de Software Orientada a Microserviços.” PhD thesis.
- Sharma, M. 2009. “RESOURCE MANAGEMENT IN DISTRIBUTED SYSTEMS.” 2009. [https://www.researchgate.net/profile/Manoj-Sharma-100/publication/328389975\\_RESOURCE\\_MANAGEMENT\\_IN\\_DISTRIBUTED\\_SYSTEMS/links/5bca0533a6fdcc03c795ff59/RESOURCE-MANAGEMENT-IN-DISTRIBUTED-SYSTEMS.pdf](https://www.researchgate.net/profile/Manoj-Sharma-100/publication/328389975_RESOURCE_MANAGEMENT_IN_DISTRIBUTED_SYSTEMS/links/5bca0533a6fdcc03c795ff59/RESOURCE-MANAGEMENT-IN-DISTRIBUTED-SYSTEMS.pdf).
- Silva, T., G. Pereira, M. Pereira, S. S. Russo, D. S. Russo, and P. Cunha. 2020. “Alta Disponibilidade Em Sistemas Distribuídos Baseados Em Arquitetura de Software Orientada a Microservicos.” In *Brazilian Symposium on Software Engineering*.
- Tanenbaum, A. S. 2007. *Sistemas Distribuídos: Principios e Paradigmas*. Sao Paulo.

- Tesser, R. K. 2011. “Monitoramento on-Line Em Sistemas Distribuidos: Mecanismo Hierarquico Para Coleta de Dados.”
- Varghese, B., and R. Buyya. 2018. “Next Generation Cloud Computing: New Trends and Research Directions.” *Future Generation Computer Systems* 79: 849–61.
- Wen, Y. et al. 2020. “CPU Usage Prediction for Cloud Resource Provisioning Based on Deep Belief Network and Particle Swarm Optimization.” *Concurrency and Computation: Practice and Experience* 32 (14): e5730.