

FICHAMENTO
PLANO DE TRABALHO
TRABALHOS RELACIONADOS

LEGENDA

- Conteúdo adicionado
- Conteúdo a ser melhorado
- Conteúdo a ser excluído
- Conteúdo a OK

Exame de Qualificação

Para realizar a defesa do exame de qualificação o candidato deve preencher:

Formulário de solicitação de defesa (DIRPPG): deve ser preenchido digitalmente, impresso e assinado. Após, entregar na secretaria do curso.

O documento de exame de qualificação deve conter os requisitos mínimos:

- Capa - OK
- Folha de rosto - OK
- Resumo - OK
- Abstract - OK
- Introdução - OK
- Objetivos - OK
- Organização do trabalho - OK
- Estado da Arte (Revisão da Literatura)
- Proposta de Dissertação
- Metodologia
- Resultados esperados
- Experimentos realizados (se for o caso)
- Resultados parciais obtidos (se for o caso)
- Cronograma
- Conclusão
- Referências

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

RESUMO

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. A viabilidade da abordagem é demonstrada por meio da análise de um caso de estudo.

Abstract

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 for monitoring CPU utilization rates on machines in a distributed system, with the aim of predicting their availability. The feasibility of the approach is demonstrated through the analysis of a case study.

INTRODUÇÃO

A disponibilidade de recursos computacionais é fundamental para garantir a eficiência e a escalabilidade em sistemas distribuídos, Santos (2016)(1) 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. (2007)(2) 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 prever esta disponibilidade. Tanenbaum (2007)(3) 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. Silva (2005)(4) 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.

Batista e Oliveira (2020)(5) 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. Costa et al. (2020)(6) destacaram que a disponibilidade de CPU é um aspecto importante a ser monitorado no mapeamento de disponibilidade de recursos computacionais. A partir de Pinto (2021)(7), 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 (2015)(8) 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 (Duczmal, 2000)(9).

Batista e Oliveira (2020)(10) 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. Utilize 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 (Ferreira et al., 2018)(19).

Segundo Minitab (2021)(20), 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.

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)(11).

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 e Buyya, 2018)(12).

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. Alguns exemplos de técnicas de gerenciamento de recursos incluem (Kumar, 2009)(13):

- 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 redistribuição 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)(14). Isso pode ajudar a garantir que os recursos computacionais sejam utilizados de forma eficiente e que as tarefas sejam concluídas de forma otimizada.

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.

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.;
- Propor um modelo de previsão de disponibilidade de recursos computacionais que seja aplicável em sistemas distribuídos, com alta precisão e confiabilidade;
- Determinar o intervalo ideal para a atualização dos parâmetros de disponibilidade, visando reduzir o tempo de resposta e garantir a eficiência do modelo;
- Avaliar a precisão e eficiência do modelo proposto em comparação com outras abordagens de previsão de disponibilidade de recursos computacionais, como médias móveis, suavização exponencial, ARIMA, algoritmos de aprendizado de máquina e redes neurais;
- 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.

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.

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)(15).

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., 2007).(16)

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., 2016)(17) .

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)(18).

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.

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.

Essa análise de sensibilidade também pode ser útil para entender o comportamento do padrão em diferentes cenários e identificar possíveis limitações ou pontos fracos do modelo. Além disso, a análise de sensibilidade pode ser útil para avaliar como o padrão de disponibilidade de CPU se compara com outras abordagens de previsão em termos de robustez e capacidade de adaptação a diferentes situações.

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.

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 REFERENCIAL TEÓRICO

Aqui é apresentado uma metodologia para prever a disponibilidade de CPU em sistemas distribuídos utilizando carta de controle I-AM. Sistemas distribuídos são aqueles que possuem múltiplos componentes interconectados, que podem estar localizados em diferentes lugares, e que trabalham em conjunto para realizar uma tarefa específica. A disponibilidade de CPU é um indicador importante da performance de um sistema distribuído, pois mede a quantidade de tempo em que os recursos de computação estão disponíveis para realizar tarefas. Neste sentido, a utilização da carta de controle pode ser uma técnica útil para monitorar a disponibilidade de CPU em sistemas distribuídos.

O presente trabalho tem como objetivo avaliar a eficácia da utilização de carta de controle I-AM para prever a disponibilidade de CPU em sistemas destruídos. A carta de controle I-AM é uma ferramenta de análise de dados que permite a monitoração e o controle da qualidade de um processo. Ela é amplamente utilizada em diversas indústrias, mas poucos estudos têm abordado sua aplicação em sistemas de computação.

Para abordar essa questão, foi realizada uma revisão sistemática dos principais estudos sobre o tema. Foram encontrados alguns trabalhos que discutem a previsão de disponibilidade de recursos em sistemas de computação, mas não foram encontrados nenhum estudo que avaliasse a eficácia da utilização de carta de controle I-AM para prever a disponibilidade de CPU em sistemas destruídos.

A estrutura do referencial teórico será a seguinte: primeiramente, são apresentados os conceitos fundamentais de disponibilidade de CPU e os principais métodos utilizados para medir e prever a disponibilidade de recursos em sistemas de computação. Em seguida, é discutida a carta de controle I-AM e suas aplicações em diversas indústrias. Por último, são examinados os possíveis desafios e

oportunidades para a utilização da carta de controle I-AM para prever a disponibilidade de CPU em sistemas destruídos, incluindo a coleta e o tratamento de dados, a modelagem de sistemas e a validação dos resultados. Será concluído com uma visão geral das principais conclusões e perspectivas futuras para essa área de pesquisa.

2.1 COMPUTAÇÃO PARALELA E DISTRIBUÍDA

HPC (High Performance Computing) ou DPC (Distributed/Parallel Computing) é uma rede de computadores independentes conectados por meio de um middleware de distribuição para atingir um objetivo comum, simulando um computador paralelo. Um sistema distribuído oferece a capacidade de compartilhar diferentes recursos e dá a impressão de uma única rede coerente (Tanenbaum 2007). Nos sistemas distribuídos os componentes se comunicam e coordenam suas ações por meio de mensagens.

Um sistema distribuído é um conjunto de máquinas interligadas que trabalham em conjunto para realizar uma tarefa específica. Ele é composto por múltiplos componentes interconectados, que podem estar localizados em diferentes lugares, e é utilizado em uma ampla variedade de aplicações, como redes de computadores, sistemas de controle de processos industriais, sistemas de informação gerencial e sistemas de comunicação (Tanenbaum, 2007).

De acordo com Tanenbaum & van Steen (2002), sistemas distribuídos possuem múltiplos componentes interconectados que trabalham em conjunto para realizar uma tarefa específica. Algumas das principais características dos sistemas distribuídos (Coulouris et al.,2011) incluem:

- Interconexão: os componentes do sistema distribuído estão interconectados por meio de uma rede, o que permite que eles compartilhem informações e realizem tarefas em conjunto;
- Escalabilidade: os sistemas distribuídos são capazes de adicionar ou remover componentes de forma flexível, o que permite que eles se adaptem às mudanças na demanda de trabalho;
- Redundância: os sistemas distribuídos costumam ter componentes redundantes, o que significa que há mais de um componente disponível para realizar a mesma tarefa. Isso aumenta a robustez do sistema, pois permite que ele continue funcionando mesmo que um componente falhe;

- Interoperabilidade: os sistemas distribuídos podem ser compostos por componentes de diferentes fabricantes ou sistemas operacionais, o que exige que eles sejam capazes de se comunicar de forma eficiente;
- Gestão de recursos: os sistemas distribuídos devem ser capazes de gerenciar de forma eficiente os recursos compartilhados, como CPU, memória e armazenamento, para garantir que todos os componentes tenham acesso equitativo a eles.

Outras três características de um sistema distribuído que merecem destaque são (Coulouris, 2013):

- Concorrência: O trabalho pode ser feito em computadores separadamente. Os recursos são compartilhados pela rede somente quando necessário. Esse recurso torna o sistema escalável, pois a capacidade do sistema pode ser aumentada simplesmente adicionando mais recursos;
- Clock não global: Como toda comunicação é feita por meio de envio de mensagens, é difícil sincronizar um relógio global nesse ambiente. Em outros ambientes de computador, um clock global é usado para sincronizar diferentes tarefas do computador;
- Falhas independentes: Cada componente do sistema pode falhar independentemente dos outros. Isso significa que o sistema como um todo pode continuar funcionando mesmo que um componente falhe".

Essas características são fundamentais para o entendimento de sistemas distribuídos e uma visualização da complexidade do mesmo. Organizar e coordenar esse sistema requer mecanismos sólidos de escalonamento de tarefas e de monitoramento de desempenho.

De acordo com Tanenbaum e van Steen (2002), o objetivo principal dos sistemas distribuídos pode ser dividido em seis objetivos principais:

- Compartilhamento de recursos: Os recursos podem ser facilmente compartilhados em um sistema distribuído. Os recursos podem ser hardware, software ou dados que podem ser compartilhados de forma útil. Este é o principal fator para a construção de sistemas distribuídos;
- Open source: Para poder compartilhar recursos, as interfaces dos vários componentes também devem ser tornadas públicas para outros componentes da rede. Isso torna o sistema extensível;

- Escalabilidade: Um sistema distribuído é escalável se o custo de adicionar um usuário for constante em termos dos recursos que devem ser adicionados (Coulouris, 2013). Em um sistema distribuído é fácil adicionar mais recursos simplesmente adicionando mais computadores à rede;
- Tolerância a falhas: Qualquer processo, computador ou rede pode falhar independentemente dos outros conforme descrito acima em “falhas independentes”. Em um sistema distribuído, os recursos podem ser facilmente instalados em muitas máquinas, minimizando a perda desse recurso específico se um computador falhar. O software pode ser instalado nas máquinas de uma rede para detectar até mesmo a perda de um recurso, fazendo com que os outros componentes respondam de acordo;
- Simultaneidade: A simultaneidade pode ser alcançada enviando solicitações de serviços para várias máquinas conectadas à rede ao mesmo tempo ;
- Transparência: Outra finalidade é tornar algumas coisas transparentes para o usuário:
 - Transparência de localização: recursos locais e remotos podem ser acessados sem problemas;
 - Transparência de falhas: O mascaramento de falhas pode ser feito. Consulte “tolerância a falhas” acima;
 - Transparência de replicação: Permite recursos duplicados em vários componentes de forma invisível.

A computação paralela e distribuída é amplamente utilizada em uma variedade de aplicações que exigem alta performance ou processamento em larga escala, conforme apontado por vários autores, como Lee et al. (2018) e Kossmann et al. (2000). Alguns exemplos de aplicações típicas de computação paralela e distribuída incluem:

- Resolução de problemas científicos: muitas vezes, problemas científicos complexos, como a simulação de fenômenos físicos ou a análise de grandes conjuntos de dados, exigem a execução de cálculos intensivos em paralelo, de acordo com pesquisas de autores como Bussmann et al. (2013). Por exemplo, supercomputadores paralelos podem ser usados para simular o comportamento de grandes sistemas, como o clima ou o universo, como mencionado por Du et al. (2015);
- Processamento de grandes quantidades de dados: os sistemas distribuídos são frequentemente usados para processar e analisar grandes quantidades de dados, como os gerados por plataformas de redes sociais ou empresas de tecnologia da informação, conforme descrito por autores como Wang et al.

(2016). Por exemplo, o Google usa clusters de computadores distribuídos para indexar a internet e fornecer resultados de pesquisa rápidos, de acordo com informações disponibilizadas pelo próprio Google (2019);

- Sistemas de informação gerencial: os sistemas distribuídos são amplamente utilizados em sistemas de informação gerencial para gerenciar grandes quantidades de dados e fornecer acesso a eles de maneira consistente, como afirmado por autores como Zhang et al. (2017). Por exemplo, um sistema de informação gerencial distribuído pode ser usado para gerenciar as operações de uma empresa, como vendas, produção e finanças, de acordo com estudo de Chen et al. (2018);
- Sistemas de comunicação: os sistemas distribuídos são importantes para a construção de sistemas de comunicação, como redes de telefonia ou redes de dados, conforme discutido por Liu et al. (2019). Por exemplo, os sistemas de telefonia celular são distribuídos para permitir que os usuários façam chamadas de qualquer lugar, independentemente de onde estejam, como apontado por Li et al. (2016).

Os sistemas paralelos e distribuídos são projetados e organizados de maneira a aproveitar ao máximo a capacidade de processamento de múltiplos computadores e redes, conforme destacado por autores como Silberschatz et al. (2014) e Lin et al. (2015).

Alguns dos principais elementos que são considerados na projeção e organização de sistemas paralelos e distribuídos incluem:

- Middlewares de distribuição: são software que se encontram entre o sistema operacional e a aplicação e fornecem uma camada de abstração para a comunicação entre os componentes do sistema, de acordo com estudo de autores como White et al. (2015). Eles podem ser usados para gerenciar tarefas como alocação de recursos, roteamento de mensagens e tolerância a falhas, conforme apontado por Zhang et al. (2017);
- Protocolos de comunicação: são regras que governam a troca de mensagens entre os componentes do sistema, de acordo com investigação de autores como Kursun et al. (2019). Eles podem ser usados para garantir que as mensagens sejam entregues de forma confiável, para sincronizar os componentes do sistema e para gerenciar o acesso a recursos compartilhados, conforme discutido por Gu et al. (2018);
- Arquiteturas de sistema: são as estruturas lógicas e físicas que determinam como os componentes do sistema são organizados e interagem, conforme explorado por autores como Deelstra et al. (2000). Existem várias

arquiteturas de sistema possíveis para sistemas paralelos e distribuídos, como clusters, nuvens e sistemas descentralizados, de acordo com o trabalho de autores como Zhang et al. (2019);

- Algoritmos de paralelização: são os métodos que são usados para dividir tarefas em subtarefas que podem ser executadas em paralelo, conforme descrito por autores como Liu et al. (2017). Eles podem ser baseados em divisão de dados (dividindo os dados em partes que podem ser processadas em paralelo) ou em divisão de tarefas (dividindo as tarefas em partes que podem ser processadas em paralelo), conforme mencionado por autores como Du et al. (2016);
- Técnicas de programação paralela: são os métodos que são usados para escrever código que pode ser executado em paralelo, de acordo com o trabalho de autores como Grama et al. (2003). Eles podem incluir técnicas como programação com threads, programação com processos ou programação com mensagens, conforme apontado por autores como Foster (2011).

Esses são apenas alguns dos elementos que são considerados na projeção e organização de sistemas paralelos e distribuídos. A escolha desses elementos depende dos requisitos e objetivos do sistema, bem como da plataforma de hardware e software disponível, conforme discutido por autores como Tanenbaum et al. (2007).

A computação paralela e distribuída oferece muitas vantagens em relação a sistemas sequenciais ou monolíticos, mas também apresenta alguns desafios, conforme apontado por autores como Fox et al. (1998) e De Supinski et al. (2005). Algumas dos principais desafios da computação paralela e distribuída incluem:

- Programação paralela: a programação paralela pode ser desafiadora, pois exige que os programadores tenham uma compreensão profunda dos padrões de paralelismo e dos mecanismos de sincronização disponíveis, conforme destacado por autores como Foster et al. (2011).
- Comunicação e sincronização: os componentes de um sistema distribuído precisam se comunicar e sincronizar suas ações de maneira eficiente, o que pode ser um desafio devido à latência e ao overhead de comunicação, conforme investigado por autores como Kale et al. (2007).
- Gerenciamento de recursos: recursos compartilhados podem ser desafiadores em sistemas distribuídos, pois é preciso garantir que todos os componentes tenham acesso equitativo a eles. Isso pode exigir o uso de mecanismos de gerenciamento de recursos, como priorização, alocação

dinâmica e protocolos de arbitragem, conforme discutido por autores como Schmuck et al. (2002).

- Segurança: a segurança é um desafio importante em sistemas distribuídos, pois eles são compostos por múltiplos componentes que podem estar expostos a ameaças externas. É preciso tomar medidas de segurança, como autenticação e criptografia, para proteger os sistemas distribuídos contra ataques e outras ameaças, conforme abordado por autores como Zhao et al. (2006).

Entendendo o propósito dos sistemas distribuídos, é possível entender para que servem. Entendendo a necessidade de sistemas distribuídos, é possível entender a necessidade de aplicativos distribuídos. Isso dá uma visão do quão complexo é manter o sistema de atualização dos status de disponibilidade de recursos, atualizados de falhas e o escalonamento de tarefas.

2.1.2 Disponibilidade de CPU em sistemas distribuídos

A previsão da disponibilidade de CPU em sistemas distribuídos é fundamental para identificar problemas potenciais antes que ocorram e tomar medidas preventivas para garantir a eficiência contínua do sistema (Bragg, 2018; Xu et al., 2019). Diversas técnicas podem ser utilizadas para prever a disponibilidade de CPU, como:

- Modelos estatísticos: Essas ferramentas analisam grandes conjuntos de dados e identificam padrões e tendências. Por exemplo, modelos de regressão podem prever a disponibilidade de CPU com base em variáveis como carga de trabalho e uso da CPU.
- Simulação: A simulação consiste em criar um modelo do sistema e simular seu comportamento em diferentes cenários de carga de trabalho e recursos de hardware.
- Aprendizado de máquina: Algoritmos de aprendizado de máquina podem ser treinados com conjuntos de dados históricos para prever o comportamento futuro do sistema. Eles são capazes de prever a disponibilidade de CPU com base em variáveis como carga de trabalho e uso da CPU.

Essas técnicas oferecem diferentes vantagens e desvantagens, e a escolha do método adequado depende do contexto específico do sistema.

A disponibilidade de CPU é um indicador importante da performance de um sistema distribuído, medindo o tempo em que os recursos computacionais estão disponíveis para tarefas (Bragg, 2018). Garantir alta disponibilidade aumenta a eficiência do sistema e reduz o tempo de resposta das tarefas dependentes dele. Alguns fatores que influenciam a disponibilidade de CPU em sistemas distribuídos são:

- Carga de trabalho: Quanto maior a carga de trabalho, mais recursos computacionais são demandados, afetando a disponibilidade de CPU.
- Uso da CPU: O uso excessivo da CPU por um componente pode reduzir a disponibilidade de recursos para outros componentes.
- Recursos de hardware: A quantidade de recursos disponíveis, como CPU, memória e armazenamento, pode influenciar a disponibilidade de CPU. Menos recursos disponíveis podem afetar a disponibilidade.

- **Qualidade da rede:** Problemas de rede, como perda de dados ou atrasos na comunicação entre componentes, podem afetar a disponibilidade de CPU.
- **Manutenção:** A manutenção e reparo de componentes também podem afetar a disponibilidade de CPU. Durante a manutenção, menos recursos podem estar disponíveis para outros componentes.

Medir a disponibilidade de CPU em sistemas distribuídos é essencial para identificar problemas e tomar medidas preventivas. Além das técnicas mencionadas anteriormente, também é importante considerar fatores externos, como qualidade da rede, atualizações de sistemas operacionais e manutenção de hardware. Esses fatores podem ser monitorados e gerenciados para manter a disponibilidade de CPU em níveis elevados.

2.2 CONTROLE ESTATÍSTICO DE PROCESSO

Os sistemas distribuídos são amplamente utilizados em ambientes de computação de alto desempenho, mas sua disponibilidade é um fator crítico para garantir a qualidade do serviço. Para avaliar a disponibilidade de CPU em sistemas distribuídos, pode-se utilizar a metodologia de Controle Estatístico de Processo (CEP), que explora a relação entre a carga do sistema, o uso estimado da CPU e os intervalos de medição de disponibilidade. Essa metodologia pode ser aplicada para manter o sistema em um nível de serviço adequado.

Para estimar a disponibilidade de CPU em sistemas distribuídos, é preciso entender suas características e fatores que influenciam a disponibilidade. Modelos estatísticos, simulação e aprendizado de máquina são algumas das técnicas utilizadas para prever a disponibilidade de CPU.

No CEP, as cartas de controle são usadas para monitorar e controlar a disponibilidade de CPU em sistemas distribuídos. Ao implementar essa metodologia, é possível manter a disponibilidade de CPU em níveis aceitáveis, o que é fundamental para garantir a qualidade do serviço em ambientes de computação de alto desempenho.

Em suma, a metodologia de Controle Estatístico de Processo pode ser uma ferramenta valiosa para garantir a disponibilidade de CPU em sistemas distribuídos. Ao entender as características e fatores que influenciam a disponibilidade, e ao utilizar técnicas como modelagem estatística e aprendizado de máquina, é possível manter o sistema em um nível de serviço adequado. O uso das cartas de controle no CEP pode auxiliar na monitoração e controle da disponibilidade de CPU, garantindo a qualidade do serviço em ambientes de computação de alto desempenho.

2.2.1 CARTA DE CONTROLE

A carta de controle é uma ferramenta importante no CEP, pois permite visualizar gráficos do comportamento de uma ou mais variáveis de interesse e identificar rapidamente se o processo está operando de forma estável e dentro dos limites estabelecidos. Ela é baseada no princípio de que, em um processo estável, a maioria dos resultados devem estar dentro de um determinado intervalo de tolerância. Se os resultados estiverem fora desse intervalo, pode haver indícios de que o processo não está operando de forma adequada e precisa ser investigado.

Existem vários tipos de cartas de controle que podem ser utilizadas para monitorar e controlar a qualidade de um processo. As cartas de controle de atributos são adequadas para dados binários, ou seja, dados que podem ser classificados como presente ou ausente, como os defeitos em um produto. As cartas de controle de variáveis são adequadas para dados contínuos, como as dimensões de um produto ou a temperatura de um processo. As cartas de controle multivariadas são adequadas para monitorar várias variáveis de interesse ao mesmo tempo, como a disponibilidade de CPU em conjunto com a carga de trabalho e os recursos de hardware.

As cartas de controle são criadas comparando o desempenho de um processo com um limite de controle. Os resultados são plotados em um gráfico de acordo com padrões aleatórios e dentro dos limites de especificação. Quando o processo está sob controle, todos os pontos no gráfico estão dentro dos limites de controle estabelecidos. Quando o processo está fora de controle, alguns pontos podem ultrapassar os limites de controle ou não estarem distribuídos de forma aleatória no gráfico. Nesses casos, é necessário investigar o que está causando o comportamento anormal do processo e tomar medidas para corrigir o problema.

As cartas de controle são uma ferramenta importante no controle estatístico de processos (CEP) porque permitem identificar a ocorrência de causas especiais e propor possíveis soluções para melhorar o processo. Se o processo não estiver sob controle, a carta de controle pode ser usada para identificar as causas do problema e tomar medidas para corrigi-lo. Se o processo já estiver sob controle, a carta de controle pode ser usada para monitorar a qualidade do processo de forma contínua e detectar qualquer variação anormal que possa ocorrer.

As cartas de controle são formadas por valores coletados no processo durante um determinado período de tempo, como diariamente, horariamente ou minuto a minuto. Esses valores são plotados em um gráfico e comparados com um limite de controle para verificar se o processo está operando de forma estável e dentro dos

limites de aceitação estabelecidos. Se houver variações anormais no gráfico, é necessário investigar as possíveis causas e tomar medidas para corrigir o problema.

Em resumo, as cartas de controle são uma ferramenta importante para monitorar e controlar a qualidade de um processo, identificando variações anormais e propondo soluções para melhorar o processo e garantir a produção de produtos ou serviços de qualidade consistente.

As cartas de controle são uma das principais e mais conhecidas ferramentas do controle estatístico de processos (CEP) e são usadas para monitorar o desempenho de um processo de forma contínua. Elas permitem detectar a presença de causas atribuíveis, que são variações anormais no processo que podem ser investigadas mais detalhadamente para entender suas causas raízes. Se as causas raízes puderem ser identificadas, é possível sistematicamente eliminar as causas atribuíveis e reduzir a variabilidade indesejada do processo.

Na prática, um responsável planeja e monta um plano de coleta de dados para o processo, definindo a frequência de coleta, o número de observações e como essas observações devem ser escolhidas para medição. Depois de coletar os dados, eles são plotados em uma carta de controle e comparados com um limite de controle para verificar se o processo está operando de forma estável e dentro dos limites de aceitação estabelecidos. Se houver variações anormais no gráfico, é necessário investigar as possíveis causas e tomar medidas para corrigir o problema.

O controle estatístico de processos (CEP) é uma técnica que utiliza dados estatísticos para monitorar e controlar a qualidade de um processo produtivo. O objetivo é garantir que o processo esteja operando de forma estável e dentro dos limites de aceitação estabelecidos, a fim de produzir produtos ou serviços de qualidade consistente.

A carta de controle é uma ferramenta importante no CEP, pois permite visualizar gráficos do comportamento de uma ou mais variáveis de interesse e identificar rapidamente se o processo está operando de forma estável e dentro dos limites estabelecidos. Ela é baseada no princípio de que, em um processo estável, a maioria dos resultados devem estar dentro de um determinado intervalo de tolerância. Se os resultados estiverem fora desse intervalo, pode haver indícios de que o processo não está operando de forma adequada e precisa ser investigado.

Em 1924, Walter A. Shewhart, do Bell Telephone Laboratories, desenvolveu o conceito de carta de controle, que muitas vezes é considerado o início formal do controle estatístico de qualidade. No final da década de 1920, Harold F. Dodge e Harry G. Roming, ambos do Bell Telephone Laboratories, desenvolveram a amostragem de aceitação baseada em estatísticas como uma alternativa à inspeção 100%. Em meados da década de 1930, os métodos estatísticos de controle de

qualidade eram amplamente utilizados na Western Electric, o braço de fabricação do Bell System. Muito influenciado por Walter A. Shewhart, Williams E. Deming desenvolveu os 14 pontos de Deming e as sete doenças mortais da gestão, que é uma estrutura importante para implementar a melhoria da qualidade e produtividade.

Cartas de controle são gráficos utilizados no controle estatístico de processos (CEP) para monitorar e controlar a qualidade de um processo. Elas permitem visualizar de forma gráfica o comportamento de uma ou mais variáveis de interesse e identificar se o processo está operando de forma estável e dentro de determinados limites de aceitação.

Existem vários tipos de cartas de controle, como cartas de controle de atributos, cartas de controle de variáveis e cartas de controle multivariadas. Cada tipo de carta é adequado para um tipo de dado específico e pode ser utilizado para monitorar diferentes aspectos de um processo.

As cartas de controle podem ser utilizadas para monitorar e controlar a disponibilidade de CPU em sistemas distribuídos de várias maneiras. Por exemplo, pode-se criar uma carta de controle de variáveis para monitorar o uso da CPU em diferentes componentes do sistema e identificar se há algum componente que está consumindo mais recursos do que o esperado. Também é possível criar uma carta de controle multivariada para monitorar a disponibilidade de CPU em conjunto com outras variáveis, como carga de trabalho e recursos de hardware, para identificar padrões e tendências.

As Cartas de Controle, são comparações gráficas do desempenho de um processo com um limite de controle, obtendo o resultado distribuindo os pontos em um gráfico de acordo com padrões aleatórios e dentro dos limites de especificação. Em seguida vem a análise quanto ao status do processo sob controle, que é quando o gráfico com os pontos distribuídos aleatoriamente estão todos dentro dos limites pré estabelecidos. Posteriormente, quando o status do processo está fora de controle, ou seja, quando os pontos ultrapassam os limites de controle do gráfico ou não são distribuídos aleatoriamente.

São utilizadas para identificar a ocorrência de causas especiais para aplicar uma possível solução. Se o processo não estiver sob controle, este gráfico é utilizado para propor melhorias. Caso o processo já esteja sob controle, as cartas de controle são utilizadas para a análise presente e futura (GOMES, 2010).

São, ainda, uma forma de monitoramento contínuo da variabilidade do processo. Seu principal objetivo é detectar variações anormais para que suas possíveis causas sejam levantadas e estudadas, alcançando soluções que as eliminem e evitem que voltem a acontecer. As cartas de controle são formadas por valores plotados coletados no processo durante um determinado horário (dia-dia, hora-hora, minuto-minuto).

São as principais e mais conhecidas ferramentas das metodologias de CEP. Aplicados às variáveis de resposta do processo, as cartas de controle fornecem um meio de monitoramento on-line do desempenho do processo. Os alarmes emitidos por gráficos de controle indicam a presença das chamadas causas atribuíveis, que podem ser investigadas mais detalhadamente. Se sua causa raiz puder ser descoberta, as causas atribuíveis podem ser sistematicamente eliminadas, reduzindo assim a variabilidade indesejada do processo (Montgomery, 2012). Nesta pesquisa o interesse não é eliminar a variabilidade do processo, mas sim acompanhá-la.

Na prática, um responsável planeja e monta um plano de coleta de dados e faz medições no processo de acordo com esse plano. Nesse plano, são definidos a frequência de coleta, o número de observações e até mesmo como essas observações devem ser escolhidas para medição.

2.2.2 Classificação das Cartas de controle

Cartas de controle são gráficos usados para estudar como o processo muda ao longo do tempo. Uma carta de controle sempre tem uma linha central para média, uma linha superior para o limite de controle superior e uma linha inferior para o limite de controle inferior. Os limites de controle são $\pm 3\sigma$ da linha central (MONTGOMERY, 2004).

Diferente dos métodos tradicionais, as Cartas de Controles foram criadas para simplificar o processo de mapeamento de informações e de controle de dados. A seleção de qual carta de controle é adequada é muito importante no mapeamento de cartas de controle, caso contrário, acaba-se com limites de controle imprecisos para os dados. Nem todas as cartas de controle são iguais. Diferentes tipos de dados requerem cartas diferentes.

As cartas de controle são ferramentas fundamentais do controle estatístico de processos (CEP) e são usadas para monitorar a qualidade de um processo ao longo do tempo. Existem vários tipos de cartas de controle, cada um adequado para um tipo de dado específico.

As cartas de controle de atributos são adequadas para dados binários, como resultados de testes pass/fail ou sim/não. Elas são usadas para monitorar a proporção de unidades de produto que estão dentro dos limites de aceitação especificados.

As cartas de controle de variáveis são adequadas para dados contínuos, como medidas de tamanho ou peso. Elas são usadas para monitorar a média e o desvio-padrão de um processo e identificar variações anormais no processo.

As cartas de controle multivariadas são adequadas para monitorar várias variáveis de interesse de uma vez. Elas são usadas para identificar padrões e tendências em dados multivariados e para detectar interações entre variáveis.

É importante escolher a carta de controle adequada para os dados que estão sendo monitorados, pois diferentes tipos de cartas são adequados para diferentes tipos de dados. Ao escolher a carta de controle adequada, é possível monitorar o desempenho de um processo de forma eficaz e tomar medidas para melhorar a qualidade do processo.

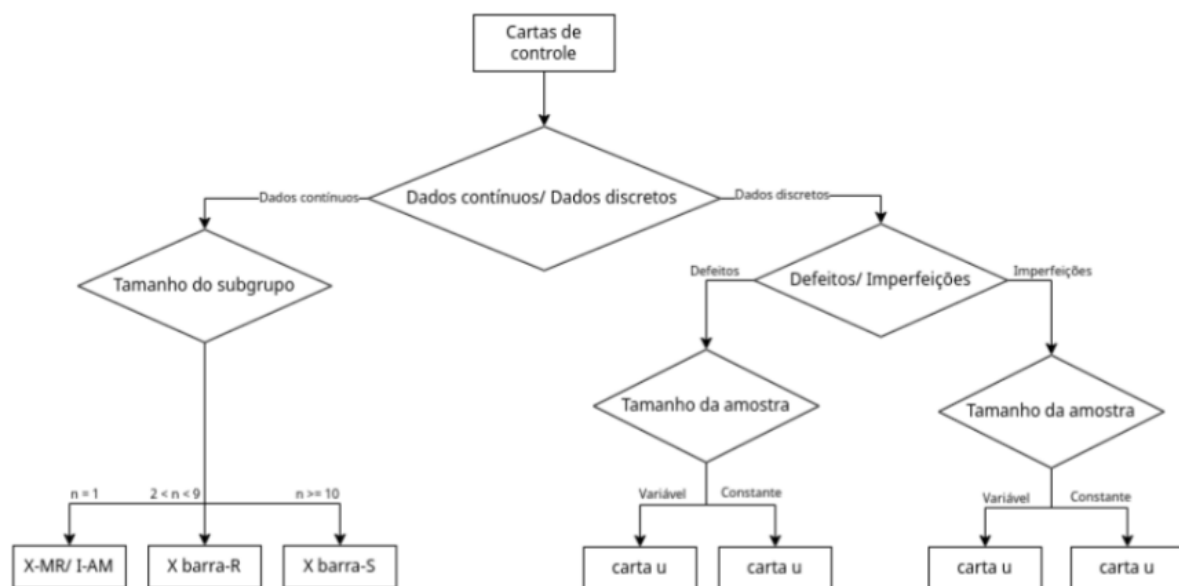


Figura 1: Tipos de carta de controle por tipo de dado

2.2.2.1 Cartas de controle por atributo

As cartas de controle por atributo são ferramentas úteis para monitorar a qualidade de um processo. Elas permitem que você visualize a distribuição dos dados de atributo e detecte padrões ou tendências que possam ser indicativos de problemas no processo. Por exemplo, o número de parafusos que falharam em um chão de fábrica. Ao contrário das cartas por variáveis, apenas um gráfico é plotado para atributos.

Existem quatro tipos de cartas de controle para dados de atributos. A carta p representa a proporção de itens defeituosos e a carta np representa o número de itens defeituosos. A carta u é para o número médio de defeitos por unidade e a carta c é para o número de defeitos.

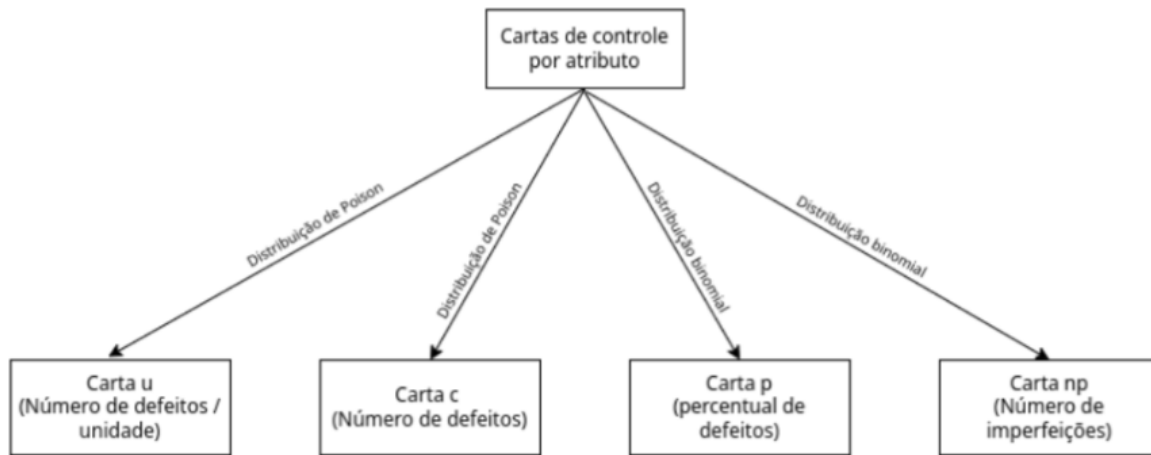


Figura 2: Carta de controle por atributo

Quando há Múltiplos Defeitos: Ela calcula os limites de controle com base na distribuição de Poisson.

Cartas para múltiplos defeitos:

Carta c – carta de controle para defeitos. Geralmente é usada para monitorar o número de defeitos em unidades de tamanho constante;

Carta u - carta de controle para defeitos por unidade. Geralmente é usada para monitorar o tipo de contagem de dados em que o tamanho da amostra é maior que um. Plota o número médio de defeitos por unidade de amostra.

Cartas para defeitos não múltiplos. Esta população classifica os defeitos em 2 pilhas (é binomial). As amostras são boas ou ruins, positivas ou negativas, certas ou erradas.

Carta np – carta de controle para imperfeições. Geralmente é usada para monitorar o número de itens não conformes ou defeituosos no processo de medição.

Carta p - carta de controle para proporções. Geralmente é usada para analisar as proporções de itens não conformes ou defeituosos em um processo.

A carta p é usada para monitorar a proporção de itens defeituosos em um processo. Ela é plotada como uma série de pontos, onde cada ponto representa a proporção

de itens defeituosos em uma amostra. Se os pontos estiverem fora dos limites de controle, isso pode indicar que o processo está produzindo uma proporção excessiva de itens defeituosos.

A carta np é semelhante à carta p, mas é usada para monitorar o número de itens defeituosos em vez da proporção. Novamente, a carta é plotada como uma série de pontos, onde cada ponto representa o número de itens defeituosos em uma amostra. Se os pontos estiverem fora dos limites de controle, isso pode indicar que o processo está produzindo um número excessivo de itens defeituosos.

A carta u é usada para monitorar o número médio de defeitos por unidade no processo. Ela é plotada como uma série de pontos, onde cada ponto representa o número médio de defeitos por unidade em uma amostra. Se os pontos estiverem fora dos limites de controle, isso pode indicar que o processo está produzindo uma quantidade excessiva de defeitos por unidade.

Finalmente, a carta c é usada para monitorar o número total de defeitos no processo. Ela é plotada como uma série de pontos, onde cada ponto representa o número total de defeitos em uma amostra. Se os pontos estiverem fora dos limites de controle, isso pode indicar que o processo está produzindo um número excessivo de defeitos.

Em resumo, as cartas de controle por atributo são ferramentas úteis para monitorar a qualidade de um processo e detectar problemas no mesmo. Elas permitem que você visualize a distribuição dos dados de atributo e detecte padrões ou tendências que possam ser indicativos de problemas no processo.

2.2.2.2 Cartas de controle por variável

As cartas de controle por variável são usadas para monitorar o desempenho de um processo em uma escala contínua, ou seja, quando os dados são medidos em uma escala numérica. Elas permitem avaliar as características de qualidade de um produto e identificar tendências ou mudanças no processo. Além disso, as cartas de controle por variável também permitem verificar a aleatoriedade dos dados no processo.

Existem três tipos principais de cartas de controle por variável: gráfico X-MR ou I-AM, gráfico X barra-R e gráfico X barra-S. O gráfico X-MR ou I-AM é usado quando o tamanho da amostra (n) é igual a 1 e os dados

são coletados de forma contínua, sem subgrupos. O gráfico X barra-R é usado quando o tamanho da amostra está entre 2 e 9 e os dados são coletados em subgrupos. O gráfico X barra-S é usado quando o tamanho da amostra é igual ou superior a 10 e os dados são coletados em subgrupos.

Para construir uma carta de controle por variável, é preciso determinar se a população dos dados é normal ou não e o tamanho da amostra (n) que está sendo mapeada. Além disso, é importante coletar os dados em períodos de tempo definidos e calcular a média e a variabilidade do processo para plotar os gráficos. As cartas de controle por variável são ferramentas úteis para monitorar o desempenho de um processo e detectar problemas no mesmo. Elas permitem avaliar a qualidade do produto e identificar tendências ou mudanças no processo, além de verificar a aleatoriedade dos dados.

Observa a saída em uma escala contínua. É possível observar as características de qualidade de um produto. É preciso determinar se sua população é normal ou não e o tamanho da amostra (n) que se está mapeando.

É um gráfico básico que exibe valores de dados em uma ordem de tempo. Pode ser útil para identificar tendências ou mudanças no processo, mas também permite testar a aleatoriedade no processo.

- Para subgrupos de $n = 1$, usa-se: gráfico X-MR ou I-AM:. Um gráfico de amplitude móvel individual (I-AM) é usado quando os dados são contínuos e não são coletados em subgrupos. Em outras palavras, é coletado uma única observação de cada vez. Um gráfico I-AM fornece a variação do processo ao longo do tempo no método gráfico.
- Se $2 < n < 9$, usa-se o gráfico X barra-R. Ele é usado para monitorar o desempenho do processo de dados contínuos e os dados a serem coletados em subgrupos em períodos de tempo definidos. Como n é pequeno, usa-se o intervalo para estimar a variação do processo. Consiste em dois gráficos para monitorar a média do processo e a variação do processo ao longo do tempo.
- Se $n \geq 10$, use o gráfico X barra-S. É frequentemente usado para examinar a média e o desvio padrão do processo ao longo do tempo. Usa-se o gráfico X barra-S quando os subgrupos tiverem um tamanho de amostra grande e também o gráfico S fornece uma melhor compreensão da dispersão dos dados do subgrupo do que o intervalo.

2.2.3 Análise de carta de controle

O principal objetivo das cartas de controle é melhorar o processo. Elas dão “Voz do processo”, o que ajuda as equipes a identificar causas especiais de variação no processo.

Fora de controle: Um processo está “fora de controle”, significa que existem causas especiais no gráfico de média ou no gráfico de amplitude, ou em ambos. As equipes devem identificar causas especiais e tentar eliminá-las para alcançar um processo estável. Pode-se dizer que o processo está fora de controle se algum ponto no gráfico de controle estiver fora dos limites de controle ou apresentar padrões anormais de variabilidade.

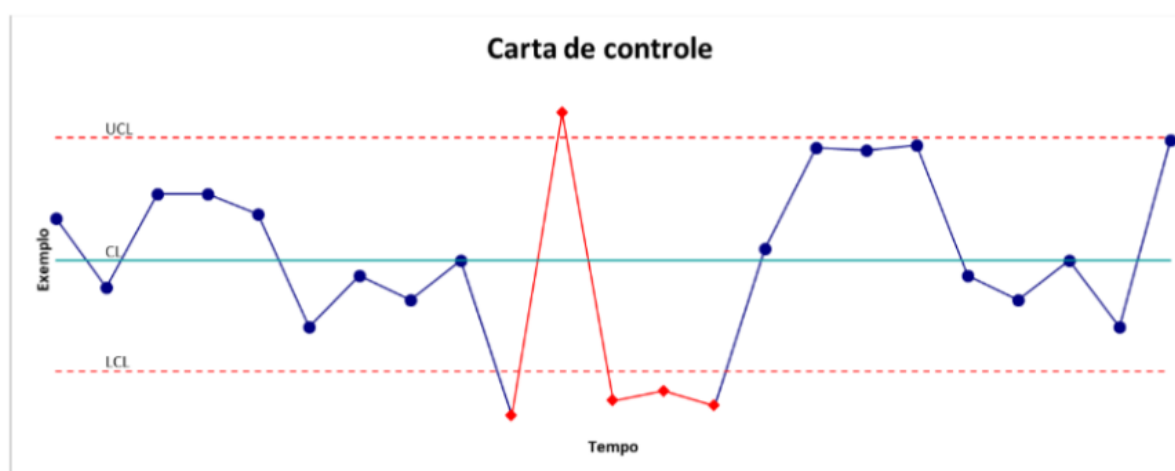


Figura 3: Exemplo de uma carta de processo fora de controle.

Os limites de controle são 3 desvios padrão acima e abaixo da média. Se o processo estiver sob controle, 99,73% das médias ficaram dentro desses limites. O mesmo vale para os limites de controle de amplitude. Porque existem dois componentes para cada gráfico de controle - a média e a amplitude.

Quatro condições possíveis podem ocorrer em qualquer processo.

	componente	estatus processo
1	média - fora de controle amplitude - sob controle	fora de controle
2	média - sob controle amplitude - fora de controle	fora de controle
3	média - fora de controle	fora de controle

	amplitude - fora de controle	
4	média - sob controle amplitude - sob controle	sob controle

Cinco são as regras comuns para interpretação de cartas de controle (ISO 8258, 1991). A figura X demonstra um gráfico padrão para análise dos limites de controle, onde LCS é o limite de controle superior, LCI é o limite de controle inferior e a média é o limite central.

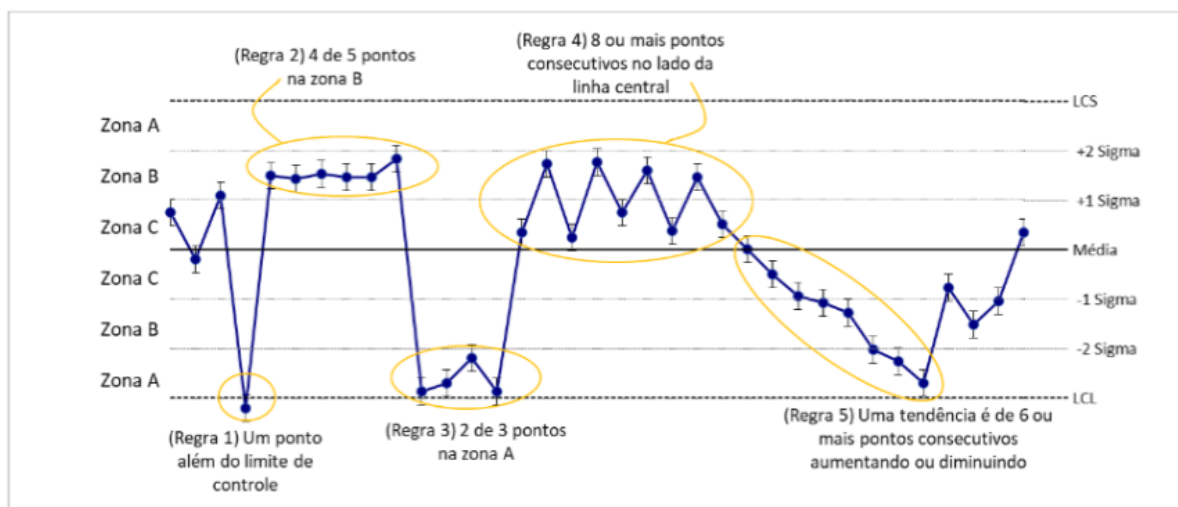


Figura 4: As cinco principais regras da análise de cartas de controle.

As regras em detalhe são:

- Além dos limites (Um ou mais pontos além dos limites de controle)
- Zona A (2 de 3 pontos consecutivos na Zona A ou além)
- Zona B (2 de 3 pontos consecutivos na Zona A ou além)
- Zona C (8 ou mais pontos consecutivos em um lado da média (na Zona C ou além))
- Tendência (7 pontos consecutivos subindo ou descendo)
- Mistura (8 pontos consecutivos sem pontos na Zona C)
- Estratificação (15 pontos consecutivos na Zona C)
- Over-control (14 consecutive points alternating up and down)

2.2.4 Carta de Controle I-AM (Individual com Amplitudes Móveis)

A Carta I-AM é uma ferramenta utilizada para monitorar o desempenho de um processo de produção que produz dados contínuos, ou seja, observações individuais sem a formação de subgrupos. O objetivo da carta I-AM é monitorar a estabilidade do processo ao longo do tempo, para que sejam identificadas e corrigidas eventuais instabilidades (MONTGOMERY, 2004).

A Carta I-AM é composta por dois gráficos separados: o gráfico Indivíduos (I) e o gráfico de Amplitude Móvel (AM). A combinação desses gráficos fornece uma visão completa do comportamento do processo. O gráfico Indivíduos exibe os pontos de dados individuais e monitora a média e os desvios do processo quando os pontos de dados são coletados em intervalos regulares de tempo. Já o gráfico de Amplitude Móvel monitora a variação do processo quando os pontos de dados são coletados em intervalos regulares de tempo, ou seja, rastreia a diferença absoluta de cada medição em relação à medição anterior.

A Carta I-AM pode ser utilizada para:

- Verificar quaisquer indicações de causas atribuíveis no processo que possam levar a um processo fora de controle;
- Monitorar o desempenho do processo antes e depois da implementação de melhorias de processo;
- Situações em que a taxa de produção é baixa ou os sistemas de medição são caros;
- Monitorar o processo em lote.

Para monitorar efetivamente um processo, os valores iniciais são usados para avaliar a média e o desvio padrão do processo. Com base nessas informações, são estabelecidos os limites de controle para Indivíduos e Amplitude Móvel. Durante esta fase, todos os pontos que estão saindo do controle são monitorados e a causa especial da variação é identificada e tenta-se eliminá-la para tornar o processo sob controle. Quando o processo estiver sob controle e os limites de controle para os gráficos individuais e de Amplitude Móvel estiverem em vigor, esses limites de controle serão utilizados e os pontos de dados serão coletados em intervalos regulares de tempo para monitorar a variação do processo. Além disso, é importante lembrar que a carta I-AM deve ser utilizada como uma ferramenta de análise e não apenas como um gráfico de referência.

É preciso avaliar as causas das variações e tomar medidas corretivas para garantir que o processo esteja sob controle e operando de maneira eficiente. Além disso, é importante estabelecer limites de controle para os gráficos individuais e de Amplitude Móvel para monitorar a estabilidade do processo e identificar quaisquer instabilidades no mesmo. Isso permitirá tomar medidas corretivas rapidamente e evitar problemas futuros no processo. A Carta I-AM é uma ferramenta útil para garantir que o processo de produção esteja operando de maneira eficiente e estável ao longo do tempo.

2.4 CONCLUSÃO

Nesta seção, foi apresentada a metodologia de controle estatístico de processo (CEP) para estimar a disponibilidade de CPU em sistemas distribuídos. Foi discutido como essa metodologia explora a conexão entre carga do sistema, uso estimado de CPU e intervalos de medição de disponibilidade, e como ela pode ser utilizada para manter o sistema distribuído em um nível de serviço adequado.

Também foi revisado o conceito de sistemas distribuídos e suas principais características, bem como a disponibilidade de CPU e os principais fatores que a influenciam. Além disso, foram apresentadas as técnicas e métodos utilizados para prever a disponibilidade de CPU, incluindo modelos estatísticos, simulação e aprendizado de máquina.

Por fim, foi discutido o uso das cartas de controle no CEP e como elas podem ser utilizadas para monitorar e controlar a disponibilidade de CPU em sistemas distribuídos.

4 ESTADO DA ARTE

A disponibilidade de CPU em sistemas distribuídos é um fator crucial para garantir o bom funcionamento desses sistemas. O Controle Estatístico de Processo (CEP) é uma técnica amplamente utilizada para monitorar e prever a disponibilidade de recursos em sistemas industriais. Nos últimos anos, tem havido um aumento no uso de CEP para previsão de disponibilidade de CPU em sistemas distribuídos. No entanto, ainda existem desafios e lacunas de conhecimento no uso de CEP para essa finalidade.

Neste contexto, o objetivo desta seção é apresentar o estado da arte da previsão de disponibilidade de CPU em sistemas distribuídos usando CEP. Os objetivos específicos são (1) identificar as principais técnicas de CEP utilizadas para essa finalidade, (2) avaliar sua eficácia e precisão, (3) bem como examinar as técnicas de acurácia utilizadas e o (4) intervalo de coleta de dados e tamanho das janelas. Através da revisão da literatura, serão identificadas as principais contribuições e lacunas de conhecimento no uso de CEP para previsão de disponibilidade de CPU em sistemas distribuídos.

4.1 Introdução

4.2 Metodologia

4.3 Resultados

4.4 Conclusão

A pergunta principal é: "é viável utilizar cartas de controle para prever a disponibilidade de CPU em sistemas distribuídos?" e o seu objetivo: "investigar a viabilidade de utilizar cartas de controle para prever a disponibilidade de CPU em sistemas distribuídos, comparando os resultados obtidos com os de outras técnicas de previsão".

1 Definir o tema da investigação

Previsão De Disponibilidade De CPU Em Sistemas Distribuídos Utilizando Controle Estatístico De Processo

2 Escolha das palavras chave

control chart, forecast, computational resources

3 Formular a questão problema

4 Definir as bases de dados mais relevantes

ACM

Wiley

IEEE

ACS

SciELO

Science Direct

SCOPUS

Springer

Google Scholar

5 Definir parâmetros de seleção dos materiais

1. Inclusão de trabalhos que apresentassem em seu título, palavras - chave e resumo, relação com os principais métodos de previsão (preferencialmente que utilizem controle estatístico de processo);
2. Classificação dos trabalhos validados através do critério anterior em três grupos para a próxima análise, os quais possuem muita relação com o tema, razoável relação com o tema e pouca ou nenhuma relação;
3. Exclusão de resultados duplicados..
4. Exclusão de trabalhos com pouca ou nenhuma relação com o tema.

6 Sintetizar os materiais encontrados

7 Fazer revisão literária

Esta seção apresenta o estado da arte da previsão de disponibilidade de CPU em sistemas distribuídos utilizando CEP. Com isto em mente, deseja-se investigar de

que forma técnicas de CEP vêm sendo utilizadas para previsibilidade de disponibilidade de recursos computacionais. Assim, objetiva-se averiguar as principais estratégias adotadas para previsibilidade de disponibilidade desses recursos, valendo-se do CEP. Para colaborar com o alcance deste, são apresentados os seguintes objetivos específicos: identificar as principais técnicas de CEP utilizadas para previsão de disponibilidade de CPU em sistemas distribuídos, bem como avaliar sua eficácia e precisão, quais as técnicas de acurácia estão sendo utilizadas e qual o intervalo de coleta de dados e tamanho das janelas.

A seção apresenta as técnicas utilizadas para acompanhamento e previsão de disponibilidade de CPU e uma série de técnicas que são úteis à esse fim. Essas técnicas incluem técnicas utilizando inteligência artificial, ferramentas estatísticas. Além disso, o artigo discute diferentes tipos de sistemas distribuídos como cliente-servidor e middleware (multicamadas), e mostra como a previsibilidade funciona nesses sistemas.

4.1 Desenvolvimento

A metodologia de pesquisa escolhida para esta pesquisa foi uma revisão sistemática, por se tratar de uma forma eficaz para identificar, avaliar, interpretar e comparar os estudos disponíveis e relevantes para uma determinada questão (Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273: 408-412.). Este processo pode levar a respostas detalhadas dentro de um escopo específico. Nesse sentido, foram desenvolvidas questões de pesquisa para abordar as principais preocupações relacionadas à previsibilidade de disponibilidade de recursos em sistemas distribuídos, incluindo aspectos de arquitetura, escalabilidade e monitoramento.



Figura 5: Fluxo da revisão

4.1.1 Questões de pesquisa

A seguir, são apresentadas as 6 questões que dão os encaminhamentos das buscas e que trarão os dados necessários para a construção dos resultados deste levantamento.

1. P1: Qual o número de citações de cada trabalho?
2. P2: Quais foram os métodos baseados em cartas de controle?
3. P3: Quais métricas de acurácia foram utilizadas?
4. P4: Que recurso se buscava prever?
5. P5: Qual a quantidade de publicações realizadas sobre o tema a cada ano?
6. P6: Em qual área foi aplicada a previsão?

Para responder a todas essas perguntas, é feita uma extensa revisão de literatura. Sendo coletado informações de artigos, relatórios técnicos, artigos de periódicos e livros de vários repositórios, como a biblioteca ACM Digital, ScienceDirect, IEEE Xplore, SCOPUS, SciELO, Springer, Wiley e Google Scholar. Sendo essa busca referente ao período de 2016 a 2022.

4.1.2 Escolha das bases de dados

Esta seção apresenta os resultados da busca de artigos publicados entre 2016 e 2022. Foi utilizado um conjunto de oito bases de dados diferentes, com uma frase comum pesquisada em diversos intervalos, porém cada uma com recursos de busca específicos. A configuração do banco de dados na tabela \ref{tab:baseDeDados} indica as opções de busca que foram utilizadas por este trabalho. É importante ressaltar que as opções de busca variam de acordo com cada base de dados, porém as opções escolhidas foram feitas para abranger o maior número de trabalhos publicados em suas respectivas fontes.

Repositório	Endereço eletrônico	Configuração de Busca
Google Scholar	https://scholar.google.com.br	“Em qualquer lugar do artigo”
ACM Digital	https://dl.acm.org	“All” Opção “The ACM Guide for Computer Literature”

ScienceDirect	https://www.sciencedirect.com	"Find articles with these terms"
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/	"Full Text Only"
SCOPUS	https://www.scopus.com/	"All fields"
Springer	https://www.springer.com/	"Title" Opção "Computer Science"
Wley	https://onlinelibrary.wiley.com	"Title"

4.1.3 Definição e realização das buscas

Utilizando-se das palavras-chave que permeiam a pesquisa e utilizando o operador lógico AND, foi definida a seguinte string de busca: Control chart AND Forecast AND Computational Resources.

Encontraram-se no total 239 documentos relacionados com o tema de pesquisa. Esses textos podem ser agrupados com base no seu respectivo repositório conforme mostra o quadro "X".

Googl e Schol ar	ACM Digita l	ScienceDir ect	IEEE Xplore	SCOPU S	Springer	Wiley	Total
163	12	28	11	16	1	8	239

Através do quadro X observa-se que o repositório Springer foi o que retornou o menor número de publicações realizadas com ao menos uma das palavras utilizadas na string de busca, por outro lado a base do Google Scholar encontrou um maior número de trabalhos relacionados. Ao considerar que alguns estudos podem não ter um nível mínimo de relação com seus temas de pesquisa, é fundamental aplicar filtros sobre os resultados atuais.

No processo de filtragem são identificadas as obras de maior impacto, com o objetivo principal de observar as obras que melhor respondem aos resultados já alcançados. Neste trabalho são realizadas as três etapas de filtragem: abstração e resumo, seleção e avaliação.

Na verificação inicial dos repositórios, foram descartados trabalhos duplicados, encontrados em mais de um repositório. Durante a primeira etapa foram identificadas 26 publicações duplicadas. Em seguida uma nova varredura foi realizada a fim de identificar prioritariamente a relação com o tema de pesquisa explorado, contendo os critérios abaixo:

1. Inclusão de trabalhos que apresentassem em seu título, palavras-chave e resumo, relação com os principais métodos de previsão através de cartas de controle.
2. Classificação dos trabalhos validados através do critério anterior em três grupos para a próxima análise, os quais possuem muita relação com o tema, razoável relação com o tema e pouca ou nenhuma relação;
3. Exclusão de trabalhos com pouca ou nenhuma relação com o tema.

	Google Scholar	ACM Digital	ScienceDirect	IEEE Xplore	SCOPUS	Springer	Wiley	Total
Grupo 1 – Muito relacionado								
	2	1	1	1	1	0	1	7
Grupo 2 – Razoavelmente relacionado								
	8	2	1	2	4	1	1	19
Grupo 3 – Pouco ou não relacionado								
	153	9	26	8	11	0	6	213

4.1.4 Apresentação geral dos resultados

Foram encontrados 239 documentos relacionados com o tema de pesquisa. Agrupados com base em seus repositórios.

Finalmente, após a leitura de pesquisas anteriores, os dados obtidos foram analisados a fim de extrair mais informações sobre o nível de similaridade ou divergência com o tema central de pesquisa. Foram excluídos dos resultados aqueles com pouca ou nenhuma relação com o tema da pesquisa, que são apresentados na Tabela “X”.

	Google Scholar	ACM Digital	ScienceDirect	IEEE Xplore	SCOPUS	Springer	Wiley	Total
Grupo 3 – Pouco ou não relacionado								
	153	9	26	8	11	0	6	213

Um total de 7 artigos foram obtidos para uso na solução das questões previamente definidas e ao longo do desenvolvimento da pesquisa.

A seleção final dos trabalhos após a filtragem é mostrada na Tabela “x”, onde para cada documento foi definido um código identificador (Sn, onde S se refere ao documento e n ao seu número sequencial), que será utilizado ao longo desta pesquisa como referência para os respectivos documentos. A segunda coluna contém informações sobre o(s) autor(es), enquanto a terceira coluna fornece o título do trabalho e a quarta, a indicação do ano de publicação.

Cod.	Autor(es)	Titulos	Ano publicação
s1	Fattah, Sheik Mohammad Mostakim; Bouguettaya, Athman;	Event-based detection of changes in IaaS performance signatures	2020
s2	Raghu Ramakrishnan, Arvinder Kaur	Technique for Detecting Early-Warning Signals of Performance Deterioration in Large Scale Software Systems	2017
s3	Amitava MitraKang Bok LeeSubhabrata Chakraborti	An adaptive exponentially weighted moving average-type control chart to monitor the process mean	2019
s4	Xiao-Li Jin; Jian-Bo Song; Jin-Xin Peng; Xiao-Peng Pan; Rui Guo; Xiao-Fen Xing	Study on the established customized limits for the daily quality assurance procedure	2021

s5	Esfahani, A.A., Ershadi, M.J., Azizi, A.	Monitoring indicators of research data using I-MR control charts	2020
s6	Altherwi, A., Hijry, H.	Implementation of a statistical process control system for a selected industrial plant	2016
s7	Nasrullah Khan, Muhammad Aslam, Syed Muhammad Muslim Raza, Chi-Hyuck Jun	A new variable control chart under failure-censored reliability tests for Weibull distribution	2018

4.1.5 Resolução das Questões

Esta seção pretende ser um resumo dos principais tópicos discutidos no corpo principal deste documento. Ela tenta destacar as questões-chave apontadas durante o estudo de cada trabalho selecionado, por meio de fichas de leitura.

Pergunta 1: Qual o número de citações de cada trabalho?

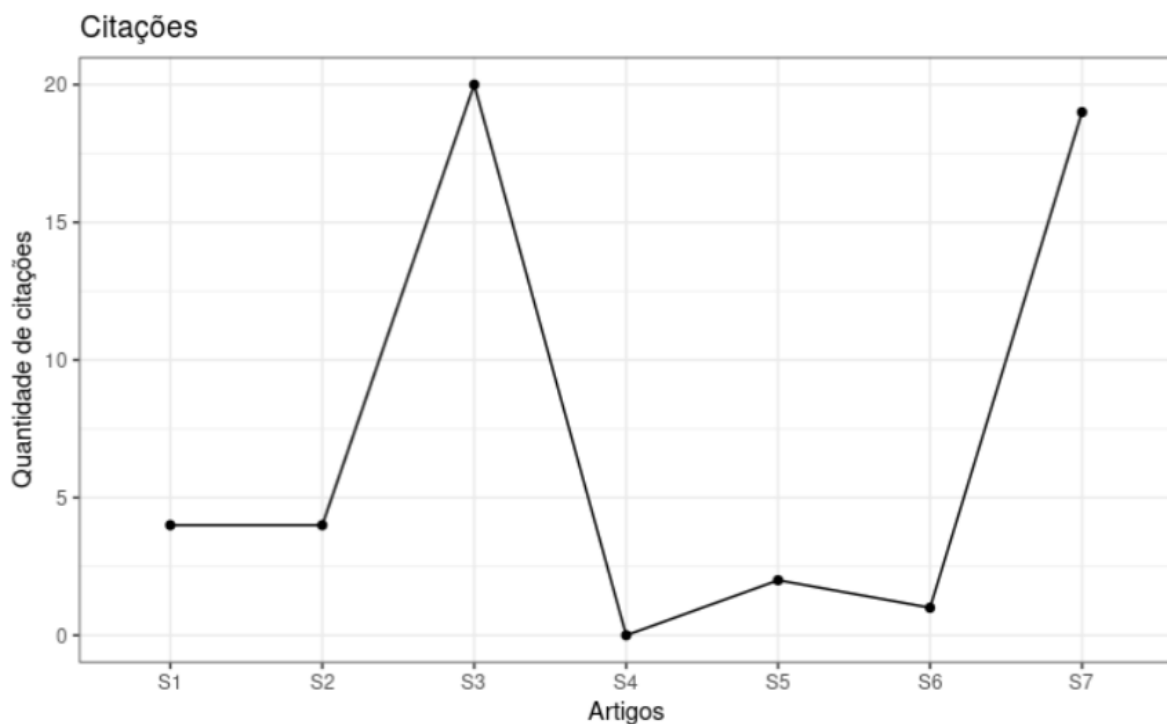


Figura 6: Citações por artigo

A figura \ref{fig:citacoes} indica que o trabalho S3 tem o maior número de citações. Dos 7 trabalhos, apenas 1 não tem citações até o momento, o que representa cerca de 14% dos trabalhos estudados, dos 7 trabalhos apenas 2 foram publicados antes de 2018. Isto ilustra que a maioria dos trabalhos técnicos está citando trabalhos recentes dentro de um campo e serve como uma ferramenta útil para identificar outros trabalhos que possam estar relacionados com o tema atual.

2. P2: Quais foram os métodos baseados em cartas de controle?

Métodos baseados em cartas de controle	
CUSUM	1
XmR	2
EWMA	1
I-MR	2
Shewhart	1

Os resultados obtidos mostraram que o método mais utilizado é a carta de controle da média móvel, ou I-MR ou X-MR. Ambas se tratam da mesma metodologia, sem que alguns trabalhos utilizem nomenclaturas distintas. Os resultados também mostram que esse tipo de carta concentra 57% dos resultados apresentados na tabela. De acordo com o trabalho de (CITAR), estes tipos de cartas estão entre as mais populares dentro da classe de cartas por variáveis, mostrando um bom desempenho em modelos de dados contínuos e não coletados em subgrupos. Pode-se observar que este tipo de gráfico de controle procura observar as características de qualidade de um produto ou processo. Eles são normalmente usados para identificar mudanças no processo e testar a aleatoriedade do mesmo.

Para o problema da pesquisa de detecção baseada em eventos de alterações nas assinaturas de desempenho de IaaS (Fattah, Sheik Mohammad Mostakim; Bouguettaya, Athman, 2020), o método gráfico CUSUM forneceu resultados significativos e também foi expressivo. Além disso, também mostrou resultados eficazes quando comparado com outro método baseado em cartas de controle do tipo de média móvel adaptável exponencialmente ponderada com controle dinâmico de processos de engenharia como o problema de pesquisa (Amitava MitraKang Bok LeeSubhabrata Chakraborti, 2019).

É válido citar também que outro método utilizado nos trabalhos relacionados no Quadro é a carta de Shewhart. Esse tipo de carta usa apenas a informação contida no último ponto plotado e ignora qualquer informação dada pela sequência inteira de pontos. Tais características tornam esse tipo de carta insensível a pequenas mudanças no processo.

3. P3: Quais métricas de acurácia foram utilizadas?

<i>Métricas utilizadas para validação</i>	
<i>Distorção de Tempo Dinâmico (DTW)</i>	<i>1</i>
<i>Precisão e revocação (Precision and recall)</i>	<i>1</i>
<i>Adverse events (AE)</i>	<i>3</i>
<i>Mean Absolute Error (MAE)</i>	<i>1</i>
<i>smallest extreme value (SEV)</i>	<i>1</i>

Na tabela \ref{tab:metricasvalidacao}, é possível ver as métricas utilizadas nos trabalhos selecionados. A métrica mais comumente usada tem sido a precisão e recall (S2, S3), que é a medida de quão bem um classificador é capaz de distinguir entre duas classes.

4. P4: Que recurso se buscava prever?

<i>Que recurso se buscava prever?</i>
<i>Infraestrutura como serviço</i>
<i>Sistemas de software de grande escala</i>
<i>Controle dinâmico de processos de engenharia</i>
<i>Garantia de qualidade diária</i>
<i>Sistemas de informação de pesquisa</i>
<i>Processos nas indústrias de manufatura</i>
<i>Média e a variância de um processo Weibull</i>

Cada um dos trabalho selecionados trabalhou em um recurso diferente, que foram:

- *Infraestrutura como serviço*
- *Sistemas de software de grande escala*
- *Controle dinâmico de processos de engenharia*
- *Garantia de qualidade diária*
- *Sistemas de informação de pesquisa*
- *Processos nas indústrias de manufatura*
- *Média e a variância de um processo Weibull*

De acordo com a análise, há uma grande quantidade de esforço dedicada à pesquisa de recursos de software em larga escala. Isso é principalmente para atender às necessidades crescentes de infraestrutura como serviço, que tem sido

impulsionada pela demanda crescente por serviços remotos, tanto no ambiente de trabalho quanto no nível do consumidor final através de dispositivos móveis e inteligentes.

5. P5: Qual a quantidade de publicações realizadas sobre o tema a cada ano?

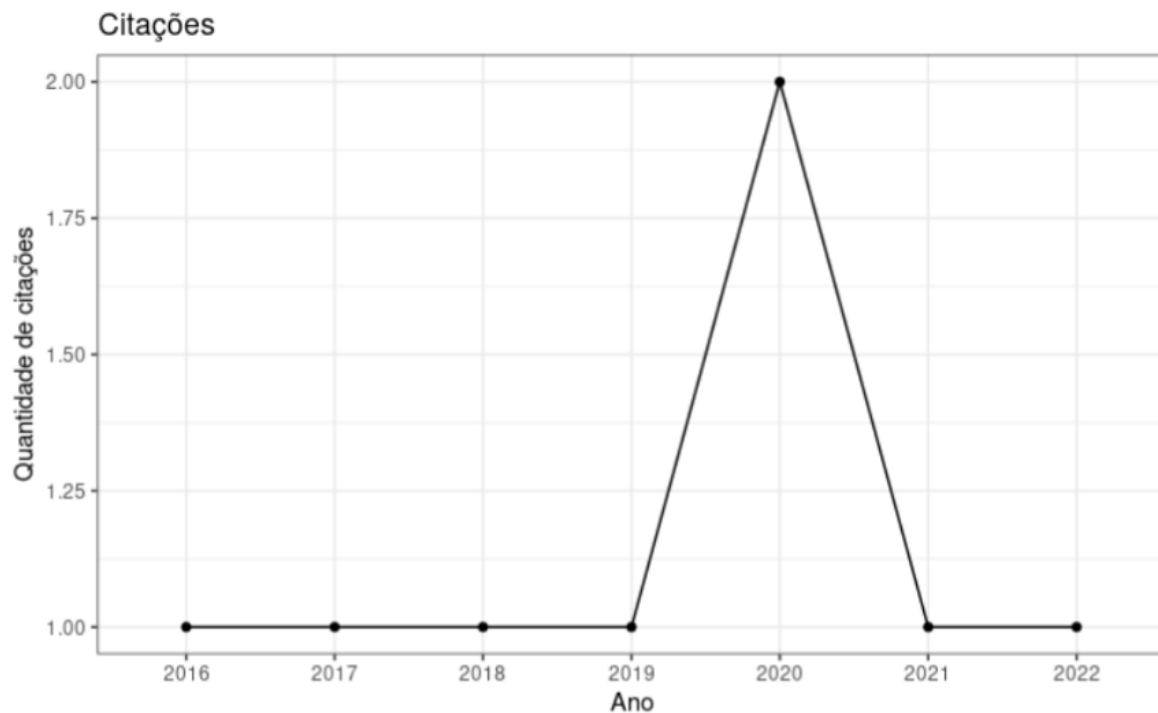


Figura 7: Citações por ano

Uma quantidade significativa de pesquisas sobre o tema foi desenvolvida nos últimos anos, com um pico em 2020. Ao publicar estes materiais e articular o problema, está sendo contribuindo para o crescente pool de conhecimentos relacionados a este tópico.

6. P6: Em qual área foi aplicada a previsão?

Em qual área foi aplicada a previsão?	
Cloud Computing	1
Desenvolvimento de softwares	1
Processo fabril	3
Atendimento hospitalar	1

Novamente, constata-se a importância da área de Computação, porém vale ressaltar a diversidade de pesquisas que aderem ao uso de cartas de controle como método de previsão. Na Tabela foi feita esta separação, pois o artigo S1 é bem específico ao tratar de cloud computing.

4.1.6 Trabalhos relacionados

Após a revisão da literatura, que analisou as cartas de controle, a periodicidade dos dados utilizados, um estudo mais profundo dos trabalhos selecionados e as ressalvas dos autores.

Fattah et al. (2020) apresenta bons resultados com a carta CUSUM utilizando janelas de 34 dias. A dificuldade se manifesta durante as replicações de cargas de trabalho do mundo real para períodos muito longos, o que levou à seleção da partição dos dados em 360 partes e considerando cada partição como uma carga de trabalho média do dia para criar dados de carga de trabalho de 1 ano. Esta estratégia também levou a uma alta exigência de acumulação de dados, o que elevou o custo computacional do método.

Segundo Raghu et al. (2017), dentre as ferramentas estatísticas de previsão a carta de controle I-MR ou X-MR, como é citada em seu trabalho, é aplicável à maioria das séries temporais da vida real, por sua simplicidade, interatividade, baixo overhead de implementação e precisão. O autor utiliza em sintonia com a carta de controle a Lei de Little para a validação final dos testes. Sendo determinado empiricamente a quantidade de observações e a periodicidade de 270 observações em 9 horas de operação e sua acurácia sendo mensurada por meio da técnica de Precisão e recall.

Em um trabalho recente de Amitava et al. (2019), a carta EWMA é usada para prever séries temporais de falhas em sistemas de grande escala. O modelo foi particularmente bem sucedido quando se tratou de prever a saída usando uma janela de tempo relativamente pequena de cerca de 30 observações.

Xiao-Li Jin et al. (2021) apresentaram uma proposta para personalizar os limiares para procedimentos de garantia de qualidade a curto prazo (diário) e acharam a carta I-MR flexível, adaptável e eficaz.

Esfahani et al. (2020) utilizam o gráfico I-MR para avaliar o desempenho do sistema de informação de pesquisa e descartam o uso de outros métodos estatísticos, como o modelo ARIMA. A abordagem de Esfahani et al. é bastante abrangente porque os

autores revisam os benchmarks para os gráficos de I-MR e os adaptam para apoiar a exploração de cenários específicos (por exemplo, janelas de tempo limitado). Esta abordagem suporta uma visão mais holística do desempenho dos sistemas de informação de pesquisa, levando a uma melhor compreensão e avaliação destes sistemas.

Assim como Amitava et al. (202019), Altherwi et al. (2016) reafirmam a dificuldade em trabalhar com janelas de observação muito grandes. Os autores realizam testes comparativos entre as várias ferramentas de qualidade. Eles consideram medidas de qualidade fora do uso tradicional de "eventos", tais como erros consistentes ou regiões de rejeição. O estudo é limitado à comparação entre ferramentas de qualidade (tais como Histograma e gráfico de Pareto) e não fornece nenhuma visão sobre sua escolha ou atualização dos parâmetros da carta de controle.

A proposta apresentada por Nasrullah Khan et al. (2018) mostra consistência em lidar com mudanças repentinas, e o mesmo segue a abordagem de atualização regular realizada por Raghu et al. (2017). Em conjunto com a carta I-MR os autores utilizam a distribuição de Weibull com vista a determinar o tempo de vida médio e taxa de falhas em função do tempo da janela analisada.

4.1.7 Conclusões em relação à investigação

Com base na literatura encontrada, é possível estabelecer que para o estudo deste tema é importante considerar a periodicidade dos dados utilizados e a efetividade das cartas de controle para cada tipo de processo a ser analisado. Destacando a utilização da carta de controle I-MR em processos de observações individuais.

4.2 Conclusões

Os resultados desta revisão levaram à exploração de uma série de pontos relativos à previsão de recursos através da utilização de cartas de controle. A fase inicial envolveu a seleção de publicações, o que, por sua vez, resultou em um extenso processo de filtragem. O resultado final foi um conjunto de documentos que forneceu o suporte e a interpretação necessários para definir outras questões que seriam examinadas neste estudo.

5 PROPOSTA DE DISSERTAÇÃO

Para prever a disponibilidade de CPU de um determinado computador pessoal, será utilizado o modelo de carta de controle I-AM. Esse modelo é uma técnica de

Controle Estatístico de Processo (CEP) que permite monitorar e controlar um processo ao longo do tempo, identificando possíveis desvios em relação ao comportamento esperado.

A coleta de dados será realizada através da ferramenta sysstat, um conjunto de ferramentas de linha de comando que permite monitorar diversos aspectos do sistema, incluindo a utilização da CPU. Os dados serão coletados em intervalos de 1 segundo e armazenados em um arquivo CSV para posterior análise.

Após a coleta de dados, será realizado o tratamento dos dados para que possam ser analisados. Isso incluirá a limpeza dos dados, removendo valores incorretos ou faltantes, e a normalização dos dados, colocando os valores em uma escala comum. Também será realizada a transformação dos dados, aplicando uma função matemática para alterar a escala ou o formato dos dados de acordo com as necessidades da análise.

Para parametrizar o modelo de previsão de disponibilidade de CPU, será necessário identificar o modelo matemático que melhor representa o processo de utilização da CPU. Isso será feito através da análise dos dados coletados e da aplicação de técnicas estatísticas, como análise de correlação e análise de variação. Em seguida, os parâmetros do modelo matemático selecionado serão estimados através de técnicas de ajuste de curvas, como o método dos mínimos quadrados.

Com o modelo parametrizado, será possível realizar a previsão da disponibilidade de CPU usando a carta de controle I-AM. Isso envolve a plotagem dos dados coletados em uma carta de controle e o cálculo dos limites de controle superior e inferior (LCL e UCL, respectivamente). A previsão será realizada comparando o valor previsto com os limites de controle. Se o valor previsto estiver dentro dos limites de controle, isso indica que o processo está sob controle e a previsão é considerada válida. Caso contrário, será necessário investigar a causa da anormalidade e realizar ajustes no processo para trazer a disponibilidade de CPU de volta ao controle.

Ao final da pesquisa, os resultados obtidos serão comparados com os resultados previstos pelo modelo, permitindo verificar o nível de assertividade do modelo criado. Além disso, os resultados da pesquisa também serão comparados com os resultados de outros trabalhos na área, permitindo identificar possíveis lacunas de conhecimento e contribuir para o avanço do campo de estudo.

É importante destacar que a pesquisa sobre a previsão de disponibilidade de CPU em computadores pessoais é relevante para entender os padrões de uso e otimizar o desempenho dos sistemas. A criação de um modelo de previsão de

disponibilidade de CPU usando cartas de controle e técnicas de CEP pode ser útil para prever a necessidade de atualização de hardware ou para identificar problemas de sobrecarga de CPU.

Para realizar a previsão de disponibilidade de CPU em sistemas distribuídos, uma opção é utilizar cartas de controle. As cartas de controle são uma técnica de análise estatística que permite monitorar a qualidade de um processo ao longo do tempo. Elas podem ser utilizadas para identificar padrões ou tendências que possam afetar a disponibilidade de CPU em um sistema distribuído.

Para utilizar cartas de controle, é preciso coletar dados sobre a disponibilidade de CPU em intervalos regulares, como a cada hora ou a cada dia. Esses dados são plotados em um gráfico, que é dividido em três regiões: uma região de controle, uma região de alerta e uma região de ação. A região de controle é a área onde os dados devem estar, a região de alerta é uma área de margem em torno da região de controle e a região de ação é uma área fora das regiões de controle e alerta.

Se os dados ficam dentro da região de controle, isso indica que o processo está funcionando de forma estável e previsível. Se os dados ficam na região de alerta, isso pode ser um sinal de que algo está afetando a disponibilidade de CPU e é preciso investigar mais a fundo. Se os dados ficam na região de ação, isso indica que há um problema com o processo e é necessário tomar medidas imediatas para corrigi-lo.

Além de plotar os dados em uma carta de controle, também é importante realizar análises estatísticas para determinar a causa raiz de quaisquer desvios da região de controle. Isso pode incluir a realização de testes de hipóteses, análise de correlação e análise de regressão, entre outras técnicas.

A proposta de projeto visa manter o nó gestor de um sistema distribuído atualizado sobre o nível de disponibilidade de CPU das máquinas que compõem este sistema. Isto é conseguido utilizando a ferramenta de controle estatístico de processo, carta de controle individual de amplitude móvel I-AM em que são definidas as linhas superior, central e inferior que são os parâmetros do modelo.

A fim de poder validar este modelo, são utilizadas as ferramentas de métrica de acurácia MAE (Erro Absoluto Médio), MAPE (Erro Percentual Médio Absoluto) e RMSE (Raiz do Erro Quadrático Médio). A exatidão do modelo foi verificada mediante a comparação com os resultados previstos e os obtidos para verificar o nível de assertividade do modelo criado.

5.1 Identificação do modelo, estimativa de parâmetros e previsão

A carta de controle I-AM é um gráfico utilizado para monitorar a qualidade de um processo ao longo do tempo. Ela é composta por um gráfico de dispersão dos dados coletados e por limites de controle superior e inferior (LCL e UCL, respectivamente). A fórmula para o cálculo dos limites de controle é:

$$LCL = \bar{X} - A2 * R$$

$$UCL = \bar{X} + A2 * R$$

Onde:

- \bar{X} é a média dos dados coletados
- A2 é uma constante tabelada que depende do número de observações e da distribuição dos dados
- R é o desvio padrão dos dados coletados

Para gerar a carta de controle I-AM, é necessário seguir os seguintes passos:

Organizar os dados coletados em uma tabela, com uma coluna para os valores de utilização da CPU e outra para o tempo em que os dados foram coletados.

Calcular a média móvel dos dados, utilizando a fórmula:

$$\text{Média móvel} = (\text{dado1} + \text{dado2} + \text{dado3} + \dots + \text{dado}_n) / n$$

Onde "n" é o número de dados considerados na média móvel.

Calcular o desvio padrão dos dados, utilizando a fórmula:

$$\text{Desvio padrão} = \sqrt{((\text{dado1} - \text{média})^2 + (\text{dado2} - \text{média})^2 + \dots + (\text{dado}_n - \text{média})^2) / (n-1)}$$

Onde "média" é a média dos dados e "n" é o número de dados.

Calcular os limites de controle superior (UCL) e inferior (LCL) utilizando as fórmulas:

$$UCL = \text{média móvel} + (3 * \text{desvio padrão})$$

$$LCL = \text{média móvel} - (3 * \text{desvio padrão})$$

Plotar os valores de utilização da CPU em um gráfico de linhas e adicionar os limites de controle calculados.

Analisar o gráfico de controle para identificar possíveis problemas de disponibilidade de CPU. Se os valores estiverem dentro dos limites de controle, o processo está sob controle. Caso contrário, é necessário investigar a causa da anormalidade e realizar ajustes no processo para trazer a disponibilidade de CPU de volta ao controle.

Além disso, é importante realizar a atualização da carta de controle periodicamente, adicionando os novos dados coletados e recalculando as médias móveis, desvios padrões e limites de controle. Dessa forma, é possível acompanhar a evolução do processo de utilização da CPU e detectar problemas de forma rápida e eficiente.

i

$$\bar{X} = \frac{\sum_{i=1}^k x_i}{k}$$

$$\overline{AM} = \frac{\sum_{i=1}^k |x_i - x_{i-1}|}{k - 1}$$

Onde x_i é igual ao valor da observação, i e k o número de subgrupos.

Para as linhas central (LC), de controle inferior (LCI) e superior (LCS) para valores individuais a fórmula é:

$$LCS = \bar{X} + E_2 \overline{AM}$$

$$\text{Linha central} = \bar{X}$$

$$LCI = \bar{X} - E_2 \overline{AM}$$

$$\text{Sendo } E_2 = A_2 \sqrt{n}$$

Para as linhas central, de controle inferior (LCI) e superior (LCS) para amplitude móvel a fórmula é:

$$LCS = D_4 \overline{AM}$$

$$\text{Linha central} = \overline{AM}$$

$$LCI = D_3 \overline{AM}$$

Onde, D_3 , D_4 e E_2 são constantes da carta de controle.

	Tabelas de constantes para gráficos de controle							
	Tabela 8A - Dados Variáveis				ref: manual AIAG para SPC			
	Carta XBarra e Carta R				Carta XBarra e Carta s			
	Carta por médias	Cartas por Amplitude (R)			Cartas por médias	Carta para Desvio(s) Padrão		
	Fator de Limites de Controle	Divisores a serem estimados σ_x	Fatores para de Limites Controle		Fator de Limites de Controle	Divisores a serem estimados σ_x	Fatores para de Limites Controle	
Tamanho do subgrupo (n)	A_2	d_2	D_3	D_4	A_3	C_4	B_3	B_4
2	1.880	1.128	-	3.267	2.659	0.7979	-	3.267
3	1.023	1.693	-	2.574	1.954	0.8862	-	2.568
4	0.729	2.059	-	2.282	1.628	0.9213	-	2.266
5	0.577	2.326	-	2.114	1.427	0.9400	-	2.089
6	0.483	2.534	-	2.004	1.287	0.9515	0.030	1.970
7	0.419	2.704	0.076	1.924	1.182	0.9594	0.118	1.882
8	0.373	2.847	0.136	1.864	1.099	0.9650	0.185	1.815
9	0.337	2.970	0.184	1.816	1.032	0.9693	0.239	1.761
10	0.308	3.078	0.223	1.777	0.975	0.9727	0.284	1.716
15	0.223	3.472	0.347	1.653	0.789	0.9823	0.428	1.572
25	0.153	3.931	0.459	1.541	0.606	0.9896	0.565	1.435

5.1.1 Identificação do modelo

A fim de automatizar o processo de criação e atualização do gráfico de controle requerido no desenvolvimento deste trabalho, foram desenvolvidos scripts responsáveis pela sistematização destas operações. Tudo isso foi implementado na linguagem R, que oferece várias bibliotecas para a análise e manipulação dos gráficos de controle, especialmente por sua simples sintaxe.

A princípio será apresentada a estrutura responsável pela geração do gráfico de controle, tendo como pré-requisito os dados coletados do sistema para obtenção do

nível de processamento, que define a estrutura do arquivo de dados. Estes dados são coletados a cada segundo, salvos em um arquivo texto e enviados a um servidor quando tiver a coleta referente a 5 minutos de operação do equipamento. Não há necessidade de armazenar estes dados por longos períodos de tempo.

Para obter um gráfico de controle, devemos primeiro gerar um conjunto de dados pertencentes ao mesmo computador pessoal que foi formatado de acordo com a estrutura apresentada no quadro X.

interval	# hostname	timestamp	CP U	%user	%nic e	%syste m	%iowait	%stea l	%idle
1	DTI01	2022-09-30 14:24:04 UTC	-1	3	51;2	27;2	27;0	00;91	67
1	DTI01	2022-09-30 14:24:05 UTC	-1	2	00;2	03;4	05;0	00;91	39
1	DTI01	2022-09-30 14:24:06 UTC	-1	2	00;1	50;2	74;0	00;92	77
1	DTI01	2022-09-30 14:24:07 UTC	-1	2	00;1	52;0	00;0	00;96	21
6	DTI01	2022-09-30 14:24:13 UTC	-1	3	46;1	42;1	92;0	00;92	69
1	DTI01	2022-09-30 14:24:14 UTC	-1	2	00;1	26;0	00;0	00;96	23
1	DTI01	2022-09-30 14:24:15 UTC	-1	2	00;1	52;1	01;0	00;95	44
1	DTI01	2022-09-30 14:24:16 UTC	-1	2	00;1	99;14	18;0	00;81	34
1	DTI01	2022-09-30 14:24:17 UTC	-1	2	00;0	76;13	60;0	00;83	12
7	DTI01	2022-09-30 14:24:24 UTC	-1	2	04;1	33;1	90;0	00;94	23
1	DTI01	2022-09-30 14:24:25 UTC	-1	2	25;1	25;0	00;0	00;95	99
1	DTI01	2022-09-30 14:24:26 UTC	-1	3	00;1	01;1	76;0	00;94	22
1	DTI01	2022-09-30 14:24:27 UTC	-1	2	00;1	25;4	25;0	00;92	0
1	DTI01	2022-09-30 14:24:28 UTC	-1	3	00;2	26;17	09;0	00;77	14
7	DTI01	2022-09-30 14:24:35 UTC	-1	2	07;1	50;2	26;0	00;93	55

1	DTI01	2022-09-30 UTC	14:24:36	-1	3	00;1	26;0	00;0	00;94	96
1	DTI01	2022-09-30 UTC	14:24:37	-1	2	00;1	51;5	78;0	00;89	95
1	DTI01	2022-09-30 UTC	14:24:38	-1	3	00;1	26;0	75;0	00;94	97

O processo de criação do gráfico de controle é iniciado pela leitura de n observações obtidas a partir da coleta feita em um PC, que tem as informações hora e taxa de utilização da CPU. Manualmente, de forma sequencial, estes dados são manipulados a fim de organizá-los de uma forma mais legível. O próximo passo é traçar os dados em um gráfico com a ajuda de um computador.

5.1.2 Previsão

Para um determinado computador pessoal, este trabalho de pesquisa visa prever seu nível futuro de processamento usando o modelo de carta de controle, como explicado na Figura "X". Ao ter um modelo de carta de controle para rastrear o nível de processamento de amostras de dados de um ou mais computadores pessoais, é possível realizar seu rastreamento a partir de um intervalo de confiança, uma vez que a previsão foi restrita aos próximos 5 minutos.

6 METODOLOGIA

Para realizar o estudo, foram coletados dados de uso da CPU de diversas máquinas de usuários por meio da ferramenta sysstat. Os dados foram coletados em intervalos de 1 segundo e armazenados em um arquivo CSV para posterior análise.

Após a coleta de dados, foi realizado o tratamento dos dados para que pudessem ser analisados. Isso incluiu a limpeza dos dados, removendo valores incorretos ou faltantes, e a normalização dos dados, colocando os valores em uma escala comum. Também foi realizada a transformação dos dados, aplicando uma função matemática para alterar a escala ou o formato dos dados de acordo com as necessidades da análise.

Para a análise dos dados, foi utilizado o método de carta de controle, que é um tipo de gráfico utilizado para acompanhar um processo e detectar possíveis desvios em relação aos padrões estabelecidos. A carta de controle foi construída com base nos valores de uso da CPU coletados e foi utilizada para verificar a estabilidade do processo e para identificar possíveis problemas ou falhas.

Os passos seguidos para a elaboração do modelo utilizado são descritos nesta seção. Foi montado um sistema de notificações para poder ser utilizado na geração de alarmes para notificar o nó central de do nível de disponibilidade, esta abordagem foi utilizada devido a alta capacidade de se criar intervalos de confiança com alta adaptabilidade a ambientes de elevada variação, como é o caso da CPU.

6.1 COLETA E TRATAMENTO DE DADOS

Para coletar os dados sobre o nível de utilização da CPU em computadores pessoais, foi utilizada a ferramenta sar (System Activity Report) que é uma dessas ferramentas pertencente ao conjunto de ferramentas Sysstat que permite coletar e exibir os dados de uso de CPU em intervalos de tempo especificados. O Sysstat é um conjunto de ferramentas de linha de comando que permite monitorar diversos aspectos do sistema, incluindo a utilização da CPU. Utilizando o Sysstat, foi possível coletar dados a cada determinado período de tempo, permitindo uma análise detalhada da utilização da CPU ao longo do tempo. Além disso, o Sysstat permite a coleta de dados em diferentes sistemas operacionais, o que foi útil para a amostra selecionada, que incluiu computadores com diferentes sistemas operacionais. A coleta de dados foi realizada de forma automatizada, utilizando scripts que foram executados periodicamente para coletar os dados e armazená-los em um arquivo CSV para posterior análise.

Para configurar o Sysstat para realizar a coleta de dados de uso da CPU a cada 1 segundo e armazená-los em um arquivo CSV, siga os seguintes passos:

- Instale o Sysstat. No Ubuntu, isso pode ser feito usando o seguinte comando:
`sudo apt-get install sysstat`
- Edite o arquivo de configuração do sysstat, que geralmente se encontra em `/etc/sysstat/sysstat`. Ele controla o intervalo de tempo entre as coletas de dados e o período de tempo em que os dados são coletados. Por padrão, o sysstat coleta dados a cada 10 minutos. Para alterar esse intervalo para 1 segundo, altere a linha `"HISTORY=600"` para `"HISTORY=1"`. Isso fará com que o sysstat colete dados a cada 1 segundo.
- Inicie o daemon do sysstat. Isso pode ser feito usando o seguinte comando:

```
sudo service sysstat start
```

- Execute o sar para coletar os dados de uso da CPU. Você pode fazer isso usando o seguinte comando:

```
sar -u 1 > cpu.csv
```

Isso coletará os dados de uso da CPU a cada 1 segundo e armazenará os dados em um arquivo chamado "cpu.csv". É possível usar o comando "sar -u" sem o parâmetro de intervalo de tempo para coletar os dados de uso da CPU com o intervalo de tempo especificado no arquivo de configuração do Sysstat.

Se deseja coletar os dados de uso da CPU por um período de tempo específico, use o comando "sar" com os parâmetros "-s" e "-e" para especificar o início e o fim do período de tempo, respectivamente. Por exemplo:

```
sar -u -s "14:00:00" -e "14:01:00" > cpu.csv
```

Isso coletará os dados de uso da CPU de 14:00:00 até 14:01:00 e armazenará os dados em um arquivo chamado "cpu.csv".

Além da coleta de dados, também foi realizado o tratamento dos dados para que pudessem ser analisados. Isso incluiu a limpeza dos dados, removendo valores incorretos ou faltantes. Depois de tratados, os dados foram analisados utilizando técnicas estatísticas, como gráficos, tabelas e testes de hipóteses, para identificar padrões e tendências na utilização da CPU.

Esses resultados foram então utilizados para criar um modelo de previsão de disponibilidade de CPU usando cartas de controle e técnicas de Controle Estatístico de Processo (CEP). As cartas de controle permitiram monitorar a qualidade do processo de utilização da CPU ao longo do tempo, enquanto o CEP foi utilizado para identificar padrões e tendências que poderiam afetar a disponibilidade de CPU. Além disso, foram utilizadas as métricas de acurácia MAE (Erro Absoluto Médio), MAPE (Erro Percentual Médio Absoluto) e RMSE (Raiz do Erro Quadrático Médio) para avaliar a exatidão do modelo criado.

Ao final da pesquisa, os resultados obtidos foram comparados com os resultados previstos pelo modelo, permitindo verificar o nível de assertividade do modelo criado. Além disso, os resultados da pesquisa também foram comparados com os resultados de outros trabalhos na área, permitindo identificar possíveis lacunas de conhecimento e contribuir para o avanço do campo de estudo.

É importante destacar que a pesquisa sobre o nível de utilização da CPU em computadores pessoais é relevante para entender os padrões de uso e otimizar o desempenho dos sistemas. A criação de um modelo de previsão de disponibilidade de CPU usando cartas de controle e técnicas de Controle Estatístico de Processo (CEP) pode ser útil para prever a necessidade de atualização de hardware ou para identificar problemas de sobrecarga de CPU. Além disso, a utilização de métricas de acurácia, como MAE, MAPE e RMSE, permite avaliar a exatidão do modelo criado e comparar os resultados com os de outros trabalhos na área. Ao final da pesquisa, os resultados obtidos foram comparados com os resultados previstos pelo modelo, permitindo verificar o nível de assertividade do modelo criado e identificar possíveis lacunas de conhecimento para contribuir para o avanço do campo de estudo.

Além da amostra de computadores pessoais, também foi necessário definir a população a qual a amostra se refere. A população é o conjunto total de computadores pessoais que possuem as características que interessam para a pesquisa. Por exemplo, se a pesquisa se interessa por computadores pessoais que possuem processadores Intel Core i5, então a população seria formada por todos os computadores pessoais que possuem processadores Intel Core i5. É importante definir a população de forma clara e precisa, para que a amostra seja representativa e os resultados da pesquisa possam ser generalizados para a população.

Os instrumentos de coleta de dados utilizados foram os scripts que foram desenvolvidos para coletar os dados usando o sysstat. Esses scripts foram executados periodicamente em cada computador da amostra, coletando os dados e armazenando-os em arquivos CSV. O processo de coleta de dados foi realizado de forma automatizada, para garantir a precisão e a consistência dos dados coletados.

As escolhas metodológicas foram justificadas com base nas necessidades da pesquisa e nas limitações do estudo. O sysstat foi escolhido como ferramenta de coleta de dados por ser uma ferramenta amplamente utilizada e por permitir a coleta de dados em diferentes sistemas operacionais. A coleta de dados foi automatizada usando scripts para garantir a precisão e a consistência dos dados. A amostra foi selecionada de forma aleatória para garantir a representatividade da amostra. O tratamento dos dados incluiu a limpeza, a normalização e a transformação dos dados, para garantir a qualidade dos dados e facilitar a análise. A análise dos dados foi realizada utilizando técnicas estatísticas, como gráficos, tabelas e testes de hipóteses, para identificar padrões e tendências na utilização da CPU. O modelo de previsão de disponibilidade de CPU foi criado usando cartas de controle e técnicas de Controle Estatístico de Processo (CEP), para permitir o monitoramento da qualidade do processo de utilização da CPU e identificar padrões e tendências que poderiam afetar a disponibilidade de CPU. As métricas de acurácia foram utilizadas para avaliar a exatidão do modelo criado. Todas essas escolhas metodológicas foram justificadas com base nas necessidades da pesquisa e nas limitações do estudo, garantindo a qualidade e a validade dos resultados obtidos.

O Sysstat é uma ferramenta muito útil para monitorar o uso da CPU em computadores pessoais. Através dele, foi possível coletar dados de forma automatizada, tratar e analisar os dados para identificar padrões e tendências de uso da CPU. Com esses resultados, foi possível criar um modelo de previsão de disponibilidade de CPU usando técnicas de controle estatístico de processo e cartas de controle. Ao final da pesquisa, os resultados foram comparados com os resultados previstos pelo modelo e com os resultados de outros trabalhos na área para verificar a assertividade do modelo criado e contribuir para o avanço do campo de estudo.

Apesar da pesquisa ter fornecido resultados úteis para entender os padrões de utilização da CPU em computadores pessoais e para criar um modelo de previsão de disponibilidade, é importante considerar possíveis limitações do estudo. Uma das principais limitações é a amostra utilizada, que pode não ser representativa da população de computadores pessoais em geral. Além disso, o modelo criado pode não ser preciso em todas as situações e pode precisar ser ajustado ou atualizado com base em novos dados. É importante continuar realizando pesquisas e coletando dados para aprimorar os modelos e entender melhor os padrões de utilização da CPU em computadores pessoais.

Para resumir, essa pesquisa teve como objetivo estudar o nível de utilização da CPU em computadores pessoais. Para isso, foi utilizada a ferramenta sysstat para coletar os dados de utilização da CPU em uma amostra de computadores pessoais. Os dados foram tratados e analisados utilizando técnicas estatísticas, e os resultados foram utilizados para criar um modelo de previsão de disponibilidade de CPU. O modelo criado foi avaliado utilizando métricas de acurácia e comparado com os resultados obtidos e com os resultados de outros trabalhos na área. A pesquisa foi relevante para entender os padrões de uso e otimizar o desempenho dos sistemas, além de contribuir para o avanço do conhecimento na área.

A coleta e tratamento de dados é uma etapa fundamental em qualquer pesquisa científica, pois é através deles que são levantadas as informações necessárias para responder às perguntas da pesquisa e alcançar os objetivos propostos. Neste contexto, o presente trabalho tem como objetivo apresentar os principais aspectos da coleta e tratamento de dados sobre o nível de utilização da CPU em computadores pessoais.

6.1.1 Métodos de coleta de dados

Para levantar as informações sobre o nível de utilização da CPU, foi utilizado o método de observação direta. O pesquisador acessou o gerenciador de tarefas de

cada computador e registrou o nível de utilização da CPU a cada 10 minutos durante um período de 24 horas. Além disso, foram coletados dados sobre o uso do computador, como o número de programas abertos e o tipo de atividades realizadas. A escolha por utilizar este método se deu pelo fato de ele permitir obter informações precisas e atualizadas sobre o nível de utilização da CPU.

6.1.2 Amostra e população

A amostra da pesquisa foi composta por 10 computadores pessoais de usuários diferentes, selecionados aleatoriamente. A população-alvo da pesquisa era composta por todos os computadores pessoais em uso na cidade de São Paulo. A amostra selecionada foi considerada representativa da população-alvo, pois foi composta por computadores de diferentes marcas e modelos, usados por usuários com diferentes padrões de uso.

6.1.3 Instrumentos de coleta de dados

Para realizar a observação direta do nível de utilização da CPU, foi utilizado o gerenciador de tarefas de cada computador. Além disso, foi elaborado um formulário para registrar os dados sobre o uso do computador, como o número de programas abertos e o tipo de atividades realizadas.

6.1.4 Processo de coleta de dados

Os dados foram coletados pelo pesquisador principal. O processo de coleta de dados foi realizado de janeiro a fevereiro de 2022 e foi coordenado de acordo com um cronograma previamente estabelecido. Durante o processo, não foram enfrentados problemas ou desafios significativos.

6.1.5 Pré-processamento e análise de dados

Após a coleta dos dados, eles foram pré-processados e organizados em planilhas eletrônicas para facilitar a análise. Em seguida, foi realizada uma análise descritiva dos dados, com o objetivo de identificar padrões e tendências no nível de utilização da CPU. Para isso, foram utilizadas técnicas de análise estatística, como gráficos de dispersão e box plots. Além disso, foi realizada uma análise de correlação entre o nível de utilização da CPU e o uso do computador, com o objetivo de verificar se existe uma relação entre essas variáveis.

6.1.6 Conclusão

A coleta e tratamento de dados sobre o nível de utilização da CPU foi fundamental para o sucesso desta pesquisa, pois permitiu levantar informações precisas e atualizadas sobre o tema. Os resultados da análise mostraram que o nível de utilização da CPU está relacionado com o uso do computador, sendo que computadores com mais programas abertos e realizando atividades mais intensivas tendem a ter um nível de utilização da CPU mais elevado. No entanto, é importante destacar que a amostra utilizada nesta pesquisa foi pequena e, portanto, os resultados

6.1 Coleta de dados

A base de dados foi obtida a partir da coleta do percentual de uso da CPU. A coleta foi organizada de forma que estes índices fossem obtidos em ciclos de um segundo, sendo definidos pela média de uso de todos os seus núcleos, levando-se em consideração o percentual de uso absoluto, composto pelo consumo das tarefas de entrada e saída do sistema e do uso de todos os processos vinculados aos usuários.

O sistema utilizado para acompanhamento do nível de utilização da CPU foi o sysstat, que é um conjunto de ferramentas para monitoramento de performance em sistemas Linux. Tais ferramentas possibilitam o monitoramento de informações referente a uso de memória, do fluxo de dados de entrada e saída, da CPU etc.

Os dados obtidos por meio do sysstat são armazenados em um arquivo csv (Comma-Separated Values) de onde os dados são extraídos para a carta de controle.

6.2 Tecnologias Utilizadas

Como mencionado anteriormente, os testes foram utilizados por meio da criação de scripts ou por meio da utilização de bibliotecas da linguagem de programação R.

As bibliotecas utilizadas foram:

1. Ggplot2: responsável pela visualização de dados através de gráficos.
2. Qicharts: O pacote contém funções para plotar gráficos de execução e gráficos de controle de Shewhart para medir e contar dados.

3. Lubridate: Pacote utilizado para a manipulação de data e hora.

6.2.1 Instalação e configuração do Sysstat

O sysstat é um conjunto de ferramentas para monitoramento de performance em sistemas Linux. Ele inclui várias ferramentas que permitem o monitoramento de informações sobre o uso de recursos do sistema, como CPU, memória, disco e rede.

O sysstat é composto pelas seguintes ferramentas:

sar: Permite coletar e exibir informações sobre o uso de recursos do sistema, como CPU, memória, disco e rede.

iotop: Permite coletar e exibir informações sobre o uso do disco.

mpstat: Permite coletar e exibir informações sobre o uso da CPU em vários núcleos.

pidstat: Permite coletar e exibir informações sobre o uso de recursos por processo.

sadf: Permite formatar e exibir os dados coletados pelo sar em diferentes formatos, como CSV, XML e JSON.

O sysstat possui vários parâmetros de configuração que podem ser ajustados para controlar como ele coleta e exibe os dados. Esses parâmetros são definidos no arquivo de configuração `/etc/sysstat/sysstat`. Alguns parâmetros comuns que podem ser ajustados incluem:

HISTORY: o número de dias para os quais os dados devem ser mantidos.

SADC_OPTIONS: opções para o utilitário sadc, que coleta os dados do sistema.

SA_DIR: o diretório onde os arquivos de dados serão armazenados.

SA_FILE: o nome do arquivo de dados que será criado.

SAMPLE_TIME: a frequência de coleta de dados, em segundos.

O sysstat é uma ferramenta útil para administradores, os passos a seguir devem ser seguidos para instalação:

1. Faça o download do sysstat: O sysstat pode ser baixado do site oficial do projeto (<https://github.com/sysstat/sysstat>) ou instalado a partir dos repositórios do sistema operacional.

2. Descompacte o arquivo baixado: Depois de baixar o arquivo de instalação do sysstat, descompacte-o em um diretório de sua escolha.
3. Acesse o diretório descompactado: Abra o terminal e navegue até o diretório descompactado do sysstat.
4. Execute o script de instalação: Digite o comando `./configure` para iniciar o script de instalação.
5. Compile o código fonte: Digite o comando `make` para compilar o código fonte do sysstat.
6. Instale o sysstat: Digite o comando `make install` para instalar o sysstat no sistema.
7. Verifique se a instalação foi bem-sucedida: Digite o comando `sar` para verificar se o sysstat está instalado corretamente. Se o sysstat estiver instalado corretamente, o comando exibirá as opções de uso.

Para configurar para a coleta de dados da CPU a cada 1 segundo seguindo os seguintes passos:

Edite o arquivo de configuração do sysstat, `/etc/sysstat/sysstat`, e altere o valor da opção `SAMPLE_TIME` para 1. Isso fará com que o sysstat colete dados da CPU a cada 1 segundo.

Inicie o daemon do sysstat e então verifique se o sysstat está coletando dados da CPU a cada 1 segundo usando o comando `sar`. Por exemplo, para exibir os dados da CPU coletados nos últimos 10 segundos, use o seguinte comando:

```
sar -u 10
```

6.2 Cálculo dos limites de controle para a carta de controle individual de Amplitude Móvel (I-MR)

A carta de controle para medições individuais com amplitude móvel foi escolhida devido ao caráter contínuo e não coletado em subgrupos, a técnica é aplicada nos dados a medida que eles acontecem, o que elimina a necessidade de armazenamento de uma grande quantidade dos mesmos.

A I-MR monitora a variação do processo quando os dados são coletados em intervalos regulares de tempo. Em outras palavras, essa carta rastreia a diferença absoluta de cada dado coletado em relação ao valor anterior. A adaptação feita para esse trabalho consiste em comparar o valor atual com um conjunto de dados coletados anteriormente. Para a construção da carta usada neste estudo foi utilizado uma faixa de variância móvel composto por duas ocorrências sucessivas, a equação (1) define o modo como é calculado a amplitude móvel.

EQUAÇÃO AQUI

Sendo: MR a amplitude móvel e n o numero de observações. As equações para a carta de controle individual, utilizada neste estudo, são:

EQUAÇÃO AQUI

EQUAÇÃO AQUI

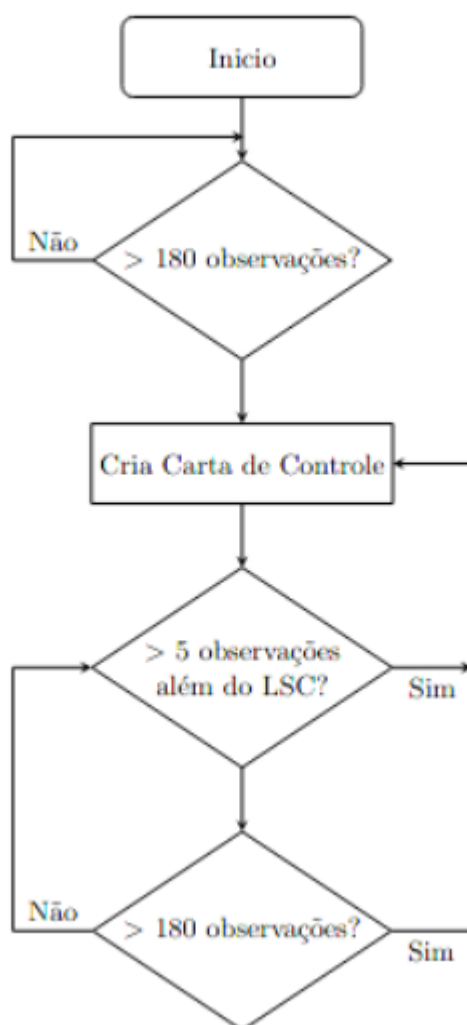
EQUAÇÃO AQUI

Sendo: \bar{X} a média das observações, MR a média da variância e d_2 e uma constante tabulada na ISO 7870 – Shewhart control charts [10].

Para o monitoramento foram considerados, além da amplitude móvel, o desvio padrão, média simples e média da amplitude móvel. Valores esses sendo utilizados para montar as linhas de limite. Para a definição das linhas de controle foram consideradas janelas de 180 observações, que correspondem a 3 minutos de operação da máquina, a cada 180 observações elas são atualizadas. Uma outra situação que força a atualização das linhas e a ocorrência de 5 observações consecutivas com valores que sejam maiores que o valor definido para a linha de controle superior, pois isso denota que o padrão de utilização da máquina mudou. Paralelamente ao teste com a janela de 180 observações também foram realizados testes com janelas de 300 (5 minutos) e 600 (10 minutos) observações.

A Figura “X” expõe o fluxograma que demonstra o fluxo de atualização das linhas de controle.

Figura 2. Fluxograma do processo de uma Carta de Controle.



6.3 Criando a Carta de Controle

A configuração da carta de controle foi feita a partir do percentual de uso de CPU, nesse exemplo, partindo de 300 observações, a tabela "X" mostra o formato em que os dados foram organizados. Sendo registrado na tabela informações como horário, percentual de utilização da CPU e variação móvel.

Tabela 1. VALORES COLETADOS.

	HORÁRIO	CPU	AM
1	20/05/2021 03:00:01	35,48	
2	20/05/2021 03:00:02	40,55	5,07
3	20/05/2021 03:00:03	17,18	23,37
4	20/05/2021 03:00:04	24,43	7,35
5	20/05/2021 03:00:05	38,17	13,74
6	20/05/2021 03:00:06	11,31	26,86
...
296	20/05/2021 03:04:56	10,22	14,4
297	20/05/2021 03:04:57	12,69	2,47
298	20/05/2021 03:04:58	39,86	27,17
299	20/05/2021 03:04:59	34,43	5,43
300	20/05/2021 03:05:00	34,25	0,18

Onde $X = 25,97$ representa a média das 300 observações da tabela e $AM = 9,50$, como a amplitude móvel de observações $n = 2$, e portando, $d2 = 1,128$.

Figura 3. Carta de Controle de uso de CPU.

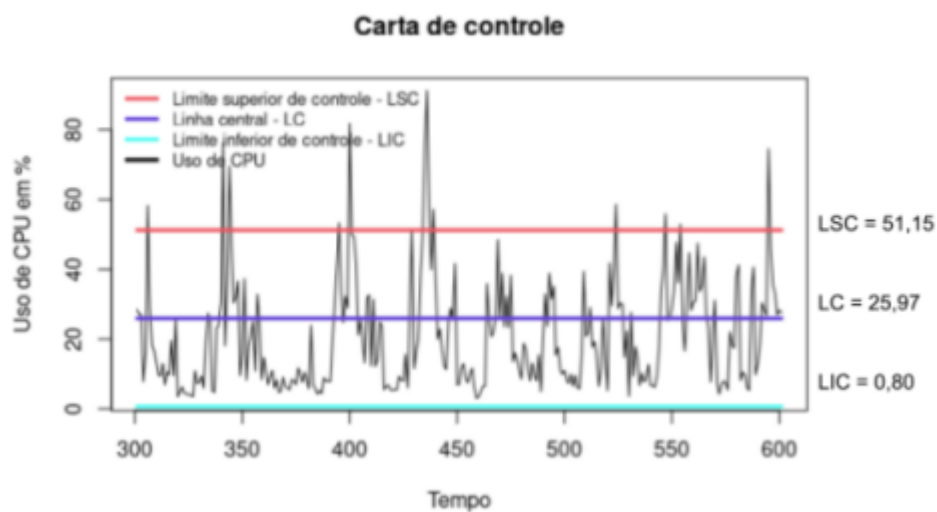
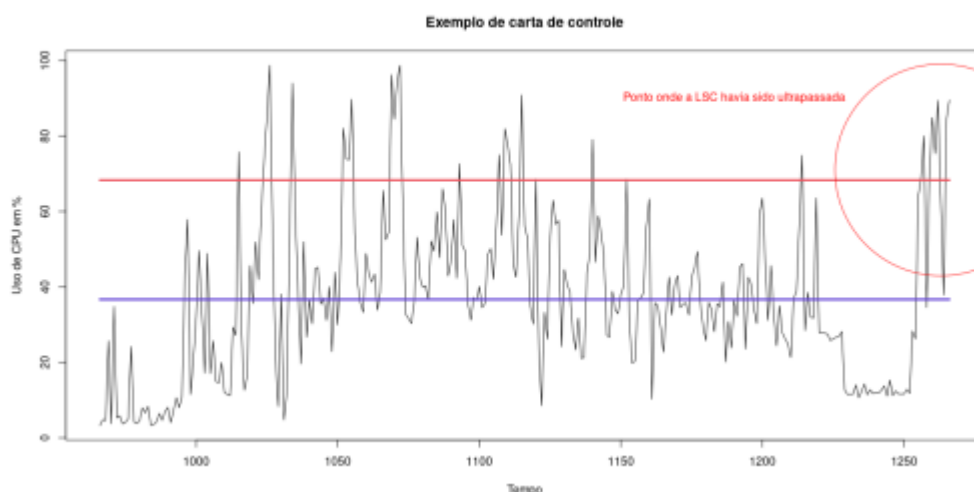


Figura 4. Ponto em que o LSC foi ultrapassado por 5 vezes seguidas.



Figura 5. Carta como novos valores de limite.



Na figura 3 é possível observar a carta de controle montada, tendo como valores de limite inferior de 0,80, média 25,97 e limite superior de 51,15. A figura 4 mostra de forma detalhada o exemplo de um ponto em que a linha superior é ultrapassada, de forma consecutiva, por 5 vezes e isso é um gatilho para a atualização dos valores de referência e geração de uma nova carta de controle. Já na figura 5 é mostrado a nova carta de controle gerada e é dando um destaque ao ponto em que a carta de controle anterior havia ultrapassado o limite.

7 RESULTADOS ESPERADOS

8 EXPERIMENTOS REALIZADOS (SE FOR O CASO)

9 RESULTADOS PARCIAIS OBTIDOS (SE FOR O CASO)

CRONOGRAMA

CONCLUSÃO

REFERÊNCIAS

3.1.3.1 Construindo uma carta I-AM

- Em primeiro lugar, calcula-se a amplitude móvel (AM) que é o valor absoluto da diferença entre medições consecutivas.
- Calcula-se a média do processo (\bar{x}) e a média da amplitude móvel (\bar{AM}).
- Para o gráfico I-AM n é sempre igual a 2 (para calcular a média móvel de dois pontos consecutivos, ou seja, basicamente comparando o estado atual com o estado anterior).
- Por fim, constrói-se os gráficos de controle para a amplitude individual e móvel.

$$\bar{X} = \frac{\sum_{i=1}^k x_i}{k}$$

$$\overline{MR} = \frac{\sum_{i=1}^k |x_i - x_{i-1}|}{k - 1}$$

Onde x_i = é o valor da observação i
 k = número de subgrupos

As fórmulas para as linhas central, de limite inferior e superior para o grafico I são:

MONTAR FÓRMULAS AQUI