

Predição de
Falhas em
Ambientes
de Big Data

Equipe: Grupo 12 – Gustavo H

Gustavo Hammerschmidt

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão

Outline

Predição de Falhas em Ambientes de Big Data

- **Problema/Motivação**
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão

Problema/ Motivação

Ambientes de Big Data possuem grande poder de processamento de dados, devido ao tamanho dos dados que precisam de tratamento. Para tal cenário, o uso de uma infraestrutura de computadores interconectados é o modo pelo qual cientistas de dados operam grandes volumes de informações.

Contudo, tais ambientes aumentam a propensão dos computadores a falhas, levando a perdas de partes de arquivos, inatividade e lentidão.

Problema/ Motivação


Alguns algoritmos de previsão de falhas de ambiente foram criadas, em decorrência disso, para detectar possíveis falhas e resolvê-las antes de afetar o ambiente como um todo.

Entretanto, algumas empresas ainda não produzem um vasto conjunto de dados, e, para aplicar algoritmos de análise, utilizam-se de conjuntos públicos ou de empresas com ambientes similares, o que pode levá-las a conclusões precipitadas devido à natureza do algoritmo utilizado,

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- **Objetivo da pesquisa**
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão



Objetivo da Pesquisa

Identificar o desenvolvimento de algoritmos de predição de falhas de ambiente e seu estado da arte.

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- **Método de pesquisa**
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão

Método de pesquisa

Estudo de casos e revisões de modelos:

[CBI]Avci C., Tekinerdogan B. e AthanasiadisI.N. “Software architectures for big data: a systematic literature review.” Em: BigData Anal 5, 5 (2020).(). DOI:<https://doi.org.ez433.periodicos.capes.gov.br/10.1186/s41044-020-00045-1>.

[DAI+]Hong-Ning Dai et al. “Big Data Analyticsfor Large-scale Wireless Networks: Chal-lenges and Opportunities.” Em:ACMComput. Surv. 52, 5, Article 99 (Septem-ber 2019), 36 pages.(). DOI:<https://doi.org/10.1145/3337065>.

[EDS]Kim E., Huh D. e Kim S. “Knowledge-based power monitoring and fault pre-diction system for smart factories.” Em:Pers Ubiquit Comput (2019).(). DOI:<https://doi.org.ez433.periodicos.capes.gov.br/10.1007/s00779-019-01348-4>.


[Kit]Barbara Kitchenham. “Procedures forPerforming Systematic Reviews.” Em:NICTA Technical Report 0400011T.1(). DOI:<https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>.

[P+]Mikalef P. et al. “Big data analytics ca-pabilities: a systematic literature reviewand research agenda.” Em:Inf Syst E-Bus Manage 16, 547–578 (2018).(). DOI:<https://doi.org.ez433.periodicos.capes.gov.br/10.1007/s10257-017-0362-y>.

[SS]Rathore S.S. e Kumar S. “A study on soft-ware fault prediction techniques.” Em:Artif Intell Rev 51, 255–327 (2019).(). DOI:<https://doi.org.ez433.periodicos.capes.gov.br/10.1007/s10462-017-9563-5>.

[Son+] Qingquan Song et al. “Tensor Comple-tion Algorithms in Big Data Analytics.”Em:ACM Trans. Knowl. Discov. Data 13,1, Article 6 (January 2019), 48 pages.(). DOI:<https://doi.org/10.1145/3278607>.

[Vu+]Tin Vu et al. “Using Deep Learning forBig Spatial Data Partitioning.” Em:ACMTrans. Spatial Algorithms Syst. 7, 1, Ar-ticle 3 (August 2020), 37 pages.(). DOI:<https://doi.org/10.1145/3402126>.



Método de
pesquisa

Esta pesquisa é uma Revisão
Sistemática de Literatura(RSL).

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- **Questões de pesquisa**
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão

Questões de Pesquisa

QP1) se o algoritmo é capaz de descrever a situação de falha a que se aplica ou como a evita;

QP2) se o grau de complexidade de complexidade do ambiente de big data impacta na complexidade do algoritmo e se influencia no seu desempenho;

QP3) se as falhas de um ambiente, cujo o algoritmo foi desenvolvido, têm impacto na avaliação deste se é aplicado em um outro ambiente.

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- **Discussão de resultados**
- Limitações da pesquisa
- Contribuições da pesquisa
- Conclusão

Discussão de Resultados

Nós observamos que algoritmos de predição de possíveis falhas estão, intimamente, integrados aos sistemas que operam sobre, logo sendo bastante diferentes entre si.

Contudo, nossa pesquisa apontou que abordagem por eles utilizadas pode ser abstraída a um padrão, ou uma lista de fatores que devem ser levados em consideração por algoritmos de predição, de forma a serem capazes de prover análises acuradas.

Discussão de Resultados

Como apontado no Artigo[EDS], a respeito do monitoramento de indústrias smart, o constante monitoramento dos equipamentos de ambientes de trabalho da indústria e suas informações foram fundamentais para que um estudo de falhas fosse feito.

Neste estudo, o consumo de energia de alguns aparelhos sofria oscilações em momentos de falha.

Com a coleta destas anormalidades de consumo, um conjunto indicativo de dados foi derivado para mapeamento de falhas, notificando os profissionais responsáveis sobre irregularidades.

Desta forma, respondendo à questão de pesquisa 1(QP1), indicando que há impacto, porém, ele sofre influencia também do conhecimento dos analistas.

Discussão de Resultados

Neste estudo[SS], contudo, Santosh indica que muitos ambientes utilizam conjuntos de dados públicos para previsão de possíveis falhas em seus ambientes e algoritmos de análise de baixa performance.

Para que um ambiente seja anti-falha, é necessário que a falha tenha sido reportada ao menos uma vez e estudada.

Para Santosh, com muitos estudos, é possível derivar um algoritmo capaz de auxiliar organizações e empresas com um histórico insuficiente de falhas, mas, para isso, seria necessário um algoritmo compatível com diferentes projetos e que houvessem similaridades entre os ambientes.

Respondendo, assim, a nossa questão de pesquisa 2(QP2), onde a complexidade do ambiente em que o algoritmo é desenvolvido pode proporcionar uma inflexibilidade deste a novos contextos.

Discussão de Resultados

Como fora mencionada, algumas empresas são incapazes de gerar um vasto histórico de falhas de ambiente, portanto, usam algoritmos de organizações que possuam ambientes similares.

Porém, as análises feitas por um algoritmo são teóricas, baseadas em pressupostos estatísticos -- que não necessariamente deduzem uma situação adequadamente.

Como indicado na questão de pesquisa 3(QP3), as falhas de ambientes genéricas podem impacatar negativamente no desempenho de um outro ambiente. Além disso, os estudos citados ressaltam a importância de gerar dados heterogêneos para uma maior conformidade dos algoritmos a diferentes ambientes ou situações.

Para Qingquan Song, et al[[Son+](#)], acredita que a expertise dos profissionais é tão relevante quanto a simples automatização da rotina, instigando uma constante análise do ambiente pelos seus profissionais também.

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- **Limitações da pesquisa**
- Contribuições da pesquisa
- Conclusão

Limitações da Pesquisa

A nossa revisão baseia-se em artigos postados nas bases SpringerLink e ACM. A seleção dos artigos fora feita:

Tabela 1. Tabela 1: Inclusão e Exclusão dos artigos selecionados

	I1	E1	E2	E3	Final
SpringerLink	508.537	20	14	2	2
ACM	49.571	20	8	3	3

- 1) com base em artigos relacionados as palavras-chave: big data, environment, fault e prediction;
- 2) foram selecionados os 20 artigos mais relevantes de acordo com as bases;
- 3) todos os artigos duplicados foram descartados;

Limitações da Pesquisa

A nossa revisão baseia-se em artigos postados nas bases SpringerLink e ACM. A seleção dos artigos fora feita:

Tabela 1. Tabela 1: Inclusão e Exclusão dos artigos selecionados

	I1	E1	E2	E3	Final
SpringerLink	508.537	20	14	2	2
ACM	49.571	20	8	3	3

4) aqueles que não escritos na língua inglesa foram removidos;

5) só foram selecionados artigos postados entre 2019 e 2020.

Limitações da Pesquisa

A nossa revisão baseia-se em artigos postados nas bases SpringerLink e ACM. A seleção dos artigos fora feita:

Tabela 1. Tabela 1: Inclusão e Exclusão dos artigos selecionados

	I1	E1	E2	E3	Final
SpringerLink	508.537	20	14	2	2
ACM	49.571	20	8	3	3

Na primeira fase de exclusão(E1), apenas 20 artigos de cada base foram selecionados com base no número de menções feitas e relevância indicados pela a respectiva base.

Na segunda fase de exclusão(E2), os artigos foram avaliados e removidos com base na duplicação entre bases e na falta de palavras-chave condizentes com a proposta de RSL.

Limitações da Pesquisa

A nossa revisão baseia-se em artigos postados nas bases SpringerLink e ACM. A seleção dos artigos fora feita:

Tabela 1. Tabela 1: Inclusão e Exclusão dos artigos selecionados

	I1	E1	E2	E3	Final
SpringerLink	508.537	20	14	2	2
ACM	49.571	20	8	3	3

Na última fase de exclusão(E3), os artigos foram lidos e avaliados com o grau de aferição de valor à pesquisa, restando apenas 5 artigos no total.

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- **Contribuições da pesquisa**
- Conclusão



Contribuições à Pesquisa

Esta pesquisa visa contribuir a todos interessados na área de falhas de ambientes e àqueles que buscam compreender o estado da arte dos algoritmos de predição.

Outline

Predição de Falhas em Ambientes de Big Data

- Problema/Motivação
- Objetivo da pesquisa
- Método de pesquisa
- Questões de pesquisa
- Discussão de resultados
- Limitações da pesquisa
- Contribuições da pesquisa
- **Conclusão**

Conclusão

Os resultados obtidos indicam que ambientes precisam sofrer falhas repetidas vezes para que um conjunto de dados seja derivado, e possa proporcionar acurácia ao algoritmo de predição.

É necessário, também, um conhecimento das especificações dos dispositivos do ambiente para que um relatório do funcionamento ideal desses seja formulado e utilizado por algoritmos de aprendizado como referência.

Todo o histórico do ambiente gerado é utilizado para a construção de algoritmos de aprendizado próprio, que auferem análises ao ambiente, reportando anormalidades conhecidas aos profissionais responsáveis.

É também importante para o ambiente que haja a integração dos resultados de predição e o uso da expertise de seus profissionais para proporcionar um melhor funcionamento do ambiente como um todo.