**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data**

**Romulos Machado**

**Phillip Furtado**

**Análise de Consequências de Alta Temperaturas em Data Center usando CIÊNCIA DE DADOS**

Belo Horizonte

2020

**Romulos Machado**

**Phillip Furtado**

**Análise de Consequências de Alta Temperaturas em Data Center usando CIÊNCIA DE DADOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2020

**SUMÁRIO**

[**1. Introdução 4**](#_Toc36150682)

[1.1. Contextualização 4](#_Toc36150683)

[1.2. O problema proposto 4](#_Toc36150684)

[2. Coleta de Dados 5](#_Toc36150685)

[3. Processamento/Tratamento de Dados 5](#_Toc36150686)

[4. Análise e Exploração dos Dados 5](#_Toc36150687)

[5. Criação de Modelos de Machine Learning 5](#_Toc36150688)

[6. Apresentação dos Resultados 6](#_Toc36150689)

[7. Links 6](#_Toc36150690)

[REFERÊNCIAS 7](#_Toc36150691)

# 1. Introdução

## 1.1. Contextualização

Para que data centers e seus respectivos equipamentos (servidores, grandes unidades de armazenamento e ativos de rede) consigam atender à necessidade dos seus usuários, é imperativo que permaneçam em funcionamento constante 24 horas por dia, 7 dias por semana. Consequentemente, assim como a maioria dos equipamentos que trabalham com processamento de informações, a temperatura nos data centers está diretamente relacionada com o bom funcionamento dos servidores, sendo fundamental que eles não superaqueçam durante a execução de suas atividades. Pois, caso isso aconteça, estes equipamentos correm o risco de se auto desligarem ou queimarem, o que pode causar uma parada não programada do data center, consequentemente a interrupção dos serviços por ele provido. Com isso, se faz necessário uma análise estatística e preditiva dos eventos de elevação da temperatura de data centers e suas consequências no funcionamento dos equipamentos.

## 1.2. O problema proposto

No que concerne a alta disponibilidade de serviços, identificar quais danos foram causados por eventos de alta temperatura dentro de um data center, a partir de dados coletados sobre o próprio ambiente e usando conceitos de ciências de dados para analisá-los, ajuda não somente a entender a extensão de problemas em equipamentos, como também a definir sistematicamente que determinadas condições estabelecidas, possivelmente irão gerar impactos negativados decorrentes e, por isso, deve ser evitados ao máximo.

Para desenvolvedor este trabalho, foram utilizados dados de temperatura de 30 servidores Dell, de modelo R720, de um determinado data center de Manaus, com o objetivo de compará-los com dados de problemas ou erros de hardware apresentados entre os anos de 2014 a 2020. Onde esta comparação determinará se existe correlação entre os eventos de superaquecimento do data center e os problemas registrados nos equipamentos no mesmo período.

# 2. Coleta de Dados

A nossa base de dados consistiu em escolher 30 servidores todos de mesmas especificações e características de hardware e, partir deles, estudar dois tipos de logs de cada um desses servidores, que foram armazenados ao longo de aproximadamente 04 anos, através de um dispositivo chamado de Integrated Dell Remote Access Controller (IDRAC). Esse componente também instalado dentro de cada um desses servidores com total independência de acesso e isolado do sistema operacional, permitiu que ficassem registrados vários alertas relacionados às mais diversas falhas relativas ao hardware dessa máquinas, assim como também registrou dados de temperatura interna e externa a estes servidores, ao longo desses 04 anos

Sendo assim, como foi indicado no parágrafo anterior, tivemos a oportunidade de trabalhar com dois tipos de logs, para cada servidor:

***-*** ***Log de Alertas (log\_servidorXX.csv)***: contendo dados sobre alertas de falhas do equipamento, tendo a seguinte estrutura:

|  |  |  |
| --- | --- | --- |
| **Nome da coluna/campo** | **Descrição** | **Tipo** |
| Severity | Indica qual a severidade do alerta (“Normal”, “Warning”, ”Critical”) | Texto |
| Date/Time | Indica a data e horário do alerta | Data/horário |
| Description | Indica a mensagem de alerta | Texto |

***- Log de Temperaturas (servidorXX\_inlettemp.csv)***: contendo dados de temperaturas registrados de hora em hora dos equipamentos, tendo a seguinte estrutura:

|  |  |  |
| --- | --- | --- |
| **Nome da coluna/campo** | **Descrição** | **Tipo** |
| Average | Indica a temperatura média em Celsius, dentro do intervalo de tempo. | Numérico |
| Peak | Indica o maior pico de temperatura em Celsius, dentro do intervalo de tempo | Numérico |
| Time | Indica a data e hora a qual foi feito a captura de temperaturas, usada para os cálculos de pico e média | Data/horário |

# 3. Processamento/Tratamento de Dados

Para obtermos uma base mais consistente de dados e facilitar as associações, resolvemos adicionar o campo “Server”, de tipo “Texto”, em cada um dos logs de alertas e também nos logs de temperaturas, para preenchê-los com as respectivas informações de nome dos servidores. Isto facilitou o processo de criarmos um dataset único, contendo os registros de todos os alertas referentes aos 30 servidores. Da mesma forma, foi criado um dataset único de registros de temperaturas de todos os respectivos servidores.

Nessa seção você deve deixar registrado todo o processamento e tratamento feitos sobre os dados obtidos. É importante que você informe a quantidade de registros obtidos, a quantidade de registros duplicados ou com informações ausentes, que tratamento você deu para cada problema encontrado em seus datasets, etc. Você deve descrever cada passo de forma minuciosa, de forma que outra pessoa consiga reproduzir o seu processamento/tratamento de forma precisa. Justifique as decisões tomadas no tratamento dos dados. Por exemplo: para os valores ausentes para o campo X eu decidi preenchê-los utilizando o cálculo da média aritmética pelo motivo ..., e então justifique sua decisão.

# 4. Análise e Exploração dos Dados

Nessa seção você deve mostrar como foi realizada a análise e exploração dos seus. Mostre as hipóteses levantadas durante essa etapa e os padrões e *insights* identificados.

# 5. Criação de Modelos de Machine Learning

Conforme o documento de instruções para o TCC, essa etapa não é obrigatória, mas é fortemente recomendada. Caso você crie modelos de *Machine Learn*ing em seu projeto, nessa seção você irá descrever as ferramentas utilizadas. Se você utilizou ferramentas visuais como Knime e Rapid Miner, coloque aqui um print do seu modelo. Caso você tenha escrito scripts em Python, por exemplo, coloque aqui o seu script. Explique as *features* utilizadas, justifique a escolha por determinado modelo, os parâmetros utilizados, etc.

# 6. Apresentação dos Resultados

Nessa seção você deve apresentar os resultados obtidos. Apresente gráficos, dahsboards, conte a sua história de forma bastante criativa. Aqui você pode utilizar os modelos de Canvas propostos por Dourard (clique [aqui](https://www.louisdorard.com/machine-learning-canvas)) ou por Vasandani (clique [aqui](https://towardsdatascience.com/a-data-science-workflow-canvas-to-kickstart-your-projects-db62556be4d0)).



# 7. Links

Aqui você deve disponibilizar os links para o vídeo com sua apresentação de 5 minutos e para o repositório contendo os dados utilizados no projeto, scripts criados, etc.

# REFERÊNCIAS

Um projeto de Ciência de Dados não requer revisão bibliográfica. Portanto, a inclusão das referências não é obrigatória. No entanto, caso você deseje incluir referências relacionadas às tecnologias ou às metodologias usadas em seu trabalho, relacione-as de acordo com o modelo a seguir.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.