

**Relatório Integração de Sistemas de informação**

**Equipa**

No.20350 – Mateus Oliveira

**Licenciatura em Engenharia Sistemas Informáticos**

**3ºano**

Barcelos | outubro, 2024

Lista de Abreviaturas e Siglas

Índice de Figuras

Figura 1 – Diagrama de Contexto 11

Índice

[1. Introdução 7](#_Toc146709708)

[2. Grupo de Trabalho 8](#_Toc146709709)

[2.1. Caracterização do Grupo 8](#_Toc146709710)

[2.2. Regulamento interno do grupo 8](#_Toc146709711)

[2.3. Sistema de Avaliação interno 8](#_Toc146709712)

[2.4. Metodologia de trabalho utilizada 8](#_Toc146709713)

[2.3 Cronograma – planificação do trabalho 9](#_Toc146709714)

[3. Proposta de Sistema 10](#_Toc146709715)

[3.1. Descrição do negócio 10](#_Toc146709716)

[3.2. Objectivos de negócio 10](#_Toc146709717)

[3.3. Domínio de aplicação do sistema 11](#_Toc146709718)

[3.4. Operações a realizar pelo sistema 12](#_Toc146709719)

[3.5. Descrição dos intervenientes 12](#_Toc146709720)

[3.6. Condições específicas 12](#_Toc146709721)

[4. Anexos 13](#_Toc146709722)

[5. Conclusão 14](#_Toc146709723)

[6. Bibliografia 15](#_Toc146709724)

# Introdução

O objetivo deste trabalho é implementar um processo de ETL (Extract, Transform, Load) para coletar, transformar e carregar dados provenientes de sensores urbanos de qualidade do ar, registos de emergências médicas, e dados meteorológicos. O propósito é integrar essas diferentes fontes de dados para analisar a correlação entre condições atmosféricas e eventos de saúde.

# Problema

O problema principal que este trabalho aborda é a **integração de dados heterogêneos** para análise da relação entre a qualidade do ar e a saúde pública. Cidades modernas enfrentam desafios significativos devido à poluição atmosférica, que pode agravar problemas respiratórios e cardíacos, além de reduzir a qualidade de vida da população.

A análise desses dados integrados pode ajudar a identificar correlações entre a qualidade do ar e os impactos na saúde pública, fornecendo informações valiosas para políticas de mitigação da poluição e cuidados com a saúde urbana.

* **Desafios Identificados**

1. **Fragmentação dos Dados**: Os dados vêm de diferentes fontes (sensores IoT, APIs de clima, registos médicos) e estão em formatos distintos (CSV, JSON). Isso requer um processo robusto de **extração e padronização**.
2. **Inconsistência e Ruído nos Dados**: Sensores de qualidade do ar podem apresentar erros de leitura ou falhas de status, como dados incompletos ou marcados como "erro", o que demanda uma fase de **limpeza de dados** para remover leituras incorretas.
3. **Correlações Complexas**: A identificação de padrões entre a qualidade do ar, fatores climáticos e emergências médicas não é direta, exigindo o uso de **técnicas de análise** para encontrar relações significativas entre as variáveis.
4. **Tempo Real e Previsão**: Além dos dados históricos, é necessário prever como as condições climáticas futuras podem impactar a dispersão de poluentes e, consequentemente, a saúde pública, o que demanda a integração de dados de previsão do tempo.

Este trabalho visa resolver esses problemas implementando um fluxo de ETL que facilita a integração e análise de dados em múltiplas dimensões, com o objetivo final de fornecer insights para a gestão da qualidade do ar e a saúde pública.

# Estratégia utilizada

A estratégia utilizada para abordar o problema da integração e análise dos dados de qualidade do ar, clima e emergências médicas baseou-se na implementação de um fluxo **ETL (Extract, Transform, Load)** robusto. O processo foi dividido nas seguintes etapas principais:

**1. Extração dos Dados (Extract)**

Para realizar a coleta dos dados necessários, foi utilizado o Pentaho Kettle (PDI), uma ferramenta de ETL que permite a automação do processo de integração de dados. A extração dos dados envolveu:

* **Sensores de Qualidade do Ar**: Extração dos dados fornecidos pelos sensores IoT em formato CSV. Os sensores coletam informações sobre poluentes (PM2.5, PM10, CO, NO2, SO2, O3), além de variáveis como temperatura e umidade.
* **Dados Meteorológicos**: Utilização de APIs como a **OpenWeather API** para obter dados de previsão do tempo, como temperatura, velocidade do vento e índice UV, os quais influenciam a dispersão de poluentes.
* **Registros de Emergências Médicas**: Importação de dados JSON contendo informações sobre condições médicas críticas (como ataques cardíacos e problemas respiratórios) relacionadas com a qualidade do ar e as condições climáticas. Os dados incluíam informações como a gravidade, data e hora, local, e a condição médica dos pacientes.

**2. Transformação dos Dados (Transform)**

Após a extração, foi necessário transformar e padronizar os dados para torná-los utilizáveis e integrados. As transformações incluíram:

* **Combinação de Dados**: Foi utilizado o passo "Merge Join" do Pentaho Kettle para combinar os dados das diferentes fontes, associando-os por **localização** e **timestamp**. Isso permitiu a correlação entre os dados de qualidade do ar, previsão meteorológica e emergências médicas.
* **Limpeza de Dados**: Utilizando o passo "Filter Rows", foi aplicada uma filtragem para remover dados inválidos ou irrelevantes, como entradas de sensores com status "error" ou com leituras inconsistentes.
* **Normalização de Dados**: Para garantir a uniformidade dos dados, foram usados passos como "Select Values" e "String Operations" para ajustar os formatos de tempo, conversão de unidades de temperatura (se necessário) e garantir a consistência nos campos de texto (localização, ID dos sensores, etc.).
* **Análise Temporal e Geográfica**: Para realizar análises por **localização** e por **intervalos de tempo**, os dados foram organizados em blocos temporais e agregados por áreas específicas (bairros ou zonas urbanas). Foi usado o passo "Group By" para calcular médias e tendências (por exemplo, a média diária de poluentes ou temperatura).

**3. Carga dos Dados (Load)**

Após a transformação e limpeza, os dados foram carregados em um formato adequado para análise futura e geração de relatórios. As etapas de carga incluíram:

* **Armazenamento em formato JSON**: Para facilitar o uso dos dados em análises posteriores e garantir a interoperabilidade entre diferentes sistemas, os dados finais foram armazenados no formato JSON, usando o passo "JSON Output" do Pentaho Kettle.
* **Criação de relatórios**: Com os dados já transformados e padronizados, foi possível gerar relatórios que relacionavam a qualidade do ar e as emergências médicas, proporcionando insights sobre como determinadas condições meteorológicas afetam a saúde pública.
* **Integração com sistemas de monitoramento**: A solução permitiu que os dados fossem exportados e integrados com plataformas de monitoramento em tempo real, possibilitando que autoridades de saúde e órgãos ambientais tomem decisões informadas sobre a qualidade do ar e a proteção da saúde pública.

**4. Ferramentas e Tecnologias**

Para realizar essas tarefas, utilizamos as seguintes ferramentas e tecnologias:

* **Pentaho Kettle (PDI)**: Ferramenta principal para construção do fluxo ETL, permitindo a automação e integração de dados.
* **APIs (OpenWeather)**: Para obter dados meteorológicos em tempo real.
* **Sensores IoT**: Fonte dos dados de qualidade do ar.
* **JSON/CSV**: Formatos de dados utilizados para entrada e saída dos dados processados.

Essa estratégia garantiu uma abordagem eficaz para a integração e análise dos dados, possibilitando a obtenção de insights relevantes sobre o impacto da qualidade do ar na saúde pública e a previsão de possíveis cenários críticos.

# Anexos

# Conclusão

# Bibliografia

**Não existem origens no documento atual.**