

**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

Este projeto tem como objetivo implementar um processo ETL (*Extract*, *Transform*, *Load*) para extrair, transformar e carregar dados provenientes de sensores urbanos de qualidade do ar, registos de emergências médicas, e dados meteorológicos. O objetivo 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. As 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 robusto. O processo foi dividido nas seguintes etapas principais:

## 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 **Open Meteo API** para obter dados de previsão do tempo, como temperatura, velocidade do vento e humidade, os quais influenciam a dispersão de poluentes.
* **Registos 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.

## 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 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).

## 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 numa base de dados**: Para facilitar o uso dos dados em análises posteriores e garantir a interoperabilidade entre diferentes sistemas, os dados finais foram armazenados numa base de dados, usando o step "Insert / Update" do Pentaho Kettle.

## 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 (Open Meteo)**: 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.
* Node-red: Ferramenta principal para a construção da UI do fluxo ETL.

Esta 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.

# Transformações

## Dados meteorológicos

Nesta transformação, através do uso de uma API coleto os dados como a temperatura, sensação térmica e humidade para guardar num ficheiro json e num outro ficheiro json guardo os dados das previsões dos próximos dias, sendo que uso o step *Select values* para escolher os dados que mais se adequam à solução.

Uma imagem com texto, captura de ecrã, diagrama, file

Descrição gerada automaticamente

## Dados da Qualidade do ar

Para extrair os dados da qualidade do ar, através do uso de um ficheiro CSV que contem vários dados de vários sensores de qualidade de ar, após escolher os dados que mais se adequam, guardo os mesmos num ficheiro json também.

Uma imagem com texto, captura de ecrã, Tipo de letra, software

Descrição gerada automaticamente

## Registos de emergências medicas

Após aceder a um ficheiro XML que contem os registos de emergências medicas do hospital da cidade, são selecionados todos os campos para serem guardados num outro ficheiro json.

Uma imagem com texto, captura de ecrã

Descrição gerada automaticamente

## Dados para emitir alertas

Para que as autoridades possam emitir alertas para prevenir problemas de saúde na população devido à qualidade do ar, temperatura entre outros, primeiro acedo ao ficheiro da qualidade do ar em json previamente referido, tendo os dados organizo os mesmos por data e filtro apenas pelos sensores que estão ativos, logo de seguida utilizo uma expressão regular (regEx) para remover as horas da data, de seguida organizo outra vez mas por localização agrupando os mesmos para fazer a media dos registos dos sensores para cada zona e junto com os dados do ficheiro json da meteorologia e por fim guardo tudo numa base de dados.

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

## Dados dos registos de emergência causados pela qualidade do ar

Para ter os registos de emergência relevantes para inserir na base de dados para futura análise, extrai mais uma vez o ficheiro json com os dados da qualidade do ar organizei por data, filtrei a condição dos sensores para mostrar apenas os ativos e também pelo AQI que fosse superior a 100, formatei a data retirando as horas da mesma, e juntei aos dados dos relatórios de emergência aos quais foram agrupados também por data e filtrados pelas condições dos pacientes possivelmente causadas pela condição do ar, por fim organizei pelos ID’s dos relatórios e removi informação duplicada.

Uma imagem com diagrama, file, Tipo de letra, captura de ecrã

Descrição gerada automaticamente

## Dados das Estatísticas Apos a junção dos dados referidos no 4.5, organizei a copia dos mesmos por localização fazendo, também, a media da temperatura, humidade e do índice de qualidade do ar para as autoridades terem uma estimativa do risco para a saúde publica em cada região.

Uma imagem com texto, file, diagrama, Tipo de letra

Descrição gerada automaticamente

# Jobs

## Criar esquema e tabelas na BD

Este Job cria um esquema e as tabelas necessárias para todos os dados analisados por meio de scripts de SQL.

Uma imagem com texto, diagrama, Tipo de letra, captura de ecrã

Descrição gerada automaticamente

## Guardar dados em Json

Foi criado um Job que utiliza as transformações previamente mencionadas (4.1, 4.2 e 4.3) para criar os ficheiros Json necessários ao mesmo tempo.

Uma imagem com diagrama, file, texto

Descrição gerada automaticamente

## Inserir dados na base de dados

Nestes Jobs correm as transformações mencionadas no 4.4, 4.5 e 4.6 para inserir os dados na base de dados.

Uma imagem com Tipo de letra, diagrama, file, Gráficos

Descrição gerada automaticamente Uma imagem com Tipo de letra, diagrama, file, design

Descrição gerada automaticamente

# Node red

Foram criados *flows* em *node red* para a visualização por parte das autoridades e outros interessados, de todos os dados obtidos em tempo real.

## Flow dos alertas

Neste flow foi criada uma UI para que as autoridades conseguissem visualizar dados como a temperatura, humidade e a qualidade do ar atual para emitir alertas e também uma lista de localizações mais especificas em estado de alerta a ser emitido também pelas mesmas.

Uma imagem com texto, file, Gráfico, captura de ecrã

Descrição gerada automaticamente

Uma imagem com texto, captura de ecrã, Tipo de letra, círculo

Descrição gerada automaticamente

## Flow das estatísticas

Para as estatísticas foram criados gráficos de barras para mostrar a media de temperatura, qualidade de ar e humidade por localização de forma a alertar as autoridades das zonas mais prováveis a ter problemas de saúde social em relação as métricas analisadas ao longo do tempo.

Uma imagem com texto, file, captura de ecrã, Tipo de letra

Descrição gerada automaticamente  
Uma imagem com texto, captura de ecrã, diagrama, Retângulo

Descrição gerada automaticamente

## Flow dos registos de emergências

Por fim temos o flow dos registos de emergência onde as autoridades consultam casos de cidadãos afetados pela qualidade do ar e ou temperatura de forma a poderem solucionar melhor futuras ocasiões analisando devidamente a informação disponibilizada.

Uma imagem com texto, captura de ecrã, file, Tipo de letra

Descrição gerada automaticamente Uma imagem com texto, número, Tipo de letra, file

Descrição gerada automaticamente

# Conclusão

# Bibliografia

**Não existem origens no documento atual.**