



UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS VII
CENTRO CIÊNCIAS EXATAS E SOCIAIS APLICADAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ADRIER JOSÉ SILVA DOS SANTOS

DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM *PIPELINE* DE
DADOS PARA MONITORAMENTO E NOTIFICAÇÃO DE EVENTOS
CLIMÁTICOS EXTREMOS

PATOS-PB
2025

ADRIER JOSÉ SILVA DOS SANTOS

DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM *PIPELINE* DE
DADOS PARA MONITORAMENTO E NOTIFICAÇÃO DE EVENTOS
CLIMÁTICOS EXTREMOS

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciência da Computação do Centro de Ciências Exatas e Sociais Aplicadas da Universidade Estadual da Paraíba como requisito parcial à obtenção do título de Bacharelado em Ciência da Computação.

Área de concentração: Engenharia de Dados.

Orientador: Prof. Dr. Rodrigo Alves Costa

PATOS-PB

2025

É expressamente proibida a comercialização deste documento, tanto em versão impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que, na reprodução, figure a identificação do autor, título, instituição e ano do trabalho.

S237d Santos, Adrier José Silva dos.

Desenvolvimento e implementação de um *pipeline* de dados para monitoramento e notificação de eventos climáticos extremos [manuscrito] / Adrier José Silva dos Santos. - 2025.
59 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Ciência da computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas, 2025.

"Orientação : Prof. Dr. Rodrigo Alves Costa, Coordenação do Curso de Computação - CCEA".

1. Pipeline de dados. 2. Monitoramento climático. 3. Processamento em tempo real. 4. AWS. 5. Análise de dados climáticos. I. Título

21. ed. CDD 005.7

ADRIER JOSÉ SILVA DOS SANTOS

DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM PIPELINE DE DADOS PARA
MONITORAMENTO E NOTIFICAÇÃO DE EVENTOS CLIMÁTICOS EXTREMOS

Trabalho de Conclusão de Curso
apresentado à Coordenação do Curso
de Ciência da Computação da
Universidade Estadual da Paraíba,
como requisito parcial à obtenção do
título de Bacharel em Ciência da
Computação

Aprovada em: 02/06/2025.

BANCA EXAMINADORA

Documento assinado eletronicamente por:

- **Rodrigo Alves Costa** (***.667.224-**), em **13/06/2025 14:23:15** com chave **1731323e487b11f0bd5c1a7cc27eb1f9**.
- **Francisco Anderson Mariano da Silva** (***.120.084-**), em **13/06/2025 14:23:40** com chave **25d6ecb6487b11f08ba72618257239a1**.
- **Giovanna Trigueiro de Almeida Araújo** (***.352.004-**), em **15/06/2025 11:27:43** com chave **e626c1d849f411f08a581a7cc27eb1f9**.

Documento emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QrCode ao lado ou acesse https://suap.uepb.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Folha de Aprovação do Projeto Final

Data da Emissão: 15/06/2025

Código de Autenticação: 82a83a



Dedico este trabalho à minha família, que sempre esteve ao meu lado com amor, apoio e incentivo incondicional.

Agradeço também aos amigos que fizeram parte dessa caminhada, compartilhando momentos de aprendizado, desafios e conquistas. Esta conquista é fruto de cada palavra de encorajamento, cada gesto de apoio e de todas as experiências que me fortaleceram ao longo dessa jornada.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, pois sem Ele nada disso seria possível. Ele foi minha força e inspiração em momentos de desafios e dificuldades, e a Ele devo minha perseverança e fé em cada etapa dessa jornada. Agradeço também a Nossa Senhora Aparecida, minha intercessora, que sempre me guiou com sua proteção e sabedoria, me oferecendo coragem nas horas de dificuldade e me fortalecendo para continuar. Sua presença constante em minha vida me trouxe paz e confiança para seguir em frente. Aos meus pais, Antônio e Hilda, que sempre me apoiaram e acreditaram em mim, mesmo nos momentos em que eu duvidei de mim mesmo. Agradeço pelo amor incondicional, pela paciência, pelas noites sem dormir e por sempre me incentivarem a seguir em frente. Vocês são minha base, e sem o apoio de vocês, nada disso seria possível. Um dia, retribuirei tudo o que fizeram por mim, com todo o amor que vocês sempre me deram.

Ao meu irmão, Adjunio, e à minha cunhada, Marinalva, minha eterna gratidão por estarem ao meu lado em todos os momentos, oferecendo não só apoio emocional, mas também conselhos valiosos. A amizade e o carinho de vocês são fundamentais para minha jornada e sou grato por cada palavra de incentivo e por sempre acreditarem em mim.

Ao meu afilhado, Emanuel, que trouxe luz e alegria aos meus dias, agradeço por me motivar a ser uma pessoa melhor. Ao ver o seu sorriso e seu crescimento, percebo o quão importante é continuar lutando pelos meus sonhos, para que, um dia, possa inspirá-lo a alcançar os seus.

Aos meus compadres, Camila e Marcos, meu agradecimento especial. Obrigado por serem como uma segunda família para mim. Em momentos de dificuldade, vocês foram verdadeiros amigos, oferecendo apoio emocional, carinho e compreensão. A amizade de vocês é um dos maiores presentes que a vida me deu.

À minha prima, Mariana, por todo o apoio emocional e pelas palavras de força nos momentos em que mais precisei. Você sempre esteve ao meu lado, me ajudando a manter a calma e a seguir em frente, mesmo nos dias mais difíceis. Sua presença fez toda a diferença nessa jornada, e sou imensamente grato por isso.

Ao meu orientador, professor Dr. Rodrigo Alves Costa, expresso minha profunda gratidão. Desde o segundo período do curso, quando a orientação parecia distante, o senhor se tornou uma fonte constante de sabedoria e apoio. Sua disponibilidade, paciência e dedicação foram essenciais para o desenvolvimento deste trabalho. Tenho plena certeza de que o senhor foi um anjo enviado por Deus, sempre disposto a ajudar e a guiar, mesmo nas horas mais difíceis. Sua orientação foi fundamental para o sucesso desse trabalho, e por isso, serei eternamente grato. Sua contribuição não se limitou ao âmbito acadêmico; a sabedoria e os ensinamentos que me proporcionou também foram determinantes para o meu sucesso profissional. Com a sua ajuda, fui capaz de amadurecer e desenvolver as habilidades necessárias para atuar com excelência na minha área, e isso é algo que levarei

para toda a minha carreira.

À toda comunidade acadêmica da UEPB, em especial à turma 2025.1, minha sincera gratidão. Compartilhamos alegrias, desafios e vitórias ao longo dessa jornada, e cada um de vocês contribuiu para meu crescimento acadêmico e pessoal. Foi um prazer imenso compartilhar esse tempo com pessoas tão incríveis e inspiradoras.

Aos meus amigos de jornada, que se tornaram mais do que colegas de curso, mas verdadeiros irmãos. Especialmente, aos meus irmãos Jennyfer e Elder, que estiveram ao meu lado, apoiando, ajudando, dando conselhos e até mesmo oferecendo os indispensáveis "puxões de orelha" quando mais precisei. Cada momento de convivência com vocês foi uma oportunidade de aprendizado e crescimento, e eu sou imensamente grato por isso.

A todos que, de alguma forma, contribuíram com suas ideias, sugestões e críticas construtivas para o desenvolvimento deste trabalho, meu muito obrigado. Cada *feedback* foi valioso para que eu pudesse aprimorar minha pesquisa e alcançar a qualidade que este trabalho reflete.

Por fim, agradeço a todos que estiveram ao meu lado durante essa caminhada, com apoio, paciência e dedicação. Cada um de vocês fez uma diferença significativa em minha vida e carreira, e eu sou eternamente grato por isso.

“Em Deus confiamos. Todos os outros devem trazer dados.” — W. Edwards Deming

RESUMO

Este trabalho apresenta o desenvolvimento de um *pipeline* de dados voltado para o monitoramento climático em tempo real, utilizando tecnologias da Amazon Web Services (AWS) e ferramentas de processamento e análise de dados. O objetivo principal foi construir uma arquitetura robusta que permita a ingestão contínua de dados meteorológicos, seu processamento eficiente e a geração de alertas automáticos para eventos climáticos extremos, além da análise histórica por meio de ferramentas de *Business Intelligence (BI)*. Para isso, foi implementada uma solução baseada em Kinesis Data Streams para captura e transmissão dos dados em tempo real, AWS Lambda para processamento imediato e Amazon Simple Notification Service (SNS) para notificação de alertas. Paralelamente, foi utilizado um *pipeline* em *batch* com armazenamento em Amazon S3, processamento com Apache Spark e armazenamento final em banco de dados PostgreSQL, orquestrado pelo Apache Airflow. O monitoramento da infraestrutura foi realizado com AWS CloudWatch, garantindo rápida identificação e resposta a falhas. Os dados processados foram disponibilizados para análises detalhadas no Metabase, possibilitando consultas como média diária de temperatura, vento médio diário, umidade média e detecção de eventos extremos. A arquitetura medalhão adotada para o processamento em camadas *Bronze*, *Silver* e *Gold* possibilitou organização, limpeza e enriquecimento dos dados, garantindo qualidade e confiabilidade. Os resultados demonstraram a eficácia da arquitetura para cenários que exigem baixa latência e alta escalabilidade, além da importância do monitoramento contínuo para manutenção da integridade do *pipeline*. O projeto contribui para o avanço das técnicas de monitoramento climático em tempo real, com potencial aplicação em diversas áreas que dependem de dados meteorológicos confiáveis.

Palavras-chave: Pipeline de Dados. Monitoramento Climático. Processamento em Tempo Real. AWS. Análise de Dados Climáticos

ABSTRACT

This paper presents the development of a data pipeline for real-time climate monitoring, leveraging Amazon Web Services (AWS) technologies and data processing and analysis tools. The main objective was to build a robust architecture enabling the continuous ingestion of meteorological data, its efficient processing, and the generation of automatic alerts for extreme weather events, in addition to historical analysis through Business Intelligence (BI) tools. To achieve this, a solution based on Kinesis Data Streams was implemented for real-time data capture and transmission, AWS Lambda for immediate processing, and Amazon Simple Notification Service (SNS) for alert notifications. In parallel, a batch pipeline was utilized with storage in Amazon S3, processing with Apache Spark, and final storage in a PostgreSQL database, orchestrated by Apache Airflow. Infrastructure monitoring was carried out with AWS CloudWatch, ensuring rapid identification and response to failures. The processed data was made available for detailed analysis in Metabase, enabling queries such as daily average temperature, daily average wind, average humidity, and detection of extreme events. The medallion architecture adopted for layered processing (Bronze, Silver, and Gold) allowed for data organization, cleaning, and enrichment, ensuring quality and reliability. The results demonstrated the architecture's effectiveness for scenarios requiring low latency and high scalability, as well as the importance of continuous monitoring for maintaining pipeline integrity. This project contributes to the advancement of real-time climate monitoring techniques, with potential applications in various areas that depend on reliable meteorological data.

Keywords: Data Pipeline. Climate Monitoring. Processing at Real-time. AWS. Climate Data Analysis.

LISTA DE ILUSTRAÇÕES

1	Arquitetura medalhão.	32
2	Arquitetura do Fluxo Experimental.	34
3	Arquitetura do <i>pipeline</i> de dados.	36
4	Função AWS Lambda responsável pela requisição da API Tomorrow. . . .	37
5	Trecho de código da função Lambda.	37
6	Exemplo de resposta da API.	38
7	Camada Bronze.	38
8	Camada <i>Silver</i>	38
9	Camada <i>Gold</i>	39
1	<i>Broker</i> responsável por receber os dados climáticos.	41
2	Métricas do <i>Broker</i>	41
3	Função Lambda responsável pela análise de eventos climáticos extremos. .	41
4	Notificação de alerta enviada por <i>e-mail</i> via Amazon SNS.	42
5	Painel do AWS CloudWatch utilizado para o monitoramento das métricas da aplicação.	45
6	Painel do Apache Airflow utilizado para o monitoramento da aplicação. . .	45
7	Visão geral dos custos dos serviços AWS durante a execução do pipeline. .	47
8	<i>Dashboard</i> no Metabase mostrando a média diária de temperatura.	48
9	Visualização no Metabase com o vento médio registrado por dia.	48
10	Gráfico de distribuição de temperaturas registrado ao longo dos períodos. .	49
11	Consulta SQL no Metabase com a umidade média registrada diariamente. .	50

LISTA DE TABELAS

5.1	Cr�terios para Identifica�o de Eventos Clim�ticos Extremos	44
-----	--	----

SUMÁRIO

	Página
1	INTRODUÇÃO 13
1.1	Problemática 13
1.2	Objetivos 14
1.2.1	Objetivo Geral 14
1.2.2	Objetivos Específicos 14
1.3	Justificativa 14
1.4	Metodologia 15
1.5	Organização do Trabalho 15
2	FUNDAMENTAÇÃO TEÓRICA 17
2.1	Introdução ao Monitoramento Climático 18
2.2	Arquitetura de <i>Pipelines</i> de Dados 19
2.3	Tecnologias para Pipelines de Dados em Tempo Real 20
2.3.1	Ferramentas e Tecnologias de Ingestão de Dados em Tempo Real 20
2.3.2	Ferramentas para <i>Streaming</i> de Dados 20
2.4	Armazenamento e Processamento de Dados Climáticos 21
2.5	Sistemas de Banco de Dados e Data Warehouses 22
2.6	Integração para Notificação e Ação em Tempo Real 23
2.7	Segurança e Controle de Acesso em Ambientes de Dados na Nuvem 23
2.7.1	Gerenciamento de Identidade e Acesso (IAM) na AWS 24
2.8	Coleta e Integração de Dados Climáticos 24
2.8.1	Fontes de Dados Climáticos 25
2.8.2	Métodos para Garantir a Precisão e Integridade dos Dados Climáticos . . . 25
2.8.3	Uso de APIs e Integração de Múltiplas Fontes de Dados 26
2.8.4	Desafios na Coleta e Integração de Dados Climáticos 27
2.9	Estudos de Caso e Aplicações de Sistemas de Alerta Climático . 28
3	METODOLOGIA 30
3.1	Tipo de Pesquisa 30
3.2	Procedimentos Metodológicos 30
3.3	Arquitetura do <i>Pipeline</i> de Dados 31
3.4	Arquitetura medalhão 31
3.5	Escolha das Tecnologias e Justificativa 32
3.6	Fluxo de Dados 34
3.6.1	Fluxo Experimental 34
3.7	Instrumentos de validação 35

4	DESENVOLVIMENTO DA APLICAÇÃO	36
4.1	Coleta de Dados e Integração com APIs Meteorológicas	36
4.2	Armazenamento e Governança dos Dados	37
5	ANÁLISE E DISCUSSÃO	40
5.1	Processamento em Tempo Real com Kinesis Data Streams	40
5.1.1	Aquisição e Transmissão Contínua de Dados com o Kinesis Data Streams .	40
5.1.2	Processamento em Tempo Real com AWS Lambda	40
5.1.3	Geração e Envio de Alertas com Amazon SNS	42
5.1.4	Benefícios da Arquitetura em Tempo Real	42
5.2	Processamento em Lote com S3, PostgreSQL e Airflow	42
5.2.1	Parâmetros para a detecção de eventos climáticos extremos	44
5.3	Monitoramento e Manutenção do Pipeline	44
5.4	Indicadores de Custos da Arquitetura na AWS	46
5.4.1	Modelos de Cobrança dos Serviços Utilizados	46
5.4.2	Estratégias para Otimização de Custos	46
5.4.3	Visualização dos Indicadores de Custos	46
5.5	Análises de Dados com Metabase	47
5.5.1	Média Diária de Temperatura	47
5.5.2	Vento Médio Diário	47
5.5.3	Distribuição de Temperaturas por Período	48
5.5.4	Análise de Umidade Média por Dia	49
5.6	Considerações Finais sobre o Desenvolvimento	49
5.7	Limitações do Projeto	50
6	CONCLUSÕES	52
6.1	Trabalhos Futuros	52
	REFERÊNCIAS	52

1 INTRODUÇÃO

Nos últimos anos, eventos climáticos extremos, como ondas de calor, secas e inundações, têm se tornado mais frequentes e severos devido ao aquecimento global e ao aumento das emissões de gases de efeito estufa, como o dióxido de carbono (CO₂). A concentração desses gases já ultrapassa os limites considerados seguros para o planeta, indicando que esses eventos tendem a ser ainda mais recorrentes. Apesar disso, há uma carência de sistemas capazes de monitorar e notificar essas ocorrências em tempo real. Essa ausência de ferramentas tecnológicas limita a capacidade de resposta rápida, agravando os impactos sociais e econômicos. Nesse contexto, torna-se essencial desenvolver soluções baseadas em dados, como *pipelines* automatizados, que possibilitem o monitoramento contínuo e a emissão de alertas, contribuindo para a mitigação dos danos causados por desastres climáticos.

Nesse cenário, torna-se fundamental o desenvolvimento de tecnologias que possibilitem o monitoramento climático e a notificação em tempo real de eventos extremos. Essas tecnologias têm o potencial de fornecer informações críticas para uma resposta rápida e eficaz, reduzindo o número de vítimas e os danos associados a esses fenômenos. Por exemplo, o uso de sistemas de monitoramento remoto, como o Cyclone Global Navigation Satellite System (CYGNSS), tem mostrado grande potencial na detecção de inundações e na análise de mudanças na cobertura de áreas úmidas, contribuindo para uma melhor compreensão das dinâmicas climáticas (Zhang *et al.*, 2021). O CYGNSS, que foi projetado inicialmente para medir a velocidade do vento sobre os oceanos, também demonstrou eficácia em monitorar a umidade do solo e a dinâmica de inundações em ambientes terrestres (Ruf *et al.*, 2018).

Em decorrência ao avanço das ferramentas de coleta e processamento de dados, como *pipelines* de dados e sistemas automatizados, a criação de soluções para monitoramento climático contínuo torna-se cada vez mais viável e necessária para auxiliar na mitigação dos impactos causados pelas mudanças climáticas.

1.1 Problemática

Assim como destacam Monteiro e Zanella (2023), a ausência de informações precisas e em tempo real limita a capacidade de adaptação e mitigação frente a desastres naturais. A falta de sistemas eficientes para monitoramento e notificação de eventos climáticos extremos representa um desafio significativo para a proteção de populações em países emergentes, como o Brasil. Embora o aumento da intensidade e frequência de desastres naturais, como tempestades, inundações e ondas de calor, seja amplamente reconhecido, a infraestrutura para acompanhar e comunicar esses eventos de forma precisa e em tempo real ainda é insuficiente. Esse déficit resulta em respostas tardias e inadequadas, aumen-

tando o número de vítimas e os danos materiais. Refletindo a gravidade desse cenário, em julho de 2023, o secretário-geral da Organização das Nações Unidas (ONU), António Guterres, alertou que entramos na era da “ebulição global”, com níveis de (CO₂) ultrapassando os limites seguros para o planeta (Sato, 2024), o que reforça a urgência na busca por soluções tecnológicas que auxiliem na mitigação dos impactos das mudanças climáticas.

Diante desse cenário, a problemática que orienta este estudo é: **Como desenvolver e implementar um *pipeline* de dados capaz de monitorar e notificar, em tempo real, eventos climáticos extremos, mitigando os impactos desses fenômenos na população em geral?**

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo desse estudo é o desenvolvimento e implementação de um *pipeline* de dados para monitorar eventos climáticos extremos e emitir alertas automáticos para autoridades e população sobre possíveis anomalias climáticas, visando melhorar a capacidade de resposta e mitigação de riscos.

1.2.2 Objetivos Específicos

- **Coletar e integrar dados:** Consolidar dados de diversas fontes, como Interface de Programação de Aplicações (API) de institutos confiáveis, estações meteorológicas e modelos climáticos, para garantir uma base de dados ampla e confiável.
- **Desenvolver algoritmos de análise:** Criar algoritmos capazes de identificar padrões e anomalias nos dados climáticos, visando detectar e prever a ocorrência de eventos extremos.
- **Implementar um sistema de alertas:** Desenvolver um sistema de notificação para alertar autoridades e população sobre a iminência de eventos climáticos extremos, permitindo uma resposta rápida e eficaz.

1.3 Justificativa

Este trabalho é justificado pela necessidade de sistemas eficazes para monitorar e notificar eventos climáticos extremos, especialmente diante da vulnerabilidade de diversas comunidades a esses problemas. Com o avanço das tecnologias de *Big Data* e automação, abre-se uma oportunidade significativa para o desenvolvimento de *pipelines* de dados que permitam o monitoramento em tempo real e a emissão de alertas automáticos para autoridades e a população em geral. Esses sistemas podem ter um papel fundamental na redução de riscos, possibilitando ações preventivas que minimizam perdas e protegem

vidas. Além disso, do ponto de vista acadêmico, o desenvolvimento de um *pipeline* de dados para monitoramento climático e análise de anomalias climáticas contribui para o avanço de pesquisas na área de Engenharia de Dados e Ciência de Dados aplicada ao meio ambiente, enquanto, no aspecto social, representa uma contribuição importante para a proteção da população contra as ameaças climáticas.

1.4 Metodologia

A metodologia seguida neste trabalho foi organizada em etapas práticas e sequenciais, com o intuito de desenvolver um *pipeline* de dados. O primeiro passo envolveu uma análise dos requisitos do projeto, considerando as necessidades de coleta, processamento e armazenamento dos dados. Com base nessa análise, foram selecionadas as tecnologias e ferramentas mais apropriadas para garantir a escalabilidade e o desempenho adequados. O desenvolvimento ocorreu em fases distintas, abrangendo a implementação das etapas de coleta, transformação e armazenamento dos dados, além da automação e orquestração das tarefas.

1.5 Organização do Trabalho

Este trabalho está organizado da seguinte forma:

- **Capítulo 1 - Introdução:** Apresenta o contexto do projeto, a problemática, os objetivos, a justificativa, a metodologia e a organização do trabalho.
- **Capítulo 2 - Fundamentação Teórica:** Aborda os conceitos teóricos e as tecnologias essenciais para o monitoramento climático e desenvolvimento de *pipelines* de dados. São apresentados temas como arquitetura de dados, ETL/ELT, bancos de dados, processamento em tempo real, mensageria, segurança em ambientes na nuvem, além de desafios na coleta e integração de dados climáticos.
- **Capítulo 3 - Metodologia:** Detalha o processo metodológico adotado, descrevendo o tipo de pesquisa, os procedimentos, o fluxo de dados, a arquitetura do pipeline baseada no modelo medalhão, e a justificativa para a escolha das ferramentas e tecnologias.
- **Capítulo 4 - Desenvolvimento:** Descreve detalhadamente a implementação do *pipeline* de dados, incluindo a arquitetura desenvolvida, os processos de ingestão de dados via APIs meteorológicas, o armazenamento dos dados na nuvem e os mecanismos de governança adotados.
- **Capítulo 5 - Análise e Discussão:** Apresenta a execução e funcionamento do pipeline, detalhando tanto o processamento em tempo real, utilizando Kinesis, Lambda e SNS, quanto o processamento em lote com S3, PostgreSQL e Airflow.

São discutidos também os parâmetros de detecção de eventos climáticos extremos, as estratégias de monitoramento e manutenção do pipeline e os dashboards desenvolvidos no Metabase. Este capítulo inclui ainda a avaliação dos resultados obtidos, as limitações identificadas e considerações sobre a eficácia da solução proposta.

- **Capítulo 6 - Conclusões e Trabalhos Futuros:** Resume os principais pontos abordados, as conclusões alcançadas, e propõe melhorias, evoluções para o pipeline e possibilidades de estudos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos e tecnologias essenciais para o monitoramento climático e a construção de *pipelines* de dados, fundamentais para os tópicos abordados nos capítulos seguintes. Na Seção 2.1, será discutida a importância do monitoramento climático, destacando sua relevância para a segurança pública, preservação ambiental e formulação de políticas públicas. Também serão abordados os desafios relacionados à coleta de dados em tempo real, como a necessidade de infraestrutura robusta e a diversidade de fontes de dados climáticos.

A Seção 2.2 explora a arquitetura de *pipelines* de dados, abordando o ciclo de vida desses *pipelines*, os tipos mais comuns (*batch*, *streaming* e híbridos) e sua importância para a ingestão e processamento de grandes volumes de dados climáticos. Em seguida, a Seção 2.3 examina as principais tecnologias utilizadas em *pipelines* de dados em tempo real, incluindo ferramentas como Apache Kafka e Amazon Kinesis, com foco nas capacidades de integração e escalabilidade dessas tecnologias.

A Seção 2.4 aborda estratégias de armazenamento e processamento de dados climáticos, destacando o uso de serviços escaláveis, como AWS S3, AWS Lambda e Apache Spark, para armazenar grandes volumes de dados e processar eventos em tempo real, facilitando a análise e a identificação de eventos meteorológicos extremos.

A Seção 2.5 explora sistemas de banco de dados e data warehouses para armazenamento e consulta de dados climáticos, com ênfase no PostgreSQL para dados geoespaciais e temporais e no Snowflake, que otimiza custos e desempenho na análise de grandes volumes de dados. A Seção 2.6 discute a integração de sistemas para notificações em tempo real, destacando o uso do Amazon SNS para envio de alertas automatizados e do Amazon Kinesis para gerenciar fluxos de dados em tempo real em situações críticas.

A Seção 2.7 aborda a segurança em ambientes de dados na nuvem, destacando o uso do IAM da AWS para controle de acesso, gestão de identidades e permissões. Enfatiza práticas como menor privilégio, autenticação multifator, rotação de credenciais e auditoria com CloudTrail, essenciais para garantir segurança, rastreabilidade e conformidade em *pipelines* de dados.

A Seção 2.8 foca na coleta e integração de dados climáticos, abordando as fontes de dados disponíveis, como APIs de serviços meteorológicos, e os métodos para garantir a precisão e a integridade das informações. O foco está no uso de tecnologias para extração, validação e normalização dos dados, além da importância de integrar diversas fontes para análises mais robustas. A seção também discute os desafios enfrentados nesse processo, como variações climáticas, influências locais e regionais, e obstáculos tecnológicos.

Por fim, a Seção 2.9 aborda estudos de caso sobre sistemas de alerta climático, destacando como a integração entre tecnologia, políticas públicas e engajamento comunitário

contribui para mitigar riscos. Mostra que a combinação de alertas móveis, conscientização, infraestrutura adaptativa e colaboração entre setores é essencial para fortalecer a resiliência frente a eventos climáticos extremos.

2.1 Introdução ao Monitoramento Climático

No mundo moderno, o monitoramento climático se tornou uma abordagem indispensável para compreender e enfrentar os desafios enfrentados pelas mudanças climáticas, sendo essencial para a preservação de ecossistemas e a segurança das populações. Esse processo envolve a coleta e análise de dados sobre variáveis como temperatura, precipitação, umidade e velocidade do vento. O principal objetivo do monitoramento é prever como essas variáveis se comportam ao longo do tempo, ajudando na compreensão das mudanças climáticas e na identificação de tendências que podem impactar ecossistemas, infraestrutura e segurança pública. Esse monitoramento é fundamental para criar políticas públicas, gerir recursos naturais, preservar vidas e mitigar desastres ambientais (Chaves *et al.*, 2018).

Eventos climáticos extremos são fenômenos que acontecem com maior frequência ou intensidade do que o normal, causando grandes impactos tanto para o meio ambiente quanto para a sociedade. Esses eventos incluem secas, inundações, tempestades, ondas de calor, ciclones e furacões, e sua ocorrência tem se tornado mais frequente nos últimos anos, principalmente devido às mudanças climáticas provocadas pelas atividades humanas (Sena e Barcellos, 2019; Dias e Pessoa, 2020). Além disso, a literatura aponta que as anomalias climáticas muitas vezes são impulsionadas por fatores como o desmatamento e a urbanização (Sena e Barcellos, 2019; Dias e Pessoa, 2020).

Com isso, os desafios no monitoramento climático em tempo real ganham ainda mais destaque, pois é preciso ter uma infraestrutura robusta que consiga lidar com o grande volume de dados gerados por diversos observatórios climáticos. Coletar dados climatológicos em tempo real exige o uso de sensores e dispositivos que precisam funcionar em condições ambientais variadas e, muitas vezes, adversas. Como aponta Correia (2024), isso demanda um grande investimento em tecnologia e manutenção, além de um planejamento cuidadoso para garantir que os dados sejam coletados de forma consistente e confiável. A diversidade de fontes de dados também torna o processo de captura mais árduo, o que torna ainda mais importante a criação de um sistema eficiente para emitir alertas quando necessário.

Além disso, é fundamental ter um sistema ágil para emitir alertas em tempo real. A capacidade de reagir rapidamente a eventos climáticos extremos pode ser decisiva para evitar danos maiores e minimizar seus impactos. Não basta apenas coletar os dados; é necessário analisá-los de forma eficaz e em tempo hábil, como: permitindo que a população receba informações precisas em tempo real. Portanto, os sistemas de alerta precoce devem ser integrados às tecnologias de monitoramento, garantindo que as informações sobre

riscos climáticos sejam disseminadas de maneira eficaz e que as comunidades possam se preparar adequadamente para enfrentar esses desafios (Correia, 2024).

2.2 Arquitetura de *Pipelines* de Dados

Um *pipeline* de dados é um conjunto estruturado de processos que permite a movimentação, transformação e armazenamento de dados de maneira eficiente e automatizada (Amazon Web Services, 2024). A arquitetura de um *pipeline* de dados geralmente envolve três etapas principais: extração, transformação e carregamento. Essas etapas formam um processo conhecido na área de dados como ETL (*Extract, Transform, Load*). No cenário de monitoramento de eventos climáticos extremos, a arquitetura de um *pipeline* de dados é essencial para integrar dados provenientes de diversas fontes, como sensores meteorológicos, dados históricos de clima e sistemas de alerta, a fim de analisar padrões e prever eventos como tempestades, ondas de calor e chuvas intensas.

A importância de um *pipeline* na Engenharia de Dados está em sua capacidade de integrar dados de várias fontes, o que é fundamental para a análise de grandes volumes de informações. No contexto do monitoramento de eventos climáticos extremos, essa integração se torna ainda mais crucial, pois exige a análise de dados de diferentes origens, como estações meteorológicas, sistemas de alerta e bancos de dados históricos. Além disso, a automação desses processos permite melhorar a eficiência e reduzir erros, garantindo que os dados sejam processados de forma íntegra, proporcionando análises confiáveis e consistentes (Haase *et al.*, 2022).

No “universo” da Engenharia de Dados, existem três tipos principais de *pipelines*: *batch*, *streaming* e híbrido. Cada uma dessas abordagens é utilizada em contextos diferentes, e sua aplicação varia conforme os requisitos do sistema de monitoramento e notificação de eventos climáticos extremos. O pipeline do tipo *batch* é apropriado para análises de dados em “lotes”, ou seja, dados agrupados e processados periodicamente, geralmente em intervalos de tempo predefinidos (Amazon Web Services, 2024). Ele é amplamente utilizado em tarefas como atualizações de relatórios e cargas de dados realizadas durante a madrugada, reduzindo possíveis impactos nos bancos de produção e aproveitando períodos de menor demanda nos sistemas de uma determinada organização. Em projetos relacionados a eventos climáticos, o modelo *batch* pode ser utilizado para processar dados históricos ou realizar análises climáticas, como a previsão de padrões em um determinado intervalo de tempo.

Já o modelo *streaming* de dados, refere-se à emissão contínua e em alta frequência de informações, projetada para viabilizar um processamento ágil e com mínima latência (Amazon Web Services, 2024). Esse tipo de *pipeline* é o ideal em cenários onde a latência mínima é crítica, como no monitoramento de eventos climáticos, monitoramento de fraudes, análise de redes sociais, processamento de logs de sistemas e dados provenientes de sensores. No caso de eventos climáticos extremos, o *pipeline* de *streaming* permite

que dados de sensores meteorológicos, como temperatura, umidade e pressão atmosférica, sejam processados em tempo real, possibilitando a emissão de alertas imediatos sobre tempestades ou outras condições climáticas críticas.

Por fim, existe a arquitetura híbrida, que combina o melhor dos dois mundos, proporcionando análises em lotes de dados e também de forma contínua. Ela é especialmente útil em organizações que precisam de análises em tempo real para ações imediatas, mas também de processamento em lotes para tarefas mais complexas, como análises históricas e previsões. Em sistemas de monitoramento de eventos climáticos extremos, a arquitetura híbrida permite combinar dados históricos para realizar previsões, análises de longo prazo, e também a emissão de alertas através de dados em *real time*.

2.3 Tecnologias para Pipelines de Dados em Tempo Real

Haustein *et al.* (2016) destacam que a análise em tempo real desempenha um papel fundamental ao fornecer informações essenciais para compreender a relação entre eventos climáticos e mudanças climáticas, facilitando respostas rápidas e bem fundamentadas. Além disso, acabaram reduzindo os impactos ambientais e sociais. Cavalcante *et al.* (2023) enfatizam que um sistema de monitoramento climático requer uma arquitetura robusta para lidar com grandes volumes de dados provenientes de sensores meteorológicos, estações de monitoramento e APIs externas. Nesse contexto, é fundamental a adoção de tecnologias que suportem cada etapa do *pipeline* de dados, abrangendo desde a ingestão e o armazenamento até a integração com bancos de dados e *data warehouses*.

2.3.1 Ferramentas e Tecnologias de Ingestão de Dados em Tempo Real

A ingestão é a etapa inicial de um *pipeline*, sendo responsável por capturar informações em tempo imediato provenientes de diversas fontes de dados. Em um sistema de monitoramento climático, essas fontes podem ser dados de sensores meteorológicos, satélites e APIs que fornecem dados como temperatura, pressão atmosférica, velocidade do vento, umidade e o índice de raios ultravioleta (UV) e ultravioleta B (UVB). Essas ferramentas devem ser robustas e escaláveis para que possam realizar o processamento desses dados em alta frequência de maneira eficiente e escalável (Silva e González, 2016).

Assim como elucida Pelle (2023), o Apache Kafka e o Amazon Kinesis estão entre as plataformas mais difundidas para *streaming* de dados.

2.3.2 Ferramentas para *Streaming* de Dados

- **Apache Kafka:** É uma plataforma de *streaming* de dados de código aberto que tem ganhado crescente popularidade devido à sua habilidade de processar grandes volumes de dados em tempo real. Ele foi desenvolvido pelo LinkedIn, o Kafka é um sistema de mensagens baseado no modelo de publicação-assinatura, que permite

a comunicação seja eficiente entre diferentes aplicações e serviços (Wang *et al.*, 2015). A arquitetura dessa tecnologia é projetada para ser distribuída e escalável, permitindo que ele gerencie um elevado volume de eventos simultâneos, o que facilita a ingestão e o processamento de dados em tempo real (Zhao *et al.*, 2015; Taranov *et al.*, 2022).

- **Amazon Kinesis:** É uma solução gerenciada da Amazon Web Services (AWS) para *streaming* de dados, permitindo que os desenvolvedores capturem, processem e analisem grandes volumes de dados em tempo real. Ela facilita a criação de aplicações que demandam processamento contínuo, como monitoramento de eventos, análises em tempo real e integração de dados de diversas fontes (Li *et al.*, 2016). O Kinesis é projetado para lidar com dados de diversas fontes, como sensores meteorológicos, logs de aplicativos e interações de usuários, permitindo respostas rápidas a eventos e tendências emergentes (Alhilal *et al.*, 2020). Além disso, oferece integração direta com outros serviços da AWS, como S3, Lambda e Redshift (Amazon Web Services, 2024). O Amazon Kinesis é composto por diferentes serviços, entre eles o Kinesis Data Streams, que permite o processamento granular e em tempo real de fluxos contínuos de dados (Amazon Web Services, 2024).

Através da utilização dessas ferramentas, é possível garantir que os dados climatológicos sejam capturados de forma confiável e contínua, permitindo o monitoramento e a análise em tempo real.

2.4 Armazenamento e Processamento de Dados Climáticos

Após a ingestão, os dados precisam ser armazenados e organizados para permitir análises rápidas e acessíveis, a AWS disponibiliza de serviços acessíveis e escaláveis para diversos casos, entre eles no armazenamento de dados obtidos através de APIs e centros meteorológicos.

- **AWS S3:** É um serviço de armazenamento de objetos que proporciona escalabilidade, alta disponibilidade, segurança e desempenho de ponta (Amazon Web Services, 2024). Ele é amplamente usado para armazenar grandes volumes de dados e para lidar com arquivos semiestruturados, como logs de sensores de observatórios climáticos, dados semiestruturados de APIs e imagens de satélite. Além disso, o S3 é ideal para armazenar dados históricos que podem ser utilizados em análises de tendências e previsões climáticas.
- **AWS Lambda:** É um serviço de computação ideal para aplicações que necessitam aumentar rapidamente a capacidade de processamento e reduzir para zero quando não estão em uso (Amazon Web Services, 2024). O Lambda é um serviço de computação ideal para aplicações que necessitam aumentar rapidamente a capacidade de

processamento e reduzir a escala para zero quando não estão em uso (Amazon Web Services, 2024). No contexto climático, o processamento dos dados pode ser realizado, já que o Lambda é uma solução *serverless* que permite a execução de funções de processamento sob demanda. Ele pode ser utilizado para consultar APIs, processar eventos gerados, detectar anomalias e encaminhar os resultados para sistemas de alerta.

- **Apache Spark:** É um framework de computação distribuída de código aberto projetado para o processamento eficiente de grandes volumes de dados (Apache Software Foundation, 2025). Ele é amplamente utilizado devido à sua capacidade de realizar análises de dados em alta velocidade, aproveitando a computação em memória. A arquitetura do Spark permite que ele opere significativamente mais rápido do que métodos antigos, com desempenho até cem vezes superior ao do Hadoop em operações de I/O (Ashkouti e Khamforoosh, 2023). Isso se deve ao seu uso de conjuntos de dados distribuídos resilientes (RDDs), que permitem que os dados sejam mantidos na memória, minimizando o tempo de leitura e escrita em disco (Aghbari *et al.*, 2020).

2.5 Sistemas de Banco de Dados e Data Warehouses

Os bancos de dados e *data warehouses* desempenham papéis cruciais no armazenamento e consulta de grandes volumes de dados de maneira estruturada e escalável.

- **PostgreSQL:** É um sistema de gerenciamento de banco de dados relacional (SGBD) de código aberto amplamente utilizado para armazenar e gerenciar dados, sendo uma solução confiável para armazenar dados geoespaciais e temporais, como a localização de estações meteorológicas e padrões climáticos regionais. O PostgreSQL suporta uma ampla gama de tipos de dados, incluindo dados estruturados e não estruturados, e oferece recursos avançados como transações ACID, suporte a JSON, replicação e extensões, como o PostGIS para dados geoespaciais (Merlo, 2024; Carlo *et al.*, 2019). Ele também permite a utilização de consultas avançadas que cruzam informações de tempo e espaço, essenciais para análises detalhadas de eventos climáticos.
- **Snowflake:** É uma plataforma de *data warehousing* na nuvem, projetada para permitir a análise de grandes volumes de dados de maneira escalável e eficiente. Sua arquitetura separa o armazenamento da computação, o que resulta na otimização de custos e no aprimoramento do desempenho em cenários de análise de dados. A sua arquitetura separa armazenamento e computação, otimizando custos e desempenho, permitindo que as organizações escalem recursos sem custos desnecessários

(Chinnathambi, 2024), tornando-o uma das escolhas mais acertadas para o armazenamento de dados analíticos climatológicos, como tendências de temperatura ou padrões de precipitação ao longo do tempo.

2.6 Integração para Notificação e Ação em Tempo Real

A integração dos sistemas de ingestão, armazenamento e processamento deve incluir notificações automáticas em tempo real, permitindo ações rápidas em situações críticas, como emergências climáticas.

- **Amazon SNS:** É um serviço gerenciado que possibilita a entrega de mensagens de publicadores para assinantes (também chamados de produtores e consumidores) (Amazon Web Services, 2024). Pode ser utilizado para o envio de alertas críticos referentes a anomalias climáticas, tanto para autoridades quanto para a população em geral.
- **Amazon Kinesis:** Já mencionado na seção *Ferramentas para Streaming de Dados*, é uma solução gerenciada pela AWS para captura, processamento e análise de grandes volumes de dados em tempo real, permitindo ações rápidas em resposta a eventos emergentes, como os relacionados ao clima.

2.7 Segurança e Controle de Acesso em Ambientes de Dados na Nuvem

Com a crescente adoção da computação em nuvem para o desenvolvimento de soluções baseadas em dados, surge a necessidade de estratégias robustas para proteção das informações sensíveis, como as relacionadas a dados climáticos. Em arquiteturas modernas de pipelines de dados, que incluem ingestão, armazenamento e processamento em tempo real, é imprescindível garantir que apenas entidades autorizadas tenham acesso aos recursos da infraestrutura em nuvem. Segundo Fernandes e Souza (2022), os principais pilares da segurança da informação em nuvem são a confidencialidade, integridade, disponibilidade e o controle de acesso. Nesse contexto, o gerenciamento eficaz de identidades e permissões se torna uma ferramenta essencial para mitigar riscos de acesso indevido, vazamentos de dados e interrupções não autorizadas em fluxos críticos.

Além disso, a governança e a rastreabilidade das ações são fatores importantes para atender às exigências de auditoria e conformidade. Nesse aspecto, destaca-se a necessidade de alinhamento com a **Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018)**, que estabelece diretrizes para o tratamento de dados pessoais, impondo responsabilidades às organizações quanto à privacidade, segurança e uso ético dos dados. A ausência de políticas claras de controle de acesso e identidade pode comprometer todo o ciclo de vida de dados sensíveis, especialmente em aplicações que envolvem monitoramento contínuo e decisões automatizadas (Lima *et al.*, 2021).

2.7.1 Gerenciamento de Identidade e Acesso (IAM) na AWS

O Identity and Access Management (IAM) é uma das soluções mais adotadas para controle de acesso baseado em identidade em ambientes de nuvem pública. Ele permite criar e gerenciar usuários, grupos, políticas e funções, possibilitando que cada entidade tenha acesso somente aos recursos necessários. De acordo com a Amazon Web Services (2024), o IAM oferece a possibilidade de gerenciar o acesso de forma segura aos serviços e recursos da AWS. Esse controle é realizado por meio de políticas no formato JSON, que determinam as ações permitidas ou negadas a cada identidade. Tais políticas podem ser aplicadas a usuários humanos, aplicações automatizadas ou serviços internos da própria AWS.

O uso de IAM se destaca em pipelines de dados que envolvem múltiplos serviços da AWS, como Amazon S3 (armazenamento), Amazon Kinesis (streaming), AWS Lambda (processamento sem servidor) e Amazon SNS (notificações). A configuração correta do IAM garante segurança operacional, permite rastrear acessos por meio de logs (como o CloudTrail) e auxilia na manutenção da conformidade com boas práticas de governança de dados e com os princípios da LGPD (Amazon Web Services, 2024).

Entre as boas práticas no uso do IAM destacam-se:

- Princípio do menor privilégio;
- Uso de autenticação multifator (MFA);
- Rotação periódica de credenciais;
- Auditoria contínua com AWS CloudTrail;
- Separação de funções (por exemplo, entre ambientes de produção e desenvolvimento).

Essas práticas são fundamentais para garantir a segurança, a privacidade e a conformidade legal em ambientes de dados na nuvem (Amazon Web Services, 2024).

2.8 Coleta e Integração de Dados Climáticos

A coleta e integração de dados referentes ao clima desempenham um papel de extrema importância em projetos que dependem de informações meteorológicas para a análise e tomada de decisão. Assim como argumenta, Esses processos envolvem a seleção de diversas fontes confiáveis, o uso de tecnologias para extração de dados e a aplicação de métodos que assegurem a qualidade e integridade dessas informações obtidas. A integração bem-sucedida dessas bases de conhecimento é essencial para fornecer análises precisas, especialmente na detecção e monitoramento de eventos climáticos extremos.

2.8.1 Fontes de Dados Climáticos

Silva e Barbosa (2022) elucidam que a utilização de dados meteorológicos recentes, acessíveis por meio de APIs (Application Programming Interfaces), podem auxiliar em diversas áreas, permitindo a tomada de decisões com base em informações climáticas precisas. Dentre as APIs mais reconhecidas, destaca-se a NOAA (National Oceanic and Atmospheric Administration), que disponibiliza um imenso acervo de dados meteorológicos históricos e em tempo real, incluindo tendências de temperatura, precipitação e eventos extremos (National Oceanic and Atmospheric Administration, 2024). Outra API popular é a OpenWeather, que fornece acesso a previsões detalhadas, mapas climáticos e dados históricos (OpenWeather, 2024).

Além dessas, há diversas APIs conhecidas no mercado, como a Climacell (atual Tomorrow.io), que oferece dados hiperlocais e modelos preditivos avançados (Tomorrow.io, 2024). A escolha da API mais adequada depende do caso de uso do projeto, pois ela deve considerar fatores como custo, limite de requisições, precisão e a cobertura geográfica desses dados.

2.8.2 Métodos para Garantir a Precisão e Integridade dos Dados Climáticos

A confiabilidade dos dados climáticos é essencial para análises precisas e tomadas de decisão. Para isso, é necessário implementar estratégias que garantam a precisão e integridade dessas informações, são elas:

- **Validação de Dados:** A validação é um processo que permite comparar os dados obtidos de diferentes fontes para identificar divergências ou valores inconsistentes. Por exemplo, diferenças significativas entre leituras de temperatura para a mesma localidade podem indicar problemas na coleta ou transmissão dos dados. Esse processo é crucial para garantir a integridade dos dados e evitar a introdução de erros que possam comprometer a análise e a interpretação (Oliveira *et al.*, 2020; Medeiros *et al.*, 2022).
- **Normalização:** A normalização é uma etapa utilizada para garantir que os dados estão homogêneos, evitando assim conflitos entre esses dados, por exemplo: temperaturas em graus Celsius ou Fahrenheit e velocidades do vento em km/h ou m/s. Moraes *et al.* (2020) destacam que a normalização é fundamental em contextos onde os dados apresentam diferentes ordens de grandeza, sendo essencial em análises que envolvem variáveis climáticas e biológicas.
- **Monitoramento Contínuo:** A implementação de um sistema automatizado de monitoramento é essencial para garantir a qualidade e a confiabilidade dos dados. Esse sistema verifica continuamente os dados coletados, identificando anomalias em

tempo real, como valores fora do intervalo esperado, inconsistências ou falhas na coleta. Narayanan (2024) argumenta que o monitoramento contínuo em *pipelines* de dados não apenas reduz custos e eleva a qualidade dos dados, mas também facilita a tomada de decisões informadas. Além disso, ao detectar problemas rapidamente, esse processo minimiza erros em análises futuras e melhora a eficiência geral do pipeline.

- **Auditoria e Controle de Versão:** A auditoria e o controle de versão são práticas fundamentais para gerenciar dados em projetos, especialmente os de longo prazo. Essas práticas permitem registrar e acompanhar o histórico das versões dos dados coletados, ajudando a identificar e lidar com mudanças na estrutura de APIs ou bases de dados. Esse acompanhamento contínuo garante que os dados permaneçam consistentes e confiáveis, mesmo diante de alterações nas fontes originais. Conforme destacam Meira *et al.* (2020), a auditoria é indispensável para o monitoramento, validação e gestão dos dados, sendo a precisão e a integridade fatores decisivos para o sucesso de qualquer análise.

2.8.3 Uso de APIs e Integração de Múltiplas Fontes de Dados

Procopio (2023) aponta que a integração com diversas bases de dados melhora a qualidade e confiabilidade das informações, possibilitando a tomada de decisões estratégicas. A combinação de dados de APIs como OpenWeather e NOAA oferece uma visão mais robusta, ao consolidar previsões de curto prazo com históricos climáticos detalhados. Essa integração de múltiplas fontes também pode compensar possíveis lacunas em uma API, fornecendo redundância nos dados e garantindo sua integridade.

Para realizar essa integração de forma eficiente, é fundamental o uso de ferramentas e arquiteturas adequadas. Tecnologias ETL são amplamente empregadas nesse contexto, permitindo a extração de dados brutos, sua transformação para formatos consistentes e a carga em bancos de dados ou sistemas de análise (Kimball e Ross, 2013). Ferramentas como Apache Airflow podem automatizar e orquestrar esses processos, facilitando a gestão e o monitoramento de *pipelines* de dados (Kepner *et al.*, 2017), enquanto sistemas baseados em AWS Lambda oferecem uma solução escalável e de baixo custo para a execução de *pipelines* (Amazon Web Services, 2023).

Em cenários onde há necessidade de lidar com grandes volumes de dados ou picos de requisições provenientes de APIs, o uso de buffers intermediários e arquitetura desacoplada se torna essencial para garantir o desempenho e a confiabilidade da ingestão e processamento dos dados.

2.8.4 Desafios na Coleta e Integração de Dados Climáticos

A coleta e integração de dados climáticos é um processo complexo, principalmente devido à variação natural dos dados e à diversidade de fatores que influenciam o clima em regiões diferentes. A seguir, serão discutidos alguns dos principais desafios enfrentados na coleta e integração de dados climáticos:

- **Variações e Mudanças nos Dados Climáticos:** A oscilação climática apresenta desafios importantes, pois as condições podem mudar de forma rápida e imprevisível. De acordo com Caldana *et al.* (2018), as condições climáticas e meteorológicas são fundamentais para a agricultura e o desenvolvimento social, mas sua natureza imprevisível pode afetar negativamente diversas atividades humanas. Essa instabilidade acaba tornando a coleta de dados e a integração de diversas fontes uma tarefa complexa.
- **Influências Locais e Regionais:** Fatores locais, como a topografia e a proximidade com corpos d'água, também dificultam a coleta e a integração de dados climáticos. Souza *et al.* (2022) afirmam que a mudança na cobertura florestal na área de estudo não pode ser considerada o único fator a influenciar as alterações climáticas, pois a relação entre esses fatores não é apenas térmica. Isso reforça a necessidade de uma abordagem que leve em conta as características regionais para garantir maior precisão.
- **Desafios Relacionados à Incerteza e à Escassez de Dados:** A incerteza sobre os impactos das mudanças climáticas e a escassez de dados confiáveis são obstáculos consideráveis. Ataíde *et al.* (2022) mencionam que há uma grande incerteza sobre os efeitos das mudanças climáticas nos recursos hídricos no Brasil. Essa incerteza representa uma dificuldade ao tentar integrar dados de fontes distintas e ao criar políticas públicas adequadas.
- **Desafios Tecnológicos e Metodológicos:** Para integrar dados climáticos de maneira eficiente, é necessário o uso de tecnologias adequadas e metodologias sólidas. Andrade *et al.* (2018) discutem a aplicação de geo estatísticas para analisar a variabilidade espacial e os padrões nas séries temporais de precipitação, destacando que esses fatores influenciam o clima local e contribuem para mudanças no sistema climático global. Isso sugere que a ausência de ferramentas adequadas pode ser um obstáculo significativo na integração dos dados climáticos.
- **Efeitos das Mudanças Climáticas:** As alterações climáticas afetam diretamente a coleta e a integração de dados. Martins e Ferreira (2010) afirmam que as áreas urbanas são especialmente vulneráveis aos efeitos das mudanças climáticas. Isso

implica que a coleta de dados precisa ser flexível e capaz de adaptar-se às mudanças contínuas nas condições climáticas.

2.9 Estudos de Caso e Aplicações de Sistemas de Alerta Climático

O desenvolvimento de sistemas de alerta para eventos climáticos extremos se tornou uma ferramenta fundamental para minimizar os impactos das mudanças climáticas. Diversos estudos e aplicações reais têm demonstrado como essas tecnologias podem ser eficazes no monitoramento e na comunicação de eventos climáticos extremos, como tempestades, inundações e ondas de calor. Esses sistemas ajudam na preparação e na resposta rápida, permitindo que as autoridades e as comunidades ajam para reduzir riscos e danos.

Um exemplo relevante é o estudo de Xavier *et al.* (2016), que apresenta a implementação de um sistema de alerta climático móvel no Brasil. A pesquisa discute o uso de um simulador de pluviômetro e a realização de testes de um sistema de alerta que envia informações em tempo real a partir de um único número de telefone. Este sistema foi projetado para alcançar áreas vulneráveis e fornecer alertas sobre eventos climáticos, como inundações, que ocorrem com frequência no Brasil. A pesquisa enfatiza a importância da integração da tecnologia móvel nos sistemas de alerta, pois ela facilita a disseminação de informações, principalmente em regiões de difícil acesso.

A pesquisa de Vaidyanathan *et al.* (2019) é outra contribuição importante, pois analisa os sistemas de alerta precoce para ondas de calor. O estudo destaca que a gestão desses riscos vai além das previsões climáticas, sendo necessário implementar soluções locais como ar condicionado e centros de resfriamento. Além disso, a pesquisa aponta que campanhas de conscientização e a comunicação de risco para a população podem reduzir significativamente os impactos dessas ondas de calor, como mortes e complicações de saúde.

Giordano *et al.* (2020) discutem, por sua vez, a relação entre as políticas públicas locais e a adaptação a eventos climáticos extremos. O estudo sugere que desastres naturais podem ser um ponto de inflexão para mudanças políticas e que a colaboração entre diferentes setores da sociedade — governo, organizações não-governamentais e comunidade — é essencial para a efetividade dos sistemas de alerta. A pesquisa reforça a importância de ações locais para melhorar a gestão dos riscos climáticos e adaptar as políticas públicas às necessidades da população.

Kreslake (2019) também contribui para a discussão, abordando como as comunidades dos Estados Unidos estão se adaptando a eventos climáticos extremos. Ele argumenta que a resiliência das comunidades depende da colaboração entre diversos setores, como governos locais, organizações e a própria população. Além de adotar soluções tecnológicas, é necessário criar uma rede de apoio local para garantir que os sistemas de alerta sejam eficazes. O autor destaca que, para que esses sistemas funcionem de forma eficiente, a participação ativa da comunidade é essencial.

Esses estudos de caso mostram que, para o sucesso dos sistemas de alerta para eventos climáticos extremos, é necessário integrar tecnologias de monitoramento com estratégias sociais, políticas e comunitárias. As soluções variam conforme o tipo de evento e as características locais, e é importante adaptar os sistemas de alerta às especificidades de cada região. Dessa forma, esses sistemas não devem ser vistos como uma única solução, mas sim como uma combinação de ferramentas e abordagens que se complementam para garantir sua eficácia.

3 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos adotados na construção da solução proposta. O estudo tem como objetivo contribuir para a compreensão da implementação de um pipeline de dados voltado ao monitoramento e à notificação de eventos climáticos extremos, utilizando tecnologias modernas de engenharia de dados em ambientes de nuvem.

A pesquisa é exploratória e aplicada: exploratória porque investiga e compreende os desafios técnicos e processuais envolvidos na criação do pipeline, e aplicada por buscar resolver um problema prático e específico, que é o monitoramento em tempo real de eventos climáticos extremos para mitigar seus impactos.

Para o tratamento dos dados meteorológicos, que são séries numéricas, a abordagem é quantitativa, envolvendo coleta, processamento e análise estatística para geração de alertas precisos. Assim, há uma convergência metodológica: a pesquisa explora qualitativamente o processo e o contexto da solução, enquanto utiliza métodos quantitativos para analisar e validar os dados tratados, garantindo um equilíbrio entre compreensão técnica e rigor analítico.

A metodologia fundamenta-se nos princípios propostos por Wazlawick (2009), que destacam a importância de uma análise flexível e adaptativa dos dados, permitindo uma melhor compreensão dos fenômenos estudados. Assim, a pesquisa evidencia o papel essencial da engenharia de dados na construção de soluções tecnológicas voltadas à previsão e mitigação de desastres naturais, reforçando sua relevância para a tomada de decisões em tempo real, visando proteger a população e minimizar impactos adversos.

3.1 Tipo de Pesquisa

Este estudo caracteriza-se como uma pesquisa aplicada, por buscar resolver um problema específico dentro do contexto da engenharia de dados. A abordagem é predominantemente quantitativa, com foco na coleta, transformação e análise de dados estruturados. Ao mesmo tempo, assume caráter exploratório e descritivo, na medida em que visa compreender e documentar os processos envolvidos na construção do pipeline de dados. Configura-se também como um estudo de caso, pois analisa a implementação de uma solução em um cenário delimitado e realista.

3.2 Procedimentos Metodológicos

Os procedimentos adotados envolvem a implementação de uma arquitetura de pipeline de dados utilizando ferramentas específicas para os processos de ETL, bem como tecnologias para o tratamento de dados em tempo real (*streaming*). A seguir, são apresentadas as etapas técnicas do desenvolvimento da solução proposta, detalhando cada componente da arquitetura e o fluxo de dados desde a origem até a geração dos alertas.

3.3 Arquitetura do *Pipeline* de Dados

A arquitetura do *pipeline* de dados foi projetada para garantir escalabilidade, eficiência e confiabilidade no processamento e armazenamento dos dados climáticos. O *pipeline* foi dividido em duas abordagens principais: processamento em tempo real e processamento em lote.

Processamento em Tempo Real: A coleta de dados meteorológicos é realizada periodicamente a partir de uma API externa, sendo os dados processados em tempo real por meio de Kinesis Data Streams. Esse serviço, em conjunto com o Lambda, permite uma análise contínua dos eventos climáticos, com a detecção de condições extremas. Quando um evento extremo é identificado, um alerta é gerado e enviado através do SNS para notificação via email.

Processamento em Lote: Para o processamento em lote, os dados são enviados diretamente para o S3, em sua camada Bronze. A partir do S3, os dados são transferidos para o PostgreSQL (camada *Silver*) através do Apache Airflow, onde são limpos e normalizados. Após o processamento, os dados são agregados e armazenados na camada *Gold* do próprio PostgreSQL, também via Airflow, para análise posterior.

3.4 Arquitetura medalhão

O armazenamento dos dados climáticos é estruturado utilizando a **Arquitetura Medalhão** (adaptado de Databricks, 2023), um modelo escalável e eficiente para a organização de dados. Essa arquitetura segue uma abordagem em camadas, facilitando a governança, a qualidade e a recuperação dos dados.

A estrutura é dividida em três camadas principais:

- **Bronze:** Armazena os dados brutos exatamente como foram ingeridos, preservando sua integridade e garantindo a rastreabilidade. Nesta camada, os dados coletados das APIs meteorológicas são armazenados sem modificações, permitindo auditoria e reprocessamento quando necessário.
- **Silver:** Contém dados já limpos, transformados e enriquecidos. Nesta fase, ocorre a remoção de duplicatas, padronização de formatos e junção de diferentes fontes para gerar um conjunto de dados mais confiável para análise.
- **Gold:** Armazena dados prontos para consumo analítico e geração de relatórios. Aqui, os dados são agregados, estruturados e otimizados para consultas, oferecendo insights valiosos para a tomada de decisão.

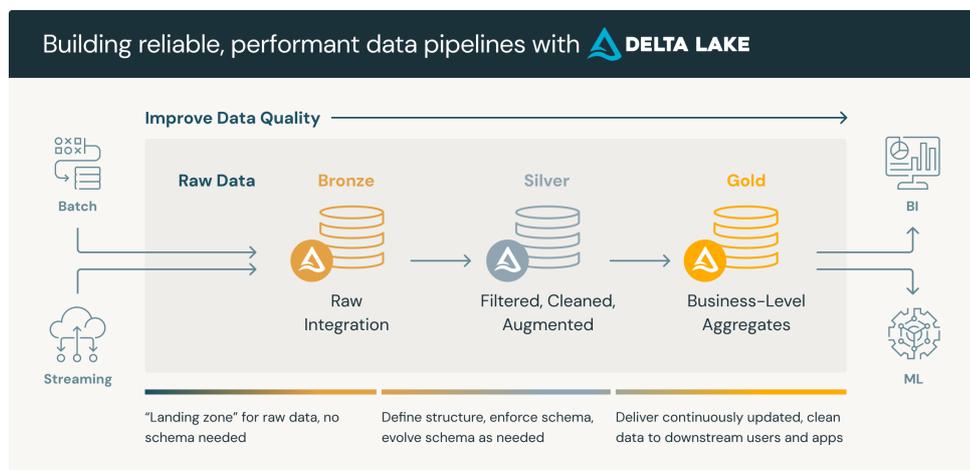
Essa organização melhora a performance das consultas e permite um processamento incremental, essencial para lidar com grandes volumes de dados climáticos. Além disso, a separação em camadas garante maior flexibilidade, possibilitando diferentes níveis de

acesso e qualidade de dados conforme a necessidade dos usuários. Essa abordagem permite que dados de diferentes naturezas e níveis de criticidade sejam tratados de forma otimizada, sem sobrecarregar os sistemas.

A arquitetura medalhão foi adotada para aprimorar tanto a performance quanto a auditoria dos dados, assegurando um controle rigoroso em cada etapa do pipeline.

As tecnologias escolhidas para cada camada foram selecionadas com base em sua escalabilidade, compatibilidade com o ecossistema AWS e capacidade de processar dados em tempo real e em lote, garantindo, assim, um pipeline confiável e eficiente.

Figura 1 – Arquitetura medalhão.



Fonte: Adaptado de Databricks (2023). Disponível em:

<https://www.databricks.com/glossary/medallion-architecture>. Acesso em: 26 mar. 2025.

Descrição: Diagrama que exemplifica a arquitetura medalhão, mostrando as camadas Bronze, Silver e Gold e o fluxo de dados entre elas, utilizado para estruturar pipelines de dados confiáveis e escaláveis.

As tecnologias adotadas em cada etapa do pipeline foram escolhidas com base em sua escalabilidade, compatibilidade com o ecossistema da AWS e capacidade de lidar com dados em tempo real e em lote. Na seção a seguir, detalha-se a justificativa técnica para a escolha de cada ferramenta utilizada.

3.5 Escolha das Tecnologias e Justificativa

A seleção dos componentes tecnológicos do pipeline levou em consideração critérios como escalabilidade, confiabilidade e custo-benefício. Abaixo, são descritas as principais ferramentas utilizadas no sistema:

- **API Meteorológica:** Para a coleta de dados climáticos em tempo real, foi utilizada exclusivamente a API da Tomorrow.io. Essa API se destaca pela sua capacidade de fornecer previsões hiperlocais e de alta resolução, integrando dados de diversas

fontes. Entre essas fontes estão: satélites meteorológicos (como GOES, Meteosat e Himawari), modelos numéricos globais e regionais (como GFS, ECMWF, HRRR, ICON e NAM), estações meteorológicas locais, sensores conectados (como dispositivos IoT e smartphones), dados de radar e modelos próprios baseados em aprendizado de máquina (Tomorrow.io, 2024). A escolha pela Tomorrow.io se deu também pelo fato de que a API da NOAA possui cobertura limitada em cidades brasileiras do interior, como Patos-PB, o que inviabilizou seu uso neste trabalho.

- **AWS Lambda:** Utilizado para coletar os dados das APIs e processá-los logo em seguida.
- **Amazon EventBridge:** Responsável por agendar a execução das funções Lambda a cada 30 minutos, garantindo a coleta periódica dos dados meteorológicos.
- **AWS IAM:** Utilizado para definir os níveis de acesso e permissões das funções Lambda, garantindo que cada função tenha acesso apenas aos recursos necessários, como leitura em APIs, envio para o Kinesis e escrita no S3.
- **Amazon Kinesis Data Streams:** Utilizado para o processamento em tempo real dos dados meteorológicos, permitindo monitoramento contínuo e identificação rápida de eventos climáticos críticos. A escolha pelo Kinesis, em vez do Apache Kafka, se deu pela sua maior simplicidade de configuração, curva de aprendizado reduzida e integração direta com outros serviços da AWS, tornando-o mais adequado para ambientes já baseados na infraestrutura da Amazon.
- **Amazon S3:** Utilizado para o armazenamento dos dados brutos e processados na camada Bronze.
- **PostgreSQL:** Utilizado para armazenar os dados após os processos de limpeza e normalização nas camadas *Silver* e *Gold*. Optou-se pelo PostgreSQL em vez do Snowflake devido ao seu custo reduzido, já que se trata de uma solução gratuita e robusta, adequada às necessidades do projeto sem gerar custos adicionais com infraestrutura ou licenciamento.
- **Apache Airflow:** Orquestra o processamento e movimentação dos dados entre as diferentes camadas do pipeline, garantindo a automação das tarefas de ETL e a integridade dos dados.
- **Amazon SNS:** Usado para enviar alertas por email quando um evento extremo é identificado.

Metabase: Utilizado para criar relatórios de BI.

3.6 Fluxo de Dados

O fluxo de dados no pipeline segue uma arquitetura híbrida, combinando processamento em tempo real e em lote para garantir eficiência e flexibilidade.

3.6.1 Fluxo Experimental

O fluxo experimental adotado no projeto é estruturado em quatro etapas principais:

Figura 2 – Arquitetura do Fluxo Experimental



Fonte: Próprio autor.

A seguir, detalha-se o funcionamento de cada etapa do pipeline implementado:

1. **Coleta de Dados (API e Lambdas):** Os dados climáticos são coletados de uma API externa por meio de Lambdas que executam a coleta a cada 30 minutos.
2. **Processamento em Tempo Real (Kinesis):** Os dados são enviados para o Kinesis Data Streams e, logo em seguida, para um Lambda, onde são analisados em tempo real. Caso um evento extremo seja detectado, um alerta é gerado e enviado por e-mail via SNS.
3. **Processamento em Lote (S3):** Os dados são enviados diretamente para o S3, que atua como armazenamento intermediário. Isso permite a ingestão eficiente dos dados, formando assim a camada *Bronze*.
4. **ETL e Armazenamento no PostgreSQL:** Utilizando o Airflow, os dados armazenados no S3 são carregados no PostgreSQL, onde passam por processos de limpeza e normalização (camada *Silver*). Em seguida, utilizando novamente o Airflow, os dados são transformados e agregados na camada *Gold*, ficando prontos para análises e geração de relatórios.
5. **Análises Climáticas utilizando o Metabase:** Após as transformações realizadas com o Airflow no PostgreSQL, os dados são disponibilizados para análise. O Metabase é utilizado para a criação de relatórios de BI, possibilitando análises mais aprofundadas dos dados climáticos.

3.7 Instrumentos de validação

Para garantir a eficácia e a qualidade do pipeline de dados desenvolvido, foram definidos alguns instrumentos de validação baseados em métricas essenciais que avaliam seu sucesso operacional. As principais métricas consideradas são:

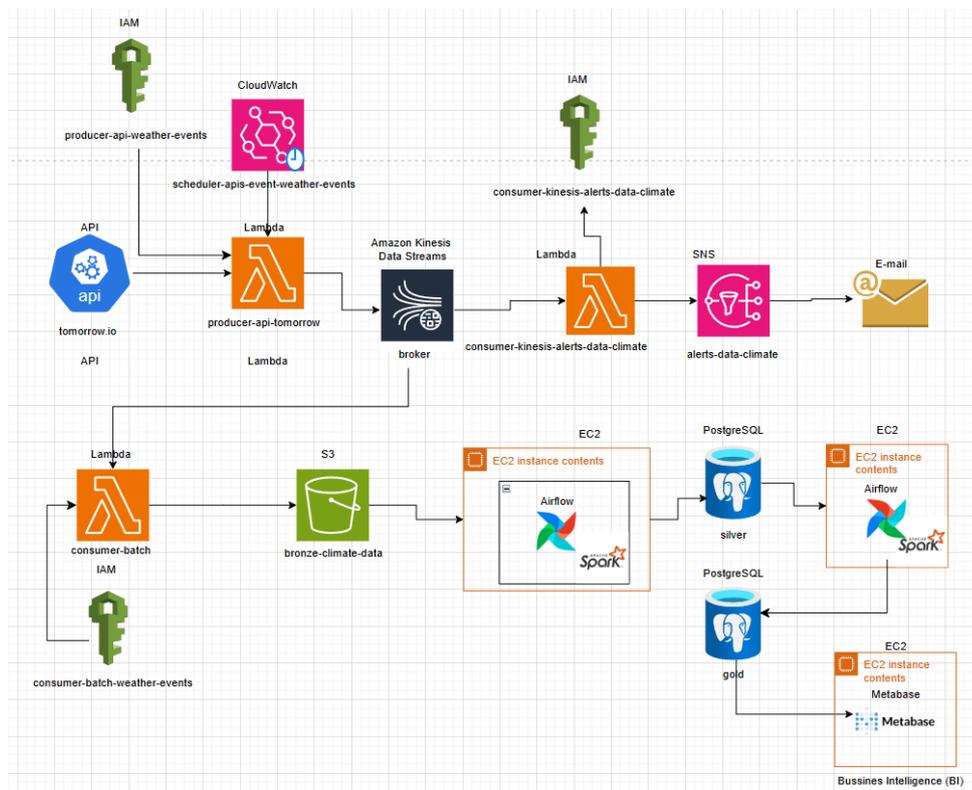
- **Latência:** Mede o tempo total entre a coleta dos dados na API e sua disponibilidade nas camadas finais do pipeline, indicando a rapidez do processamento e a capacidade de entrega em tempo adequado para análises e alertas.
- **Taxa de erro:** Avalia a quantidade de falhas ocorridas durante as etapas de ingestão, processamento e armazenamento dos dados. Essa métrica é fundamental para monitorar a confiabilidade do pipeline e identificar pontos críticos que necessitam de ajustes.
- **Custo operacional:** Refere-se ao gasto total com os recursos computacionais e de armazenamento utilizados, incluindo serviços na nuvem, como AWS Lambda, S3 e demais componentes. O controle do custo é essencial para garantir que o pipeline seja financeiramente viável e escalável.

Essas métricas são monitoradas continuamente por meio de ferramentas de observabilidade integradas ao pipeline, permitindo a rápida identificação de anomalias e a implementação de melhorias contínuas.

4 DESENVOLVIMENTO DA APLICAÇÃO

Este capítulo descreve o desenvolvimento do *pipeline* de dados climáticos, abordando sua arquitetura, a coleta de dados via API meteorológica e a organização dos dados em camadas. O processo inclui a integração com a API Tomorrow IO, o armazenamento estruturado nas camadas *Bronze*, *Silver* e *Gold*, além da aplicação de boas práticas de governança e automação para garantir escalabilidade e confiabilidade no processamento dos dados.

Figura 3 – Arquitetura do *pipeline* de dados.



Fonte: Próprio autor.

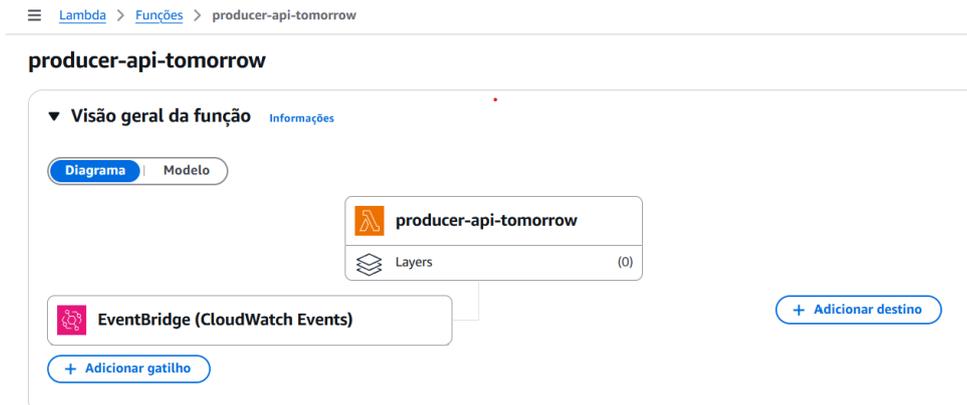
Descrição: Diagrama ilustrativo da arquitetura geral do *pipeline* de dados, demonstrando os principais componentes e o fluxo desde a ingestão até o armazenamento e processamento dos dados.

4.1 Coleta de Dados e Integração com APIs Meteorológicas

O primeiro passo do desenvolvimento do pipeline foi a configuração da integração com a API meteorológica. Utilizando serviços da AWS, como o Lambda, os dados são coletados periodicamente a partir da API externa, sendo a coleta realizada a cada 30 minutos. A API utilizada foi selecionada com base em sua cobertura geográfica, precisão e capacidade de fornecer dados em tempo real que possibilitem a identificação de eventos climáticos extremos, tais como tempestades, inundações, rajadas de vento e ondas de calor.

A API de coleta de dados utilizada foi a Tomorrow IO, que disponibiliza dados em tempo real e históricos, possibilitando assim, previsões detalhadas. Para a aplicação, serão utilizado apenas os dados atuais provenientes dessa API. A integração com essa aplicação foi realizada por meio de uma função AWS Lambda, que executa a coleta de dados conforme o agendamento definido no CloudWatch Events.

Figura 4 – Função AWS Lambda responsável pela requisição da API Tomorrow.



Fonte: Próprio autor. O código-fonte está disponível em: <https://github.com/adrierjs/climate-data-pipeline>. Acesso em: 25 maio 2025.

Descrição: Interface de configuração da função Lambda que orquestra chamadas à API Tomorrow para coletar dados climáticos em tempo real.

Figura 5 – Trecho de código da função Lambda.

```

STREAM_NAME = "broker"
REGION = "us-east-1"

kinesis_client = boto3.client('kinesis', region_name=REGION)

def lambda_handler(event, context):
    response = requests.get(url, headers=headers)
    weather_data = response.json()

    kinesis_client.put_record(
        StreamName=STREAM_NAME,
        Data=json.dumps(weather_data),
        PartitionKey="partition_key"
    )

    return {
        'statusCode': 200,
        'body': json.dumps('Dados enviados ao Kinesis com sucesso')
    }

```

Fonte: Próprio autor.

Descrição: Fragmento do código-fonte da função Lambda, demonstrando a implementação da lógica de requisição e tratamento dos dados recebidos da API.

4.2 Armazenamento e Governança dos Dados

A organização e governança dos dados seguem a arquitetura medalhão, conforme discutido na seção de metodologia, com a utilização das camadas Bronze, *Silver* e *Gold* para

Figura 6 – Exemplo de resposta da API.

```

1 |
2 |   "data": {
3 |     "time": "2025-05-04T03:24:00Z",
4 |     "values": {
5 |       "cloudBase": 2.2,
6 |       "cloudCeiling": 2.2,
7 |       "cloudCover": 81,
8 |       "dewpoint": 16.1,
9 |       "freezingRainIntensity": 0,
10 |      "humidity": 84,
11 |      "precipitationProbability": 9,
12 |      "pressureSeaLevel": 1012.83,
13 |      "pressureSeaLevel": 1012.49,
14 |      "rainIntensity": 0,
15 |      "sleetIntensity": 0,
16 |      "snowIntensity": 0,
17 |      "temperature": 37.1,
18 |      "temperatureApparent": 37.1,
19 |      "uvIndex": 0,
20 |      "visibility": 9.44,
21 |      "weatherCode": 1902,
22 |      "windDirection": 204,
23 |      "windGust": 3.5,
24 |      "windSpeed": 0.6
25 |     }
26 |   },
27 |   "location": {
28 |     "lat": 15.3333

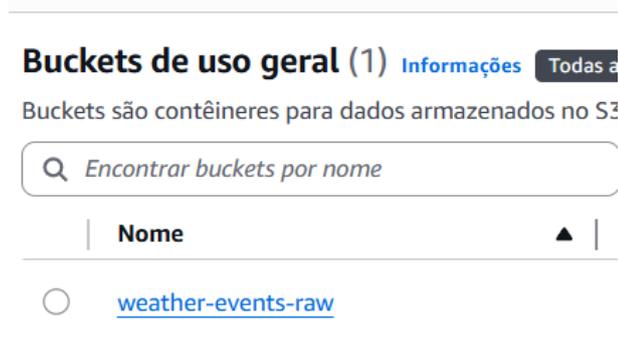
```

Fonte: Próprio autor.

Descrição: Exemplo do formato JSON da resposta retornada pela API Tomorrow, contendo dados climáticos utilizados no pipeline.

garantir a integridade e rastreabilidade dos dados ao longo de todo o pipeline. O armazenamento nas diferentes camadas permite que os dados sejam facilmente acessados, processados e analisados conforme suas características e maturidade.

Figura 7 – Camada Bronze.



Fonte: Próprio autor.

Descrição: Representação da camada Bronze, responsável por armazenar os dados brutos recebidos diretamente do *broker*, sem qualquer tipo de tratamento. Essa camada atua como repositório inicial, garantindo a persistência e rastreabilidade dos dados originais.

Figura 8 – Camada *Silver*.



Fonte: Próprio autor.

A camada Silver apresenta os dados já tratados e limpos, prontos para análises mais específicas. Nessa etapa, realiza-se o enriquecimento dos dados e a padronização de formatos, promovendo uma estrutura mais organizada e confiável.

Figura 9 – Camada *Gold*.

Fonte: Próprio autor.

Ilustração da camada Gold, que contém os dados prontos para consumo por dashboards, relatórios e análises de negócio. Os dados aqui já estão agregados, contextualizados e estruturados de acordo com os requisitos das áreas usuárias.

Cada camada da arquitetura medalhão desempenha um papel crucial:

- **Bronze:** Armazena dados brutos, coletados diretamente da API Tomorrow.IO.
- **Silver:** Armazena os dados refinados e limpos, com tratamento e enriquecimento, recebidos da camada Bronze.
- **Gold:** Armazena os dados da camada *Silver*.

5 ANÁLISE E DISCUSSÃO

Este capítulo apresenta a análise e discussão dos processos envolvidos na construção do pipeline de dados para monitoramento climático, destacando tanto as etapas de processamento em tempo real quanto em lote. Serão detalhadas as soluções adotadas para ingestão, processamento, armazenamento e monitoramento dos dados, utilizando serviços como Amazon Kinesis, AWS Lambda, Amazon SNS, Amazon S3, PostgreSQL e Apache Airflow. Além disso, serão discutidos os critérios utilizados para a detecção de eventos climáticos extremos, bem como os benefícios, desafios e resultados obtidos com a implementação da arquitetura proposta.

5.1 Processamento em Tempo Real com Kinesis Data Streams

Após a coleta dos dados, a etapa de processamento em tempo real foi configurada utilizando o Kinesis Data Streams. Esse serviço permite que os dados meteorológicos sejam processados continuamente à medida que são coletados, sem a necessidade de esperar o término de um ciclo de coleta completo. Por meio da utilização do AWS Lambda, é realizada a análise em tempo real dos dados, detectando eventos climáticos extremos e permitindo a geração imediata de alertas.

Sempre que um evento extremo é identificado, o Lambda aciona a criação de um alerta, que é enviado via Amazon SNS para notificação por e-mail. Isso permite que os usuários sejam informados sobre condições adversas com rapidez e precisão, garantindo uma resposta ágil para mitigar impactos.

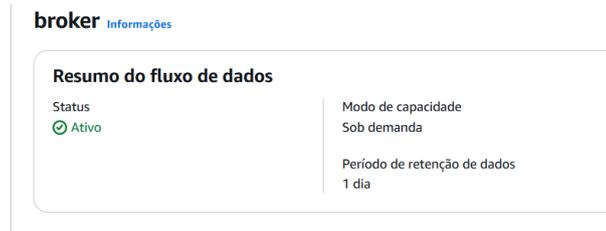
5.1.1 Aquisição e Transmissão Contínua de Dados com o Kinesis Data Streams

O Kinesis Data Streams, atuando como *Broker* foi utilizado para ingerir dados meteorológicos de forma contínua, possibilitando o envio imediato das informações coletadas para o processamento subsequente. Essa abordagem em *streaming* permitiu o tratamento quase instantâneo dos dados, reduzindo a latência e aumentando a capacidade de resposta do sistema em tempo real.

5.1.2 Processamento em Tempo Real com AWS Lambda

A função AWS Lambda foi configurada para ser acionada automaticamente a cada novo dado recebido pelo Kinesis. Essa função executa a lógica de detecção de eventos climáticos extremos, como altas temperaturas, chuvas intensas e ventos fortes. Ao identificar essas condições, a função Lambda prepara os dados relevantes para o disparo de alertas, garantindo uma análise ágil e automatizada.

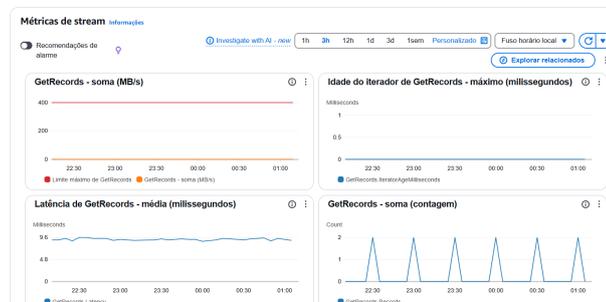
Figura 1 – *Broker* responsável por receber os dados climáticos.



Fonte: Próprio autor.

Descrição: A imagem ilustra o uso do Amazon Kinesis como *broker* para ingestão dos dados climáticos em tempo quase real. O serviço atua como ponto central de recepção antes do armazenamento definitivo no S3.

Figura 2 – Métricas do *Broker*.



Fonte: Próprio autor.

Descrição: As métricas apresentadas refletem o volume de dados trafegado no *broker*, incluindo número de registros processados, taxa de entrada e eventuais falhas, permitindo o monitoramento da performance e estabilidade do pipeline.

Figura 3 – Função Lambda responsável pela análise de eventos climáticos extremos.



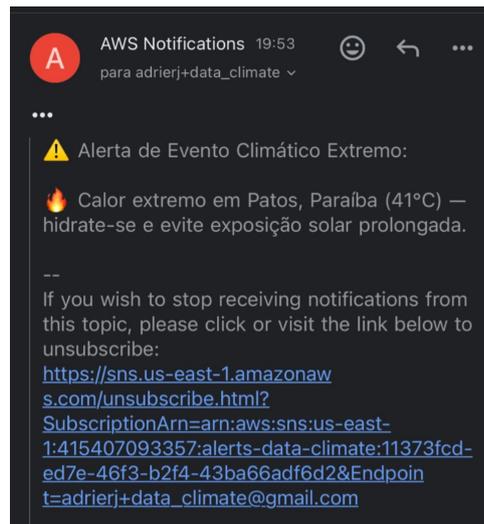
Fonte: Próprio autor.

Descrição: Representação da função AWS Lambda que processa eventos climáticos extremos recebidos via Kinesis e SNS, realizando análises em tempo real para alertas e monitoramento.

5.1.3 Geração e Envio de Alertas com Amazon SNS

Após a detecção de um evento extremo, a função Lambda envia uma notificação ao serviço SNS. O SNS é responsável por distribuir os alertas para os canais configurados, como *e-mail* ou SMS. Essa integração permite que as notificações cheguem aos usuários em poucos segundos, facilitando ações preventivas e respostas rápidas.

Figura 4 – Notificação de alerta enviada por *e-mail* via Amazon SNS.



Fonte: Próprio autor.

Descrição: Exemplo de notificação de alerta enviada por e-mail usando o serviço Amazon SNS, utilizada para comunicação rápida de eventos críticos no pipeline de dados climáticos.

5.1.4 Benefícios da Arquitetura em Tempo Real

A arquitetura baseada em Kinesis, Lambda e SNS se mostrou altamente vantajosa. Entre os principais benefícios observados estão: a escalabilidade automática, a eliminação de processos em batch, a redução de custos operacionais e a maior precisão no envio de alertas. Essa solução em tempo real permitiu uma rápida resiliência e escalabilidade na detecção de eventos climáticos extremos.

5.2 Processamento em Lote com S3, PostgreSQL e Airflow

A parte do pipeline que lida com dados em lote foi projetada para garantir eficiência no armazenamento e no processamento dos dados históricos. Inicialmente, os dados coletados são enviados para o Amazon S3, compondo a camada Bronze — onde ficam armazenados como dados brutos, prontos para serem limpos e enriquecidos.

Na camada Silver, os dados brutos são limpos, normalizados e enriquecidos com a adição de duas colunas relacionadas à data atual, facilitando a auditoria no futuro. Todo esse processo é realizado utilizando o Apache Spark com PySpark. Em seguida, os dados

são carregados no banco de dados PostgreSQL, preparando-os para análises posteriores. Após esse processamento, os dados são movidos para a camada Gold, onde são agregados, estruturados e otimizados para consultas analíticas e geração de relatórios gerenciais.

O Apache Airflow foi fundamental para a orquestração do pipeline, garantindo que os dados fossem processados de forma automática e eficiente. Ele foi configurado para gerenciar as tarefas de ETL, movendo os dados entre as camadas *Silver* e *Gold* de acordo com a lógica definida para cada etapa. Além disso, também foi configurado para monitorar e registrar o status de cada tarefa, permitindo a detecção e correção de falhas rapidamente. Isso assegura uma execução confiável e sem interrupções.

Após as transformações realizadas nas referidas camadas, os dados ficam disponíveis para análise. O Metabase é utilizado para a criação de relatórios de BI, possibilitando análises aprofundadas e a geração de insights com base nos dados climáticos.

5.2.1 Parâmetros para a detecção de eventos climáticos extremos

Tabela 5.1 – Critérios para Identificação de Eventos Climáticos Extremos

Parâmetro	Valor de Referência	Descrição	Fonte
Precipitação (diária)	≥ 100 mm	Associada a inundações e desastres ambientais em regiões vulneráveis.	Guimarães <i>et al.</i> , 2021
Velocidade do vento	> 15 km/h (impactante); ~ 100 km/h (ciclone)	Pode causar danos severos em eventos como tempestades e ciclones.	Guimarães <i>et al.</i> , 2021
Rajadas de vento	> 20 km/h	Rajadas intensas agravam danos em estruturas.	Mohr <i>et al.</i> , 2017.
Intensidade da chuva	> 10 mm/h (forte); > 20 mm/h (extrema)	Chuvas fortes podem causar alagamentos.	Dwyer & O’Gorman, 2017
Intensidade da neve	> 5 mm/h (equiv. água)	Nevascas intensas impactam mobilidade e infraestrutura.	Critério técnico (WMO/NOAA)
Temperatura alta	$> 40^{\circ}\text{C}$	Associada a ondas de calor e riscos à saúde.	Guimarães <i>et al.</i> , 2021
Temperatura baixa	$< -5^{\circ}\text{C}$	Representa riscos de frio intenso.	Guimarães <i>et al.</i> , 2021
Índice UV	> 7	Nível alto de radiação solar prejudicial à saúde.	Santos <i>et al.</i> , 2016
Baixa visibilidade	< 1 km	Reduz segurança em transportes em nevoeiros ou tempestades de areia.	Critério técnico (WMO/NOAA)
Pressão atmosférica	< 1000 hPa	Indica presença de sistemas de baixa pressão.	Guimarães <i>et al.</i> , 2021

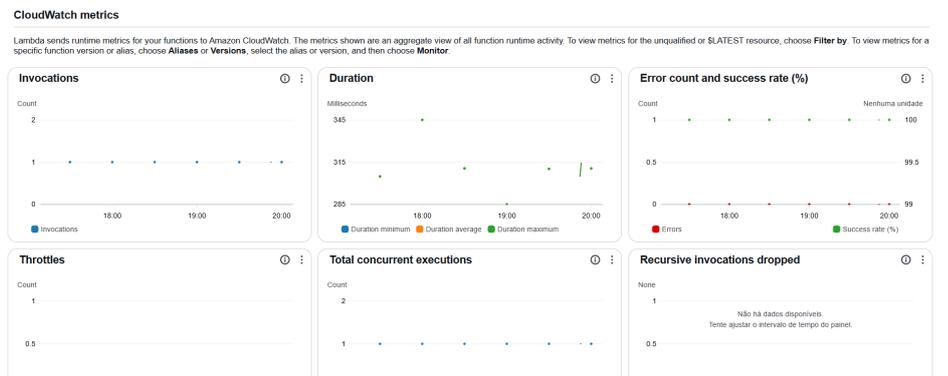
5.3 Monitoramento e Manutenção do Pipeline

O monitoramento contínuo e a manutenção eficaz do pipeline de dados são essenciais para garantir que o processo funcione sem interrupções e que qualquer falha seja detectada e corrigida rapidamente. Para isso, foram adotadas duas ferramentas cruciais: Amazon CloudWatch e Apache Airflow.

O Amazon CloudWatch foi utilizado para monitorar os serviços da AWS, como Amazon S3, AWS Lambda, Amazon Kinesis, entre outros. Ele oferece métricas em tempo real e a capacidade de criar alarmes para detectar problemas, garantindo que os serviços da AWS estejam operando de acordo com o esperado.

Por outro lado, o Apache Airflow foi empregado para monitorar a execução da limpeza e migração dos dados entre as tabelas Bronze e *Gold*. O Airflow orquestra o fluxo

Figura 5 – Painel do AWS CloudWatch utilizado para o monitoramento das métricas da aplicação.

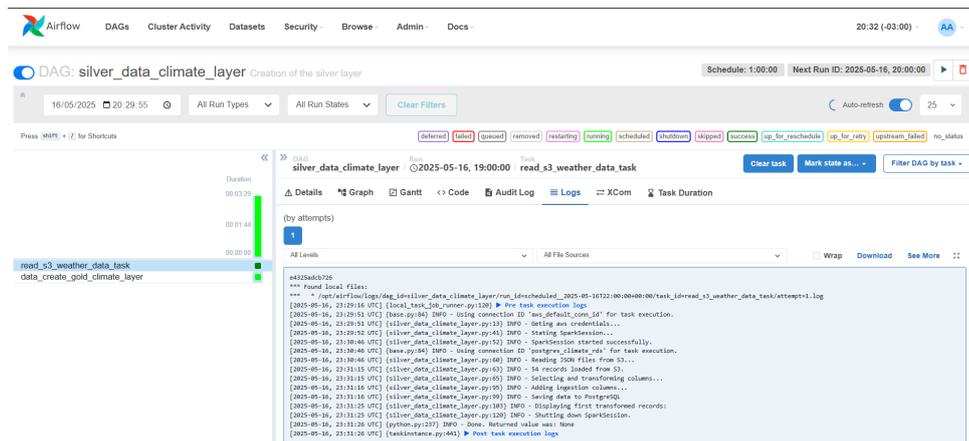


Fonte: Próprio autor.

Descrição: O painel do AWS CloudWatch exibe em tempo real as principais métricas de desempenho e logs da aplicação, permitindo a visualização de tendências, identificação de anomalias e acompanhamento da saúde geral do sistema de processamento de dados.

de trabalho, garantindo que cada tarefa de migração e limpeza seja executada conforme o planejado, e fornece recursos como logs detalhados e notificações em caso de falhas. Esses recursos permitem a intervenção imediata caso haja qualquer problema na execução das tarefas. A combinação dessas ferramentas proporciona um sistema robusto de

Figura 6 – Painel do Apache Airflow utilizado para o monitoramento da aplicação.



Fonte: Próprio autor.

Descrição: O painel do Apache Airflow ilustra o monitoramento do fluxo de trabalho, exibindo o status de execução de cada tarefa, logs detalhados, e permitindo a identificação e resolução de falhas no pipeline de dados, garantindo a orquestração e execução eficaz das operações.

monitoramento e manutenção, assegurando que o pipeline de dados seja mantido eficiente e funcionando corretamente ao longo do tempo.

5.4 Indicadores de Custos da Arquitetura na AWS

A análise dos custos envolvidos na implementação e operação do pipeline de dados é fundamental para garantir a viabilidade econômica da solução, especialmente quando se trabalha com serviços na nuvem que possuem cobrança baseada no consumo. Neste projeto, foram utilizados diversos serviços da AWS, tais como Amazon Kinesis Data Streams, AWS Lambda, Amazon SNS, Amazon S3, PostgreSQL (via Amazon RDS) e Apache Airflow, cada um com seu modelo específico de cobrança.

5.4.1 Modelos de Cobrança dos Serviços Utilizados

O processamento em tempo real com Kinesis e Lambda apresenta custos proporcionais ao volume de dados ingeridos e ao número de invocações das funções Lambda. Já o processamento em lote utilizando S3 para armazenamento e PostgreSQL para consulta possui custos mais estáveis e previsíveis, porém pode demandar maior espaço de armazenamento e tempo computacional durante as execuções das tarefas orquestradas pelo Airflow.

5.4.2 Estratégias para Otimização de Custos

Para controlar os custos, adotaram-se algumas estratégias, tais como:

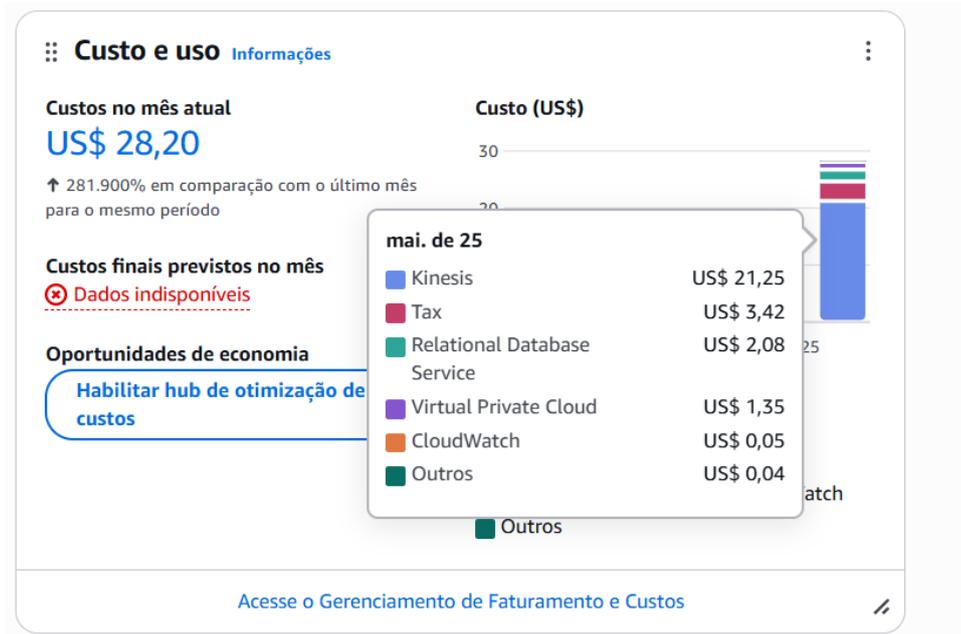
- Uso de funções Lambda com duração otimizada, evitando execuções prolongadas que aumentam o custo.
- Armazenamento de dados em camadas no S3, permitindo arquivamento de dados menos acessados em classes de menor custo.
- Monitoramento constante das métricas de uso via Amazon CloudWatch para antecipar possíveis picos e ajustar a capacidade dos serviços.
- Orquestração das tarefas de processamento em lote via Apache Airflow, evitando execuções desnecessárias e otimizando janelas de execução para minimizar custos.

5.4.3 Visualização dos Indicadores de Custos

A Figura ?? apresenta um *printscreen* do painel de custos da AWS referente ao período de implementação do pipeline, demonstrando a distribuição dos gastos entre os serviços utilizados.

Essa análise reforça a importância do planejamento financeiro em projetos de engenharia de dados na nuvem, destacando como a escolha e o uso consciente dos serviços impactam diretamente no custo total da solução, permitindo a escalabilidade sem perda de eficiência econômica.

Figura 7 – Visão geral dos custos dos serviços AWS durante a execução do pipeline.



Fonte: Próprio autor.

Descrição: O gráfico apresenta a distribuição dos custos gerados pelos diferentes serviços da AWS empregados na execução do pipeline de dados, oferecendo uma análise detalhada do consumo de recursos e suas respectivas despesas ao longo do período de monitoramento.

5.5 Análises de Dados com Metabase

Com os dados processados e organizados na camada *Gold* do *pipeline*, foi possível realizar análises exploratórias e gerar insights relevantes utilizando consultas SQL dentro da plataforma Metabase. A seguir, são apresentados alguns exemplos de análises desenvolvidas com base nas informações climáticas armazenadas:

5.5.1 Média Diária de Temperatura

Esta análise apresenta a variação da temperatura ao longo dos dias, permitindo a identificação de tendências de aquecimento, resfriamento e possíveis anomalias climáticas. O fluxo de dados foi executado durante um período de dois dias, utilizando a cidade de Patos-PB como referência para esta observação.

5.5.2 Vento Médio Diário

A análise da velocidade média diária do vento foi realizada durante dois dias na cidade de Patos-PB. Esse estudo é fundamental para identificar padrões regionais de vento, auxiliando na detecção precoce de eventos extremos, como tempestades e rajadas intensas, e contribuindo para ações preventivas em áreas como agricultura, energia e defesa civil.

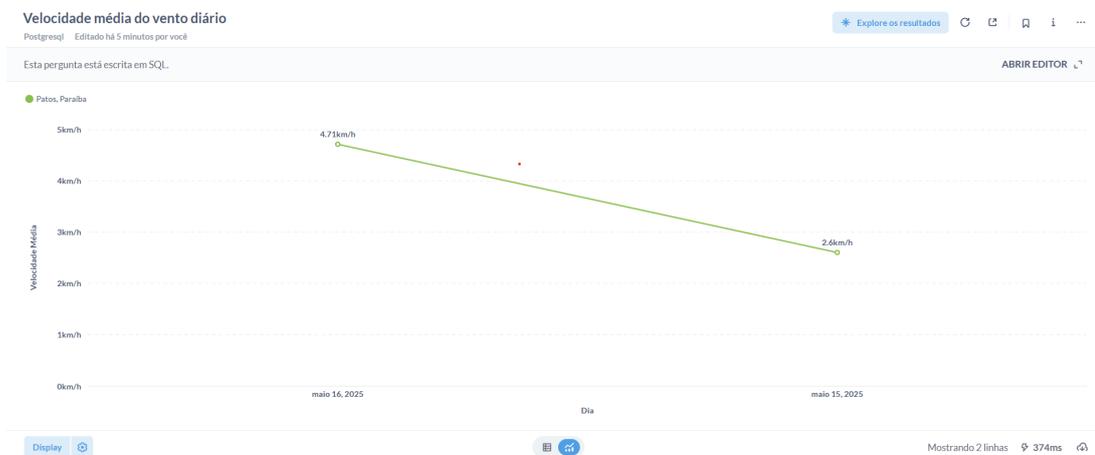
Figura 8 – *Dashboard* no Metabase mostrando a média diária de temperatura.



Fonte: Próprio autor.

Descrição: O dashboard no Metabase exibe graficamente a média diária de temperatura, permitindo a visualização de tendências e variações ao longo do tempo. Este painel é crucial para a análise e monitoramento dos dados climáticos coletados.

Figura 9 – Visualização no Metabase com o vento médio registrado por dia.



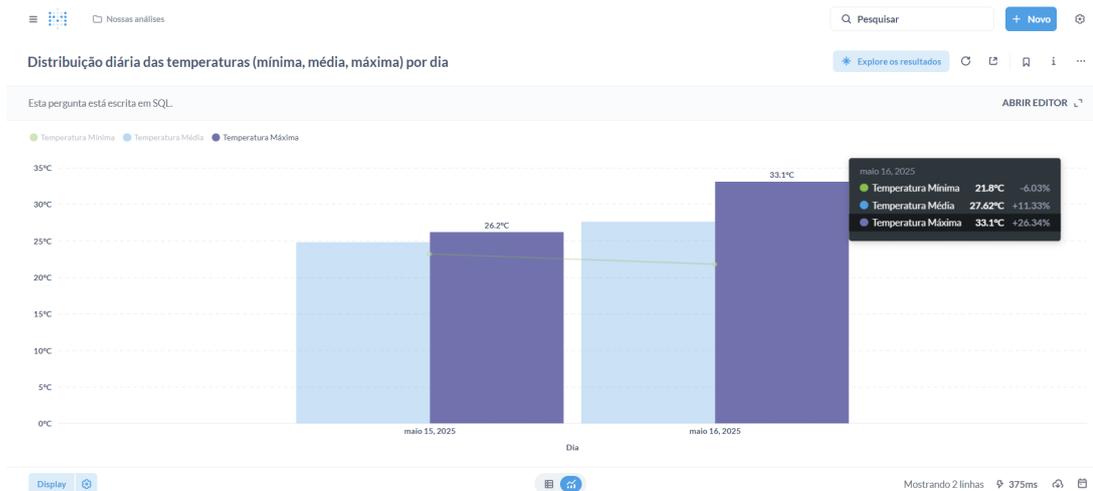
Fonte: Próprio autor.

Descrição: O gráfico no Metabase apresenta a média diária da velocidade do vento, ilustrando as variações ao longo do tempo. Essa visualização é fundamental para a análise de padrões climáticos e o monitoramento das condições ambientais.

5.5.3 Distribuição de Temperaturas por Período

Esta visualização mostra como as temperaturas estão distribuídas ao longo do tempo, permitindo compreender a frequência de ocorrências de temperaturas extremas e sua concentração em determinados horários ou dias. A análise foi realizada durante dois dias na cidade de Patos-PB, proporcionando uma visão localizada das variações térmicas ao longo do período.

Figura 10 – Gráfico de distribuição de temperaturas registrado ao longo dos períodos.



Fonte: Próprio autor.

Descrição: O gráfico de distribuição de temperaturas ilustra a frequência e a variação dos dados climáticos ao longo dos períodos analisados, permitindo identificar padrões, picos e as faixas de temperatura mais comuns registradas pelo pipeline.

5.5.4 Análise de Umidade Média por Dia

A umidade média é uma variável essencial para avaliar o conforto térmico e a possibilidade de formação de chuvas. Esta análise acompanha a evolução da umidade ao longo do tempo e foi realizada durante dois dias na cidade de Patos-PB, permitindo identificar padrões e variações que impactam diretamente nas condições climáticas locais.

Essas análises permitiram transformar dados brutos em informações valiosas para a tomada de decisão, além de facilitar o monitoramento de condições meteorológicas com base em evidências históricas. O uso do Metabase também viabilizou a criação de dashboards interativos, que podem ser compartilhados com diferentes públicos de forma simples e visual.

5.6 Considerações Finais sobre o Desenvolvimento

O desenvolvimento deste *pipeline* de dados seguiu uma abordagem iterativa e adaptativa, garantindo que cada etapa fosse otimizada de acordo com as necessidades do projeto. A integração de dados em tempo real com o processamento em lote, aliado ao uso das ferramentas da AWS e do Apache Airflow, permitiu construir uma solução robusta e escalável, capaz de fornecer alertas precisos e oportunos sobre eventos climáticos extremos.

A arquitetura medalhão, por sua vez, garantiu que os dados fossem organizados de forma eficiente e escalável, possibilitando análises de grandes volumes de dados sem comprometer a performance. Esse desenvolvimento oferece um modelo flexível e aplicável a outros contextos de análise de dados em tempo real, especialmente em cenários que

Figura 11 – Consulta SQL no Metabase com a umidade média registrada diariamente.



Fonte: Próprio autor.

Descrição: A imagem apresenta uma consulta SQL no Metabase, calculando a umidade média diária. Este tipo de consulta é fundamental para a análise de dados climáticos, permitindo extrair informações específicas para visualização e monitoramento da umidade ao longo do tempo.

exigem resposta rápida e assertiva.

5.7 Limitações do Projeto

Durante o desenvolvimento do *pipeline* de dados, foi necessário realizar algumas adaptações para garantir maior simplicidade, compatibilidade e viabilidade técnica. Inicialmente, os serviços estavam planejados para serem hospedados em uma instância EC2 na AWS, o que permitiria um ambiente controlado e escalável para a execução do pipeline. No entanto, devido aos custos associados à manutenção dessa infraestrutura na nuvem, optou-se pelo desenvolvimento local, o que facilitou a implementação e os testes iniciais do pipeline.

Uma das principais adaptações foi a utilização do Apache Spark para viabilizar a ingestão dos dados provenientes do Amazon S3. O PostgreSQL não oferece suporte nativo à leitura direta de arquivos armazenados no S3, exigindo uma solução intermediária. O Spark foi utilizado para preencher essa lacuna, atuando na extração, transformação e carga dos dados brutos armazenados no S3. Após o processamento, os dados foram inseridos no PostgreSQL, onde foram organizados nas camadas *Silver* e *Gold*.

Ao manter o S3 como repositório principal da camada bronze, a arquitetura pôde aproveitar as vantagens de um *data lake* — como escalabilidade e baixo custo de armazenamento — enquanto o PostgreSQL serviu como ambiente para armazenamento e análise dos dados tratados. Essa combinação resultou em um pipeline de dados eficiente, rastreável e de fácil manutenção.

Vale destacar que a execução do *pipeline* foi limitada a dois dias por questões de custo.

Durante esse período, os serviços utilizados na AWS geraram uma cobrança aproximada de 28 dólares. Essa limitação foi necessária para viabilizar o projeto dentro de um orçamento reduzido, sem comprometer a qualidade e os objetivos da análise.

Por fim, embora o *pipeline* atenda plenamente aos objetivos propostos, recomenda-se como trabalhos futuros a reavaliação da arquitetura com foco em escalabilidade, bem como a implementação de testes automatizados e pipelines de CI/CD para fortalecer a confiabilidade e a robustez da solução.

6 CONCLUSÕES

Este trabalho teve como foco o desenvolvimento e a implementação de um *pipeline* de dados voltado ao monitoramento e à notificação de eventos climáticos extremos. A proposta surgiu diante da crescente ocorrência desses fenômenos e da necessidade de soluções tecnológicas capazes de identificar e reportar situações de risco com agilidade e precisão.

Ao longo do projeto, foram utilizadas tecnologias amplamente empregadas na área de Engenharia de Dados, como Amazon Kinesis para ingestão e processamento em tempo real, Amazon S3 e PostgreSQL para armazenamento, Apache Airflow para automação e orquestração das etapas do pipeline, e AWS Lambda para execução de funções sob demanda, de forma escalável e sem a necessidade de provisionamento de servidores. A integração com a API meteorológica permitiu a coleta contínua de dados, que foram tratados e analisados conforme parâmetros predefinidos, resultando na geração automática de alertas.

A arquitetura adotada seguiu o modelo em camadas do tipo medalhão, o que possibilitou uma organização mais eficiente dos dados e maior controle sobre as transformações aplicadas. No entanto, é importante destacar algumas limitações metodológicas do sistema desenvolvido. A principal delas é a dependência de conectividade com a internet, tanto para a coleta de dados via API quanto para o funcionamento dos serviços em nuvem. Além disso, há uma dependência dos termos e condições dos provedores de dados meteorológicos, como o Tomorrow.io, que pode alterar seu modelo de preços, limites de requisições ou disponibilidade de serviços, impactando diretamente na operação e nos custos do pipeline.

6.1 Trabalhos Futuros

Como trabalhos futuros, propõe-se a avaliação de uma arquitetura alternativa baseada em tecnologias open-source, como Apache Kafka para *streaming* de dados e Grafana para visualização e monitoramento em tempo real. Além disso, recomenda-se explorar abordagens de *edge computing*, visando reduzir a dependência de conectividade e permitir o processamento de dados diretamente na borda, próximo às fontes de coleta. Também se sugere a inclusão de modelos preditivos baseados em aprendizado de máquina, bem como a ampliação do sistema para abranger outros tipos de monitoramento ambiental, como enchentes, queimadas e qualidade do ar. A integração de múltiplas APIs meteorológicas poderia ainda aumentar a robustez, a confiabilidade e a precisão das análises. Espera-se que a solução desenvolvida contribua para ações preventivas e auxilie na tomada de decisão em situações de emergência climática.

REFERÊNCIAS

AGHBARI, Z. A.; ISMAIL, T.; KAMEL, I. **Sparkknn: a distributed in-memory data partitioning for knn queries on big spatial data**. *Data Science Journal*, v. 19, 2020.

Disponível em: <https://doi.org/10.5334/dsj-2020-035>.

ALHILAL, A.; BRAUD, T.; HUI, P. **Distributed vehicular computing at the dawn of 5G: a survey**. 2020. Disponível em: <https://doi.org/10.48550/arxiv.2001.07077>.

AMAZON WEB SERVICES. **AWS Lambda**. Disponível em: https://docs.aws.amazon.com/pt_br/lambda/latest/dg/welcome.html. Acesso em: 22 nov. 2024.

AMAZON WEB SERVICES. **AWS Lambda: introdução**. Disponível em: https://docs.aws.amazon.com/pt_br/lambda/latest/dg/welcome.html. Acesso em: 22 nov. 2024.

AMAZON WEB SERVICES. **AWS Lambda: Serverless Compute**. Disponível em: <https://aws.amazon.com/lambda/>. Acesso em: 27 set. 2024.

AMAZON WEB SERVICES. **Amazon Kinesis Features**. Disponível em: <https://aws.amazon.com/kinesis/features/>. Acesso em: 28 abr. 2025.

AMAZON WEB SERVICES. **Amazon Kinesis Data Streams Developer Guide**. Disponível em: <https://docs.aws.amazon.com/streams/latest/dev/what-is-kinesis-data-streams.html>. Acesso em: 28 abr. 2025.

AMAZON WEB SERVICES. **Novas informações no console do IAM ajudam a seguir as melhores práticas do IAM**. Disponível em: <https://aws.amazon.com/blogs/security/newly-updated-features-in-the-aws-iam-console-help-you-adhere-to-iam-best-practices/>. Acesso em: 9 maio 2025.

AMAZON WEB SERVICES. **O que é um pipeline de dados?**. 2024. Disponível em: <https://aws.amazon.com/what-is/data-pipeline/>. Acesso em: 28 set. 2024.

AMAZON WEB SERVICES. **O que é processamento em lote?**. Disponível em: <https://aws.amazon.com/pt/what-is/batch-processing/>. Acesso em: 28 set. 2024.

AMAZON WEB SERVICES. **O que são dados de streaming?**. Disponível em: <https://aws.amazon.com/pt/what-is/streaming-data/>. Acesso em: 28 set. 2024.

AMAZON WEB SERVICES. **Práticas recomendadas de segurança no IAM.** Disponível em: https://docs.aws.amazon.com/pt_br/IAM/latest/UserGuide/best-practices.html. Acesso em: 9 maio 2025.

AMAZON WEB SERVICES. **Recomendações de controle de segurança para gerenciamento de identidade e acesso.** Disponível em: <https://docs.aws.amazon.com/prescriptive-guidance/latest/security-controls-by-caf-capability/identity-and-access-controls.html>. Acesso em: 9 maio 2025.

AMAZON WEB SERVICES. **What is Amazon Kinesis?.** Disponível em: <https://docs.aws.amazon.com/streams/latest/dev/introduction.html>. Acesso em: 28 abr. 2025.

ANDRADE, A. R. S. d. et al. **Geoestatística aplicada à variabilidade espacial e padrões nas séries temporais da precipitação no agreste pernambucano.** *Journal of Environmental Analysis and Progress*, p. 126-145, 2018. Disponível em: <https://doi.org/10.24221/jeap.3.1.2018.1668.126-145>.

APACHE SOFTWARE FOUNDATION. **Apache Spark Documentation.** 2025. Disponível em: <https://spark.apache.org/docs/latest/index.html>. Acesso em: 16 maio 2025.

ASHKOUTI, F.; KHAMFOROOSH, K. **A distributed computing model for big data anonymization in the networks.** *Plos One*, v. 18, n. 4, e0285212, 2023. Disponível em: <https://doi.org/10.1371/journal.pone.0285212>.

ATAIDE, L. C. P. d. et al. **Aplicação da análise fatorial para determinação da vulnerabilidade hidrometeorológica para o sub médio São Francisco.** *Revista Brasileira De Meteorologia*, v. 37, n. 4, p. 405-417, 2022. Disponível em: <https://doi.org/10.1590/0102-77863740021>.

AWS - AMAZON WEB SERVICES. **Getting started with Amazon Simple Storage Service (Amazon S3).** Disponível em: <https://aws.amazon.com/pt/s3/getting-started/#:~:text=0%20Amazon%20Simple%20Storage%20Service,de%20qualquer%20lugar%20na%20Web>. Acesso em: 30 set. 2024.

CALDANA, N. F. d. S. et al. **Ocorrências de alagamentos, enxurradas e inundações e a variabilidade pluviométrica na bacia hidrográfica do rio Iguaçu.** *Revista Brasileira De Climatologia*, v. 23, 2018. Disponível em: <https://doi.org/10.5380/abclima.v23i0.60584>.

CARLO, D.; ANDRADE, D.; LIBERATO, R.; SCHWERZ, A. L. **Um estudo exploratório das ferramentas de código aberto para a replicação de dados no Post-**

greSQL. In: ESCOLA REGIONAL DE BANCO DE DADOS (ERBD 2019), 15., Florianópolis, 2019. Anais [...]. Disponível em: <https://doi.org/10.5753/erbd.2019.8474>.

CAVALCANTE, A. B.; MEDEIROS, V. W. C. D.; GONÇALVES, G. E. **Detecção de anomalias em dados meteorológicos do sertão de Pernambuco utilizando isolation forest e dbscan**. In: WORKSHOP DE COMPUTAÇÃO APLICADA À GESTÃO DO MEIO AMBIENTE E RECURSOS NATURAIS (WCAMA 2023), 14., João Pessoa, 2023. Anais [...]. p. 1-6. Disponível em: <https://doi.org/10.5753/wcama.2023.230058>. Acesso em: 22 nov. 2024.

CHAVES, M. E. D.; CONCEIÇÃO, K. V. D.; GUERRERO, J. V. R. **Geoestatística no monitoramento da temperatura no estado da Bahia**. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, v. 13, n. 5, p. 652, 2018. Disponível em: <https://doi.org/10.18378/rvads.v13i5.6203>.

CHINNATHAMBI, J. K. **Amplifying big data utilization in healthcare analytics through cloud and Snowflake migration**. *European Journal of Computer Science and Information Technology*, v. 12, n. 6, p. 15-23, 2024. Disponível em: <https://doi.org/10.37745/ejcsit.2013/vol12n61523>.

CORREIA, I. d. S. et al. **Aplicação do módulo de monitoramento da qualidade do ar no sistema telemétrico para monitoramento ambiental nos sistemas de transportes públicos**. *Studies in Engineering and Exact Sciences*, v. 5, n. 1, p. 2957-2972, 2024. Disponível em: <https://doi.org/10.54021/seesv5n1-148>.

DATABRICKS. **Medallion Architecture**. 2023. Disponível em: <https://www.databricks.com/glossary/medallion-architecture>. Acesso em: 26 maio 2025.

DIAS, E. M. S.; PESSOA, Z. S. **Percepções sobre os riscos das mudanças climáticas no contexto da região semiárida do Rio Grande do Norte, Brasil**. *Desenvolvimento e Meio Ambiente*, v. 55, 2020. Disponível em: <https://doi.org/10.5380/dma.v55i0.73783>.

DWYER, J. G.; O'GORMAN, P. A. **Changing precipitation extremes: scaling and shifts in distributions**. *Geophysical Research Letters*, v. 44, n. 23, p. 12,066–12,074, 2017. Disponível em: <https://doi.org/10.1002/2017gl072855>.

GIORDONO, L.; BOUDET, H.; GARD-MURRAY, A. **Local adaptation policy responses to extreme weather events**. *Policy Sciences*, v. 53, n. 4, p. 609-636, 2020. Disponível em: <https://doi.org/10.1007/s11077-020-09401-3>.

GUIMARÃES, A. C. F. et al. **Clima e desastres naturais no Brasil: uma análise dos eventos extremos e suas implicações socioambientais**. *Research, Society*

and Development, v. 10, n. 9, e1781092021, 2021. Disponível em: <https://doi.org/10.33448/rsd-v10i9.17882>.

HAASE, C.; RÖSELER, T.; SEIDEL, M. **Metl: a modern etl pipeline with a dynamic mapping matrix**. 2022. Disponível em: <https://doi.org/10.48550/arxiv.2203.10289>.

HAUSTEIN, K. et al. **Real-time extreme weather event attribution with forecast seasonal ssts**. *Environmental Research Letters*, v. 11, n. 6, p. 064006, 2016. Disponível em: <https://doi.org/10.1088/1748-9326/11/6/064006>.

KEPNER, J. et al. **Apache Airflow: Orchestrating Data Pipelines**. O'Reilly Media, 2017. Disponível em: <https://www.oreilly.com/library/view/apache-airflow/9781492055361/>. Acesso em: 27 set. 2024.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**. 3. ed. Wiley, 2013. Disponível em: <https://www.wiley.com/en-us/The+Data+Warehouse+Toolkit%2C+3rd+Edition-p-9781118530801>. Acesso em: 1 nov. 2024.

KRESLAKE, J. M. **Perceived importance of climate change adaptation and mitigation according to social and medical factors among residents of impacted communities in the United States**. *Health Equity*, v. 3, n. 1, p. 124-133, 2019. Disponível em: <https://doi.org/10.1089/heq.2019.0002>.

LI, H.; DONG, M.; OTA, K.; GUO, M. **Pricing and repurchasing for big data processing in multi-clouds**. *IEEE Transactions on Emerging Topics in Computing*, v. 4, n. 2, p. 266-277, 2016. Disponível em: <https://doi.org/10.1109/tetc.2016.2517930>.

MARTINS, R. D.; FERREIRA, L. d. C. **Oportunidades e barreiras para políticas locais e subnacionais de enfrentamento das mudanças climáticas em áreas urbanas: evidências de diferentes contextos**. *Ambiente e Sociedade*, v. 13, n. 2, p. 223-242, 2010. Disponível em: <https://doi.org/10.1590/s1414-753x2010000200002>.

MEDEIROS, M. S. d.; GALLANI, J. d. A.; SWIECH, J. N. D. **Índices glicêmicos em pacientes diabéticos antes e após o uso do sistema de monitorização flash da glicose**. *Research, Society and Development*, v. 11, n. 15, e176111536858, 2022. Disponível em: <https://doi.org/10.33448/rsd-v11i15.36858>.

MEIRA, S. R. C.; OLIVEIRA, A. D. S. B.; SANTOS, C. O. **A contribuição da auditoria para a qualidade da gestão dos serviços de saúde / The audit's contribution to the quality of health service management**. *Brazilian Journal of Business*, v. 3, n. 1, p. 1021-1033, 2021. Disponível em: <https://doi.org/10.34140/bjbv3n1-058>.

MERLO, A. et al. **Estudo comparativo entre MongoDB e PostgreSQL usando embedding documents**. In: BRAZILIAN E-SCIENCE WORKSHOP (BRESKI 2024), 18., 2024. Anais [...]. p. 64-71. Disponível em: <https://doi.org/10.5753/breski.2024.244119>.

MOHR, S.; KUNZ, M.; RICHTER, A.; RUCK, B. **Statistical characteristics of convective wind gusts in Germany**. *Natural Hazards and Earth System Sciences*, v. 17, n. 6, p. 957–969, 2017. Disponível em: <https://doi.org/10.5194/nhess-17-957-2017>.

MONTEIRO, J. B.; ZANELLA, M. E. **A metodologia estatística dos eventos extremos de precipitação: uma proposta autoral para análise de episódios pluviométricos diários**. *Revista Brasileira de Climatologia*, v. 32, p. 494-516, 2023. Disponível em: <https://doi.org/10.55761/abclima.v32i19.16323>.

MORAES, R. A.; NETO, E. Q.; LAMPARELLI, R. A. C. **Relação entre o número de larvas do mosquito transmissor da dengue e elementos climáticos / Relationship between number of mosquitos larvae dengue transmitters and climate elements**. *Brazilian Journal of Animal and Environmental Research*, v. 3, n. 3, p. 2670-2675, 2020. Disponível em: <https://doi.org/10.34188/bjaerv3n3-172>.

NARAYANAN, S.; S, M.; ZEPHAN, P. **Real-time monitoring of data pipelines: exploring and experimentally proving that the continuous monitoring in data pipelines reduces cost and elevates quality**. *ICST Transactions on Scalable Information Systems*, 2024. Disponível em: <https://doi.org/10.4108/eetsis.5065>.

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION (NCDC). **Climate Data Online Web Services v2**. Disponível em: <https://www.ncdc.noaa.gov/cdo-web/webservices/v2#stations>. Acesso em: 10 nov. 2024.

NOMS. **IEEE/IFIP Network Operations and Management Symposium**, 2023. Disponível em: <https://doi.org/10.1109/noms56928.2023.10154377>.

OLIVEIRA, Á. R. G.; JUNIOR, A. J. P. R.; SILVA, N. L. **Avaliação do consumo de energia do hospital de clínicas da universidade federal do triângulo mineiro**. In: SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 2020, Goiânia. Anais [...]. Disponível em: <https://doi.org/10.48011/sbse.v1i1.2426>.

OPENWEATHERMAP. **OpenWeatherMap API**. Disponível em: <https://openweathermap.org/>. Acesso em: 10 nov. 2024.

PELLE, I. et al. **A comprehensive performance analysis of stream processing with Kafka in cloud native deployments for IoT use-cases**. In: IEEE/IFIP NETWORK OPERATIONS AND MANAGEMENT SYMPOSIUM (NOMS 2023), 2023,

Miami, FL. Anais [...]. Disponível em: <https://doi.org/10.1109/noms56928.2023.10154377>.

PROCÓPIO, D. F. et al. **Desvendando o futuro financeiro: a revolução dos sistemas de informação na contabilidade gerencial empresarial.** *Contribuciones a Las Ciencias Sociales*, v. 16, n. 11, p. 28800-28816, 2023. Disponível em: <https://doi.org/10.55905/revconv.16n.11-242>.

RUF, C. S. et al. **A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation.** *Scientific Reports*, v. 8, n. 1, 2018. Disponível em: <https://doi.org/10.1038/s41598-018-27127-4>.

SANTOS, L. A. R. dos et al. **Relação entre índice UV e variáveis meteorológicas em região tropical.** *Revista Brasileira de Meteorologia*, v. 31, n. 1, p. 51-58, 2016. Disponível em: <https://doi.org/10.1590/0102-778620130001>.

SATO, H. et al. **Complex plant responses to drought and heat stress under climate change.** *The Plant Journal*, v. 117, n. 6, p. 1873-1892, 2024. Disponível em: <https://doi.org/10.1111/tpj.16612>.

SENA, A.; BARCELLOS, C. **Informação sobre seca e saúde: subsídio para ações preventivas.** *P2p e Inovação*, v. 6, p. 16-32, 2019. Disponível em: <https://doi.org/10.21721/p2p.2019v6n1.p16-32>.

SILVA, M. C. B. D.; GONZÁLEZ, S. M. **Selfbi: tomada de decisão sob demanda do usuário utilizando dados da web.** In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS (SBBDD 2016), 31., Salvador, 2016. Anais [...]. Disponível em: <https://doi.org/10.5753/sbbd.2016.24334>.

SILVA, M. F. DA; BARBOSA, R. V. R. **Regime de ventos em cidades de diferentes regiões geográficas de Alagoas a partir de dados meteorológicos recentes.** *Revista Brasileira de Climatologia*, v. 31, p. 509-538, 2022. Disponível em: <https://doi.org/10.55761/abclima.v31i18.15606>.

TARANOV, K.; BYAN, S.; MARATHE, V. J.; HOEFLER, T. **Kafkadirect: zero-copy data access for Apache Kafka over RDMA networks.** In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2022, Philadelphia, PA. Proceedings [...]. p. 2191-2204. Disponível em: <https://doi.org/10.1145/3514221.3526056>.

TOMORROW.IO. **Tomorrow.io Data Sources.** Disponível em: <https://app.tomorrow.io/tomorrow-data-sources>. Acesso em: 3 maio 2025.

VAIDYANATHAN, A. et al. **Assessment of extreme heat and hospitalizations to inform early warning systems.** *Proceedings of the National Academy of Sciences*, v. 116, n. 12, p. 5420-5427, 2019. Disponível em: <https://doi.org/10.1073/pnas.1806393116>.

WANG, G. et al. **Building a replicated logging system with Apache Kafka.** *Proceedings of the VLDB Endowment*, v. 8, n. 12, p. 1654-1655, 2015. Disponível em: <https://doi.org/10.14778/2824032.2824063>.

WAZLAWICK, Raul. **Metodologia de Pesquisa para Ciência da Computação.** 2. ed. São Paulo: Ática, 2009.

XAVIER, A. L.; BONATTI, D.; CELASCHI, S. **Rain gauge simulator and first tests with a new mobile climate alert system in Brazil.** *Journal of the Brazilian Computer Society*, v. 22, n. 1, 2016. Disponível em: <https://doi.org/10.1186/s13173-016-0042-7>.

ZHANG, S. et al. **Using CYGNSS data to map flood inundation during the 2021 extreme precipitation in Henan Province, China.** *Remote Sensing*, v. 13, n. 24, p. 5181, 2021. Disponível em: <https://doi.org/10.3390/rs13245181>.

ZHAO, S.; CHANDRASHEKAR, M.; LEE, Y.; MEDHI, D. **Real-time network anomaly detection system using machine learning.** In: INTERNATIONAL CONFERENCE ON THE DESIGN OF RELIABLE COMMUNICATION NETWORKS (DRCN), 11., Kansas City, MO, 2015. Anais [...]. Disponível em: <https://doi.org/10.1109/drcn.2015.7149025>.