

@WeatherNit: uma Plataforma Orientada a Dados para Monitoramento de Chuvas e Ocorrências de Eventos Climáticos*

Marcos Lage¹, Fabio Victorino¹, Gustavo Muller Moreira¹, Bruno Cunha Sá¹, Aline Paes¹, Annie Amorim¹, Deborah Cholodoysky¹, Kaio Pereira¹, Gabriel Assis¹, Arthur Poustka¹, Paulo Alves¹, Andressa Nemirovsky², Nathalia Moura², Daniel de Oliveira¹

¹Instituto de Computação – Universidade Federal Fluminense (UFF)

²Secretaria Municipal de Defesa Civil e Geotecnia de Niterói

{mlage, alinepaes, danielcmo}@ic.uff.br

Resumo. *O número de eventos climáticos extremos têm aumentado em todo planeta, impactando fortemente grandes centros urbanos, em especial aqueles que cresceram de forma desordenada. Nessas regiões, enchentes e deslizamentos de terras são responsáveis por muitas mortes, todos os anos. Por isso, planejamentos que ajudem a se antecipar, reagir e evitar tais eventos são de fundamental importância. Neste artigo, apresentamos uma plataforma chamada @WeatherNit para o monitoramento de chuvas e eventos climáticos na cidade de Niterói. A plataforma permite a visualização interativa de dados históricos e de tempo real de volumes acumulados de chuvas e de ocorrências de enchentes e deslizamentos, integrados e armazenados usando um Data Lakehouse. A plataforma foi avaliada com um estudo de caso utilizando dados do CEMADEN e da Prefeitura de Niterói que demonstrou o potencial da abordagem no monitoramento de crises e no apoio ao desenvolvimento de políticas públicas.*

1. Introdução

O aumento crescente no número de eventos climáticos extremos ocorrendo no planeta tem motivado o desenvolvimento de soluções nas mais diversas áreas relacionadas ao clima [Mizutori and Guha-Sapir 2020, De Frenne et al. 2021]. Tais eventos podem variar desde inundações e secas até furacões e tornados, dependendo da região. O impacto desses eventos comumente é de grandes proporções, e seus efeitos podem ser ainda mais devastadores nos grandes centros urbanos. Devido ao crescimento desordenado, grandes centros urbanos tendem a ter muitas moradias em áreas de risco e áreas extensas impermeabilizadas, que fazem com que o fluxo de água da chuva corra rapidamente para rios (que podem estar canalizados) que não possuem capacidade de escoamento suficiente, causando assim inundações e deslizamentos [Thorndahl and Willems 2008]. Assim, torna-se uma prioridade para a administração pública de grandes centros urbanos o planejamento para se antecipar e/ou reagir à ocorrência de tais eventos.

Algumas abordagens encontradas na literatura [de Souza et al. 2022] propõem modelos matemáticos para realizar previsões relacionadas ao clima. Porém, esses trabalhos têm foco em eventos específicos, e em regiões pequenas (*i.e.*, micro climas). Quando consideramos centros urbanos com áreas comumente extensas (*e.g.*, a cidade de Niterói, usada como estudo de caso nesse artigo, possui 129,3 km² de área), tais abordagens podem não

*O presente trabalho foi realizado com apoio da Prefeitura de Niterói (via edital PDPA), CNPq e FAPERJ.

apresentar resultados satisfatórios. Em especial, se considerarmos a modelagem de eventos climáticos extremos (*e.g.*, grandes tempestades como as que ocorreram em Petrópolis em 2022), a complexidade do fenômeno pode tornar muito difícil a criação de um modelo matemático. Uma alternativa nesse caso é usar dados históricos tanto para analisar e entender os eventos (e saber reagir às suas consequências) quanto para treinar modelos preditivos. Algumas abordagens já vem sendo propostas nesse sentido na área de aprendizado de máquina [Kumar et al. 2021, Rolnick et al. 2022]. Um grande desafio é como treinar modelos com dados heterogêneos (comumente multimodais), em múltiplas granularidades e diferentes escalas temporais.

Entretanto, para que modelos de aprendizado de máquina possam ser aprendidos a partir de fontes heterogêneas, existem desafios anteriores ao treinamento do modelo, que são a obtenção, integração, tratamento e disponibilização dos dados. Estima-se que 80% do tempo do desenvolvimento de aplicações no contexto de Ciência de Dados são gastos com tarefas relacionadas à preparação dos dados a serem analisados e/ou usados no treinamento. Dessa forma, desenvolver plataformas que sejam capazes de obter e integrar dados que possam ser usados para análise e treinamento de modelos preditivos passa a ser uma prioridade. Esse dados podem variar desde índices pluviométricos disponibilizados por estações pluviométricas (sistemas de aquisição de dados por meio de sensores que fornecem o índice pluviométrico em uma dada área) até imagens de câmeras de monitoramento e de satélite. A maior dificuldade dessa integração é que os dados disponibilizados comumente se encontram em granularidades, formatos e resoluções diferentes. Por exemplo, as estações pluviométricas da prefeitura de Niterói capturam os dados a cada 15 minutos enquanto que o CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais) captura os dados a cada 10 minutos, quando há ocorrência de precipitação, e a cada hora, quando não há precipitação. Além disso, mesmo que os dados tenham sido tratados e integrados, o usuário especialista deve possuir modos de analisá-los. Por mais que uma representação tabular seja útil, muitas análises são facilitadas se apoiadas por técnicas de visualização [Chan 2006, Diehl et al. 2015], *e.g.*, identificar áreas com altos acumulados de chuva a cada hora em um mapa de uma região.

Esse artigo apresenta a @WeatherNit, uma plataforma para monitoramento de chuvas e eventos climáticos, desenvolvida em parceria com a Secretaria Municipal de Defesa Civil e Geotecnia de Niterói. A @WeatherNit permite a visualização interativa de dados de chuvas e ocorrências associadas à eventos climáticos por meio de dados integrados e armazenados em um *Data Lakehouse*. Um *Data Lakehouse* é a união dos conceitos de *Data Lake* e *Data Warehouse*, e sua proposta é prover um repositório integrado de dados em grande volume e em formatos diferentes, ao mesmo tempo que proporciona garantias de transações ACID (acrônimo em inglês dos termos *Atomicity*, *Consistency*, *Isolation*, *Durability*). Assim, dados estruturados como índices pluviométricos são armazenados em *Data Warehouses* (DW), e dados não estruturados são armazenados em seus formatos brutos para que possam ser posteriormente analisados e processados. A @WeatherNit foi avaliada com um estudo de caso de integração e análise dos dados do CEMADEN e da Prefeitura de Niterói. Resultados mostraram o potencial da abordagem proposta. O restante deste artigo está organizado da seguinte forma: Na Seção 2, discutimos alguns trabalhos relacionados. Na Seção 3 é apresentada a plataforma @WeatherNit, enquanto que na Seção 4 realizamos um estudo de caso simplificado. Finalmente, na Seção 5, concluímos o trabalho.

2. Trabalhos Relacionados

Existem alguns trabalhos na literatura que propõem ferramentas e plataformas para integração de dados climáticos, em especial, de dados pluviométricos. [Esplugues et al. 2013] propõem a aplicação de técnicas de aprendizado de máquina em dados pluviométricos para analisar os períodos de seca/chuva na Espanha. Nessa abordagem, os dados pluviométricos são coletados na AEMET (Agência Estatal de Meteorología da Espanha) e é utilizado o SPI (*Standard Precipitation Index*), para o monitoramento de condições associadas a secas e excesso de chuva. Os dados pluviométricos são integrados em um ambiente de DW, como na @WeatherNit. [Morais and Ferreira 2015] propõem um banco de dados para armazenar de forma integrada dados pluviométricos do Estado de Goiás. Os dados pluviométricos são coletados da ANA (Agência Nacional de Águas) e os dados de imagens do *Tropical Rainfall Measuring Mission* (TRMM), controlado pela NASA. Finalmente, a abordagem proposta por [Salas et al. 2020] introduz um *framework* para realizar a fusão de múltiplas fontes de dados heterogêneos de chuva. A abordagem coleta dados hidrológicos de múltiplas fontes e os processa usando o sistema de gerência de *workflows* VisTrails. Por meio do VisTrails, o usuário é capaz de personalizar *scripts* para processamento dos dados. Diferentemente da @WeatherNit, a abordagem de [Salas et al. 2020] não modela seus dados como DWs, e não pré-agrega os dados, o que faz com que consultas se tornem computacionalmente intensivas.

3. A Plataforma @WeatherNit

A plataforma @WeatherNit tem como objetivo obter dados multimodais (*e.g.*, dados estruturados, imagens) de múltiplas fontes externas e com diferentes granularidades, e integrá-los em um único repositório para posterior consulta e análise. A arquitetura da @WeatherNit é apresentada na Figura 1, e é composta de quatro camadas principais: (i) Fontes de Dados Externas, (ii) Camada de Armazenamento/Integração, (iii) Camada de Análise/Visualização, e (iv) Camada de Treinamento de Modelos.

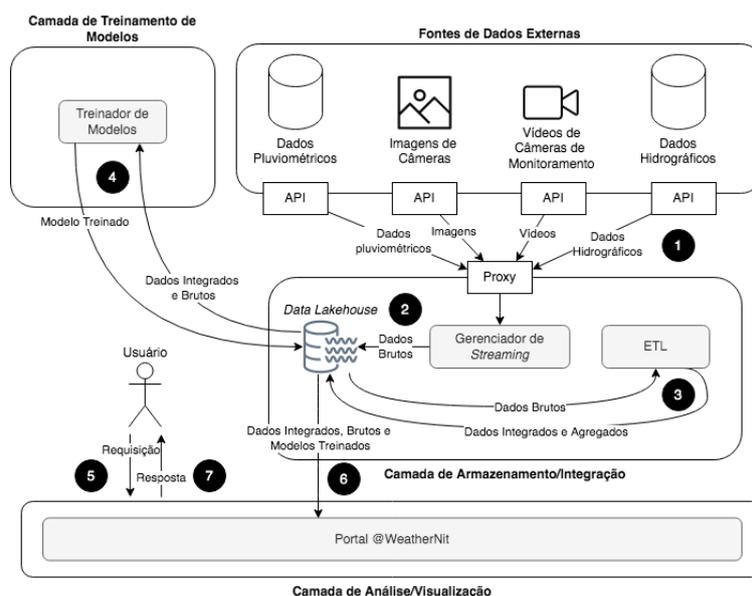


Figura 1. Arquitetura da plataforma @WeatherNit.

A camada das Fontes de Dados Externas é onde os usuários podem obter os dados brutos para processamento e análise. Encontram-se nessa camada os principais provedores

de informação, como o CEMADEN, a ANA, imagens e vídeos de câmeras de monitoramento. Na plataforma @WeatherNit assumimos que cada fonte de dados possui uma API própria que é usada pela @WeatherNit para a obtenção dos dados (passo ① na Figura 1). Essa obtenção de dados é agendada, de forma que possa ser realizada automaticamente sem necessidade de intervenção do usuário. Na Camada de Armazenamento/Integração, os dados coletados (estruturados e não-estruturados) são repassados a um componente *Gerenciador de Streaming* (que pode ser implementado por meio de plataformas de processamento de *streams* como o Kafka) e armazenados no *Data Lakehouse* (passo ②). Após o armazenamento dos dados brutos (e.g., em formato CSV, JSON, MP4 etc.), o componente de ETL (*Extract, Transform and Load*) é capaz de carregar os dados e realizar agregações de forma a armazená-los em um *Data Mart* dentro do *Data Lakehouse* (passo ③). O *script* que executa o processo de ETL deve ser fornecido pelo usuário. Os dados brutos e pré-agregados podem ser então consumidos na Camada de Treinamento de Modelos para treinar modelos preditivos por meio de técnicas de aprendizado de máquina (passo ④). O componente *Treinador de Modelos* é um *script* definido pelo usuário que realiza o treinamento de acordo com o objetivo definido. O modelo treinado é carregado no *Data Lakehouse* assim que gerado. Finalmente, na Camada de Análise/Visualização o usuário pode submeter requisições ao Portal @WeatherNit (passo ⑤). O Portal @WeatherNit então acessa os dados brutos, pré-agregados e os modelos treinados (passo ⑥) e responde ao usuário (passo ⑦) com uma visualização ou algum resultado de classificação ou predição (e.g., a partir de uma imagem de câmera, informar se uma região está alagada ou não). A plataforma @WeatherNit se encontra em homologação por parte de especialistas e pode ser acessada em <http://pdpachuvvas.ic.uff.br>. Em sua versão atual, as funcionalidades de carga e análise se encontram disponíveis, e o volume de dados carregados atualmente no sistema é em torno de 6GB. Ainda, o treinamento de modelos de aprendizado de máquina (para previsão de chuvas e identificação de inundações) se encontra em desenvolvimento.

Em relação as tecnologias utilizadas no desenvolvimento da plataforma, as camadas de Armazenamento/Integração e Análise/Visualização de dados foram implementadas utilizando as linguagens JavaScript e Python. Mais especificamente, a camada de Armazenamento/Integração foi construída utilizando o *framework Express.js*. O gerenciador de *Streaming* e o componente ETL foram escritos em Python e o *Data Lakehouse*, além dos dados brutos, contém o PostGIS. Já o Portal @WeatherNit foi construído utilizando o *framework* Angular e as bibliotecas Mapbox e Vega-lite para a construção de mapas e visualizações. Por fim, os componentes da Camada de Treinamento de Modelos estão sendo escritos em Python e R.

4. Estudo de Caso

Nesta seção, apresentamos um estudo de caso simplificado para apoiar as especialistas da área de Defesa Civil, co-autoras deste artigo. Com esse estudo de caso, mostramos como a @WeatherNit é capaz de auxiliá-las em suas tarefas diárias de monitoramento e análise de chuvas e eventos climáticos. O estudo usa como base dados carregados do CEMADEN e da Prefeitura de Niterói entre os anos de 2014 e 2022. Os dados do CEMADEN e das estações da prefeitura de Niterói são acessados via APIs próprias. O estudo se concentra na avaliação de chuvas na cidade de Niterói na data 07/06/2016, e foi tomado como horário-base às 00:00. A partir do horário-base, calculamos o índice acumulado nos últimos 15 min, 30 min, 1 hora, 6 horas, 12 horas, 24 horas, 48 horas, 72 horas, 96 horas e 1 mês. Conforme apresentado na Figura 2, o usuário é capaz de definir a data e hora-base para a consulta, e

visualizar no mapa as estações pluviométricas e os respectivos índices pluviométricos.

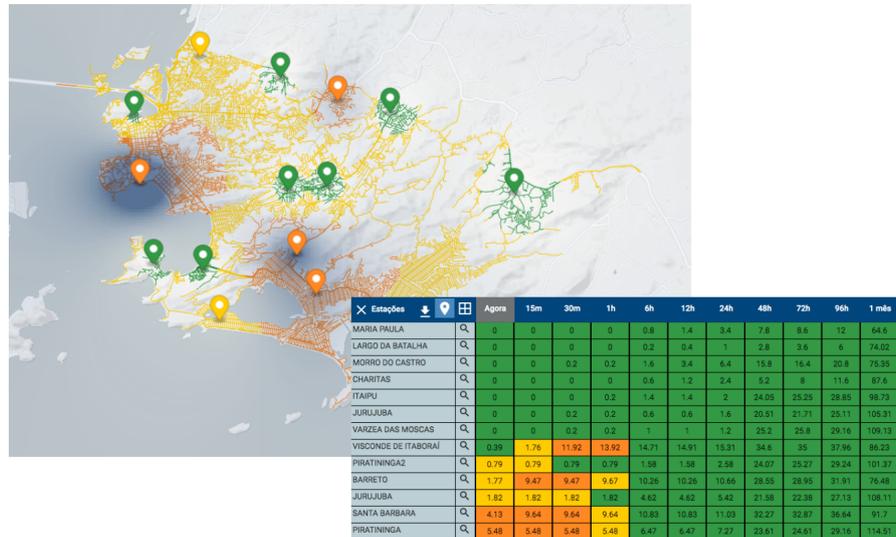


Figura 2. Monitoramento dos dados pluviométricos na plataforma @WeatherNit. Os pins representam as estações pluviométricas e suas cores a intensidade de chuva no momento. A Tabela de estações apresenta o acumulado de chuva para cada estação nos últimos 15 min, 30 min, 1h, 6h, 12h, 24h, 48h, 72h, 96h e 1 mês.

É importante ressaltar que apesar da @WeatherNit possuir no *Data Lakehouse* somente dados pluviométricos associados à determinados pontos do espaço, *i.e.*, onde se encontram as estações (o que faz com que os dados sejam especialmente esparsos), por meio do uso de técnicas de interpolação baseadas em distância [Lu and Wong 2008], a plataforma constrói uma grade com valores acumulados de chuva cobrindo toda a região geográfica de interesse (no caso a cidade de Niterói). Utilizando estes valores interpolados, podemos visualizar índices pluviométricos através de mapas de calor, como os exibidos na Figura 2. Em especial, na data escolhida para o estudo de caso, fortes chuvas começaram a atingir a cidade de Niterói. É possível visualizar índices pluviométricos mais elevados na região do bairro do Ingá (estação de cor laranja mais a esquerda na Figura 2), e a área mais escura em torno da estação é produto da interpolação realizada. Além das estações, os usuários podem escolher apresentar na interface as ocorrências (*e.g.*, deslizamentos, quedas de barreira) que estão acontecendo por conta das chuvas.

Além da visualização dos dados pluviométricos no mapa, a plataforma @WeatherNit também é capaz de apresentar os dados de forma tabular para o usuário, conforme apresentado na Figura 2. O usuário pode então exportar esses dados (em formato tabular ou a imagem gerada) ou salvar suas consultas na própria plataforma para posterior análise. Além das análises espaciais de chuvas, a @WeatherNit já é capaz de importar imagens de câmeras de monitoramento para o *Data Lakehouse*. Essas imagens estão sendo usadas para o treinamento de modelos que possam identificar inundações ou problemas em tempo real na cidade, mesmo em locais onde não há nenhuma estação pluviométrica. Esse tipo de identificação permitirá que a Defesa Civil possa responder o mais rápido possível. É importante ressaltar que as consultas que são submetidas ao *Data Mart* da @WeatherNit foram otimizadas para que executassem em até 5 segundos, uma vez que tempos de execução de consultas altos podem reduzir a eficácia e at inviabilizar a construção de sistemas de exploração e análise visual de dados.

5. Conclusões e Trabalhos Futuros

Tradicionalmente, estudos sobre o clima necessitam processar um grande volume de dados, comumente heterogêneos, multimodais e representados em diferentes granularidades temporais e espaciais. Esses dados podem ser obtidos a partir de uma miríade de fontes, e integrá-los de forma que possam ser analisados, visualizados ou consumidos por outras abordagens, não é uma tarefa trivial. Este artigo apresenta a plataforma @WeatherNit, que tem como objetivo carregar dados de fontes externas, integrá-los de acordo com as necessidades dos usuários e disponibilizar mecanismos analíticos e de visualização. De forma a avaliar a plataforma @WeatherNit, executamos um estudo de caso simplificado que integrou dados das bases da Prefeitura de Niterói e do CEMADEN, entre os anos de 2014 e 2022. As consultas executadas e as visualizações geradas mostram o potencial da @WeatherNit. Trabalhos futuros incluem adicionar mecanismos de identificação de inundações a partir de imagens de câmeras de monitoramento e dados pluviométricos, além da sugestão de rotas viárias otimizadas para atendimento às ocorrências (*e.g.*, deslizamentos), que são consequências de eventos climáticos. Além disso, pretendemos desenvolver um mecanismo para detectar falhas nos dados de origem, como por exemplo identificar possíveis *outliers* ou completar dados faltantes, a fim de aumentar a qualidade dos dados.

Referências

- Chan, W. W.-Y. (2006). A survey on multivariate data visualization. *Department of Computer Science and Engineering, Hong Kong University of Science and Technology*, 8(6):1–29.
- De Frenne, P., Lenoir, J., Luoto, M., Scheffers, B., et al. (2021). Forest microclimates and climate change: Importance, drivers and future research agenda. *Global Change Biology*, 27(11):2279–2297.
- de Souza, C. V. F., da Cunha Luz Barcellos, P., Crissaff, L., Cataldi, M., Miranda, F., and Lage, M. (2022). Visualizing simulation ensembles of extreme weather events. *Computers & Graphics*, 104:162–172.
- Diehl, A., Pelorosso, L., Delrieux, C., Saulo, C., Ruiz, J., Gröller, M. E., and Bruckner, S. (2015). Visual analysis of spatio-temporal data: Applications in weather forecasting. In *Computer Graphics Forum*, number 3 in 34, pages 381–390.
- Esplugues, F. B., Gramaje, M. d. C. P., and García-Haro, F. J. (2013). Técnicas de minería de datos para el análisis de periodos de sequía en españa. *Revista Tiempo y Clima*, 5(30).
- Kumar, P., Chandra, R., Bansal, C., Kalyanaraman, S., Ganu, T., and Grant, M. (2021). Micro-climate prediction - multi scale encoder-decoder based deep learning framework. KDD, page 3128–3138.
- Lu, G. Y. and Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, 34(9):1044–1055.
- Mizutori, M. and Guha-Sapir, D. (2020). Human cost of disasters 2000-2019. Technical report, United Nations Office for Disaster Risk Reduction.
- Morais, L. d. and Ferreira, N. C. (2015). Banco de dados pluviométricos integrados por dados do sensor trmm e estações pluviométricas no estado de goiás. *Anais Eletr.*, 17.
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2022). Tackling climate change with machine learning. *ACM Comput. Surv.*, 55(2).
- Salas, D., Liang, X., Navarro, M., Liang, Y., and Luna, D. (2020). An open-data open-model framework for hydrological models’ integration, evaluation and application. *Environ. Model. Softw.*, 126:104622.
- Thorndahl, S. and Willems, P. (2008). Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series. *Water Research*, 42(1):455–466.