

Análise de Dados Pluviométricos Multi-fonte Baseada em Técnicas OLAP e de Visualização: uma Abordagem Prática*

Lorena Christ'na Nascimento¹, Lucas Knust¹, Ramon Santos¹
Bruno Cunha Sá¹, Gustavo Muller Moreira¹, Fabiola de Souza Freitas²,
Nathalia Moura², Marcos Lage¹, Daniel de Oliveira¹

¹Instituto de Computação – Universidade Federal Fluminense (UFF)

²Secretaria Municipal de Defesa Civil e Geotecnia de Niterói

{lorennachrist, lucasknust, santosramon, bcunha, gustavomuller}@id.uff.br

{mlage, danielcmo}@ic.uff.br

Abstract. *Climate studies have gained relevance due to the increase in climatic events with severe impacts observed in the last decade, especially in urban areas. For example, large volumes of precipitation can cause floods and landslides, impacting city traffic and even costing citizens' lives. In order to make it possible to monitor rainfall volumes, pluviometers are installed throughout the country. However, such stations are controlled by multiple organizations and produce data in different space/time resolutions and formats. This paper proposes TEMPO, a system that uses OLAP (Online Analytical Processing) techniques to propose efficient storage, query, and analysis mechanisms to handle pluviometers data. To evaluate the tool, we present a case study showing the integration and analysis of data from CEMADEN and Alerta Rio.*

Resumo. *Estudos sobre o clima têm ganhado relevância devido ao aumento do número de eventos climáticos com impactos severos observados na última década, em especial em áreas urbanas. Por exemplo, ocorrências de grandes valores acumulados de chuva podem causar inundações e deslizamentos de terra, impactando o trânsito da cidade e até mesmo custando a vida de cidadãos. Para possibilitar o monitoramento de volumes de chuva, estações pluviométricas se encontram instaladas pelo país. Entretanto, tais estações são controladas por múltiplas organizações e oferecem dados em formatos distintos. Neste artigo propomos a ferramenta TEMPO (sisTema dE Monitoramento Pluviométrico) que utiliza técnicas OLAP (Online Analytical Processing) para propor mecanismos de armazenamento, consulta e análise eficientes de dados pluviométricos. Para avaliar a ferramenta, apresentamos um estudo de caso de integração e análise dos dados do CEMADEN e do Alerta Rio.*

1. Introdução

Os estudos relacionados ao clima têm se tornado cada vez mais importantes nos últimos anos principalmente devido ao aumento do número de eventos climáticos que vêm ocorrendo (e.g., inundações, secas, etc.) em múltiplas regiões do planeta, incluindo o Brasil [Mizutori and Guha-Sapir 2020]. Em especial, em grandes centros urbanos o impacto

*Os autores gostariam de agradecer a FAPERJ, CAPES e CNPq por financiarem o artigo.

de eventos climáticos pode ser severo. Devido ao crescimento não ordenado, centros urbanos tendem a possuir grandes áreas impermeabilizadas, que fazem com que o fluxo de água da chuva corra rapidamente para locais como rios e baixadas, que não possuem capacidade de escoamento para o volume de água recebido, causando assim transbordamentos e inundações [Thorndahl and Willems 2008].

Tomemos como exemplo o caso da cidade de Niterói (esse exemplo será usado consistentemente ao longo do artigo). Diversas vias importantes da cidade sofrem com inundações frequentes como a Av. Marquês do Paraná. Apesar dos esforços da prefeitura para realizar intervenções nesses locais, a capacidade de escoamento ainda não é suficiente. Com a ocorrência de tempestades repentinas ou pancadas de chuva intermitentes, essas ruas se tornam intransitáveis e impactam o trânsito de toda a cidade (já que são as principais saídas em direção ao Rio de Janeiro). Além disso, as inundações impactam fortemente as pessoas que residem em áreas de risco (*e.g.*, encostas, margens de rios, *etc.*), como foi o caso do deslizamento no Morro do Bumba, ocorrido em 2010¹, em que 47 morreram. Dessa forma, monitorar o acúmulo de água de chuva em cidades como Niterói é uma tarefa prioritária para órgãos governamentais (*e.g.*, a Defesa Civil).

De forma a permitir tal monitoramento, um conjunto de estações pluviométricas (sistemas de aquisição de dados por meio de sensores que fornecem o índice pluviométrico em uma dada área) se encontram instaladas por todo o país. Apesar dessas estações representarem um avanço no que se refere ao monitoramento das chuvas (já que fornecem dados quase que em tempo real), ainda existem desafios na coleta, tratamento e análise desses dados. De fato, as estações existentes são controladas por múltiplas organizações. No caso de Niterói, um conjunto de estações localizadas na cidade são gerenciadas pela prefeitura enquanto que um outro subconjunto é gerenciado pelo CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais) [CEMADEN 2021]. Os órgãos que monitoram as chuvas devem ter acesso aos dados de todas as estações de forma a aumentar a área de cobertura do monitoramento, fazendo com que as ações para mitigar ocorrências como alagamentos e deslizamentos de encostas sejam mais efetivas. Entretanto, tal integração de dados de múltiplas estações (multi-fonte) não é uma tarefa trivial.

Os dados disponibilizados por essas estações podem se encontrar em granularidades e resoluções diferentes, *e.g.*, as estações da prefeitura de Niterói capturam os dados a cada 15 minutos enquanto que o CEMADEN captura os dados a cada 10 minutos, quando há ocorrência de precipitação, e a cada hora, quando não há precipitação. Além disso, podem existir diferenças na representação da localização de cada estação. De fato, algumas estações informam suas posições geográficas em latitude e longitude, enquanto outras fornecem apenas o nome do bairro onde estão instaladas, sem uma localização exata. Finalmente, mesmo que os dados tenham sido tratados e integrados, o usuário especialista deve possuir modos de analisar tais dados. 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. Sendo assim, prover um ambiente em que o usuário especialista possa realizar consultas sobre dados de múltiplas estações (independente de quem gerencia a mesma) e utilize técnicas de visualização para apoiar

¹<https://glo.bo/2PmFyBJ>

seu trabalho é uma tarefa relevante e desafiadora.

De forma a oferecer mecanismos para armazenamento, consulta e análise eficientes de dados pluviométricos, esse artigo apresenta a ferramenta TEMPO (sistema de Monitoramento Pluviométrico). A TEMPO é baseada em técnicas OLAP (*Online Analytical Processing*) [Cuzzocrea 2015] para representação multidimensional dos dados. Técnicas OLAP permitem representar os dados de forma que consultas analíticas tradicionalmente *CPU-bound* sejam executadas de forma mais eficiente. Um *Data Warehouse* (DW) [Kimball and Ross 2002] é proposto para representação dos dados pluviométricos capturados em múltiplas fontes heterogêneas. Além disso, estratégias de visualização de dados espaciais foram acopladas ao DW para a exibição de mapas com os índices pluviométricos das estações selecionadas para análise. A TEMPO foi avaliada com um estudo de caso de integração e análise dos dados do CEMADEN e do Alerta Rio. Resultados mostraram o potencial da abordagem proposta. A ferramenta TEMPO se encontra no contexto do projeto “Niterói Organizada e Segura: Estudo do Impacto das Chuvas”, realizado em parceria com a Prefeitura de Niterói no contexto do Programa de Desenvolvimento de Projetos Aplicados (PDPA).

Esse artigo se encontra organizado em quatro seções além dessa introdução. Na Seção 2, conceitos de DW são apresentados, assim como os trabalhos relacionados. Na Seção 3 é apresentada a ferramenta TEMPO, enquanto que na Seção 4 a mesma é avaliada. Finalmente, na Seção 5, concluímos o presente artigo.

2. Referencial Teórico e Trabalhos Relacionados

Essa seção apresenta os conceitos básicos de *Data Warehouses* e discute os trabalhos relacionados à abordagem usada na ferramenta TEMPO.

2.1. OLAP e *Data Warehousing*

Data Warehouses são repositórios para grandes volumes de dados coletados de múltiplas fontes por meio de processos apropriados de Extração, Transformação e Carregamento (ETL) [Kimball and Ross 2002]. Segundo [Inmon et al. 2005], os DWs devem lidar com os dados de uma maneira variável no tempo, não volátil e orientada ao assunto, o que permite que um usuário possa consultar os dados em seu grão mais fino, mas também agregados por níveis distintos de granularidade. Os DWs possuem características diferentes dos bancos de dados tradicionais uma vez que são desenvolvidos para otimizar a inserção e a consulta aos dados. Os DWs dependem fortemente da *modelagem dimensional* [Kimball and Ross 2002] para manter sua estrutura lógica compreensível, bem como para otimizar o desempenho da consulta.

O modelo dimensional é baseado no conceito de *tabelas fato* e *tabelas de dimensões* para armazenar os dados. As tabelas fato contêm os valores de interesse de consulta (*e.g.*, os índices pluviométricos), enquanto que as dimensões qualificam esse dado quantitativo. Um modelo dimensional tradicionalmente segue um dos seguintes padrões: (i) esquema estrela (*star schema*) e (ii) floco de neve (*snowflake*) [Han et al. 2012]. O esquema estrela baseia-se no projeto de uma tabela fato central, enquanto as dimensões restantes são projetadas como tabelas de dimensão que se associam com a fato por meio de chaves estrangeiras. Já no esquema floco de neve, as tabelas de dimensão são normalizadas com relacionamentos “um-para-muitos” entre os atributos normalizados. Dessa

forma, o esquema estrela e o floco de neve são padrões concorrentes a serem escolhidos de acordo com a disponibilidade de disco e poder de processamento computacional [Kimball and Ross 2002].

Um DW pode ser implementado de três formas [Han et al. 2012]: (i) Processamento Analítico *Online* Relacional (ROLAP), (ii) Processamento Analítico *Online* Multidimensional (MOLAP) e (iii) Processamento Analítico *Online* Híbrido (HOLAP). No ROLAP, o modelo multidimensional é implementado em um Sistema de Gerência de Banco de Dados (SGBD) relacional (*e.g.*, MySQL, PostgreSQL). No MOLAP, o modelo multidimensional é implementado em estruturas vetoriais que permitem indexação rápida de dados pré-agregados. No HOLAP, existe a combinação do armazenamento do ROLAP com os cálculos rápidos do MOLAP. Nesse artigo, utilizaremos a abordagem ROLAP com o SGBD PostgreSQL. Uma vez modelados, os DWs devem ser carregados com os dados de múltiplas fontes. Essa carga é realizada por meio de um processo chamado de ETL (Extração, Transformação e Carga). Após a ingestão de dados, os DWs são capazes de oferecer consultas multidimensionais em níveis distintos de granularidade por meio de relatórios dinâmicos e interativos [Han et al. 2012].

2.2. Abordagens Existentes para Gerência de Dados Pluviométricos

Podemos encontrar algumas abordagens na literatura que propõem o uso de técnicas OLAP e de visualização para dados pluviométricos. A abordagem proposta por [Salas et al. 2020] propõe uma infraestrutura para realizar a fusão de múltiplas fontes de dados heterogêneos por meio de uma arquitetura de dados abertos. A abordagem é capaz de coletar dados hidrológicos de múltiplas fontes e processá-los por meio de um sistema de *workflow* (VisTrails), onde o usuário é capaz de personalizar *scripts* para processamento dos dados. Diferentemente da TEMPO, a abordagem de [Salas et al. 2020] não considera o uso de *Data Warehouses* para manter dados pré-agregados nem disponibiliza estratégias de visualização de dados espaciais.

A abordagem proposta por [Morais and Ferreira 2015] propõe um banco de dados para armazenar informações sobre o comportamento pluviométrico no Estado de Goiás. Os dados pluviométricos foram coletados na ANA (Agência Nacional de Águas) e os dados de imagens do *Tropical Rainfall Measuring Mission* (TRMM), controlado pela NASA. Apesar dos autores explicitarem a metodologia utilizada para elaborar o trabalho, o *schema* do banco de dados não foi apresentado, nem foi discutido se a abordagem considera dados multi-fonte. A abordagem proposta por [Esplugues et al. 2013] propõe a aplicação de técnicas de mineração de dados pluviométricos para analisar os períodos de seca na Espanha. Os dados pluviométricos foram coletados através da AEMET (Agência Estatal de Meteorología) e foi utilizado o SPI (*Standard Precipitation Index*), comumente utilizado para o monitoramento de condições associadas a secas e excesso de chuva. Apesar de os autores terem explicado os métodos utilizados para elaboração do trabalho, assim como no caso da abordagem anterior, o banco de dados e a ferramenta de ETL utilizados para o desenvolvimento não foram discutidos, limitando o entendimento da proposta apresentada.

3. Abordagem Proposta: TEMPO

A TEMPO é uma ferramenta que tem como objetivo captar dados pluviométricos multi-fonte com diferentes granularidades e integrá-los em um único repositório para posterior

consulta e análise. A arquitetura da TEMPO é apresentada na Figura 1, e é composta de três camadas principais: (i) Fontes de Dados, (ii) Camada de Armazenamento e (iii) Camada de Análise.

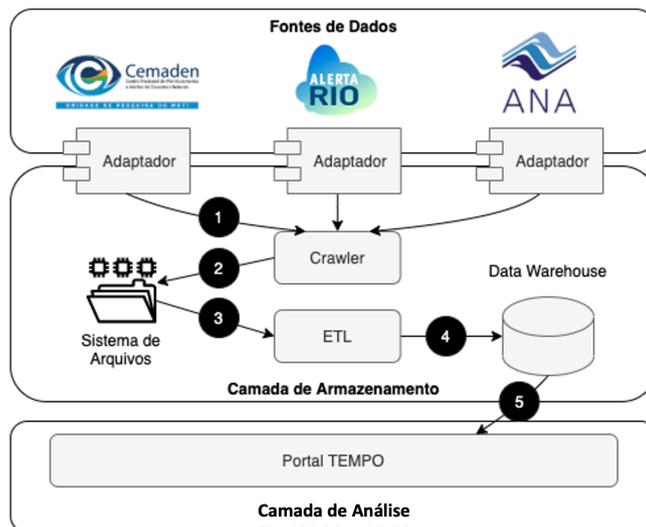


Figura 1. Arquitetura da TEMPO.

A camada das Fontes de Dados é onde podemos obter os dados brutos para processamento e análise. Nessa camada se encontram os principais provedores de informação como o CEMADEN e o Alerta Rio (exclusivo da cidade do Rio de Janeiro). É importante ressaltar que cada um desses provedores disponibiliza seus dados de diferentes formas. Assim, cada provedor possui um *Adaptador* associado dentro da TEMPO. O componente *Crawler* realiza o *download* dos dados invocando os Adaptadores (passo 1), que por sua vez acessam APIs públicas, servidores FTP, *etc.* Os dados coletados são armazenados no *Sistema de Arquivos* (área de *staging*) da Camada de Armazenamento em seu estado bruto (passo 2). Além de armazenar os dados brutos (*e.g.*, em formato CSV, JSON, *etc.*), na área de *Staging* são criadas tabelas *Staging* para cada fonte de dados consideradas. Essas tabelas possuem os mesmos atributos que constam em cada um dos arquivos captados nas fontes de dados.

Uma vez que os dados brutos coletados nas fontes e as tabelas de *Staging* se encontram disponíveis, o componente *ETL* realiza as devidas transformações nos dados (*e.g.*, ajuste de granularidade, padronização de endereços e localizações, agregações, *etc.* - passo 3). Em sua versão atual, a TEMPO utiliza a ferramenta *Pentaho Data Integration* versão 9.1 para carregar os dados no DW, que chamamos de DW-TEMPO (passo 3). O DW-TEMPO segue o esquema estrela. Conforme mencionado anteriormente, esse esquema consiste em uma tabela fato, que fica localizada em uma posição central do modelo, ligada por meio de múltiplos relacionamentos *1:n* com as chamadas tabelas de dimensão. O esquema estrela foi o escolhido uma vez que as dimensões são desnormalizadas, o que facilita o processo de ETL e a agregação dos dados. Optamos por seguir a abordagem ROLAP, e instanciamos o DW-TEMPO no SGBD PostgreSQL versão 4.26.

O *schema* do DW-TEMPO é apresentado na Figura 2. Esse *schema* possui uma tabela fato e quatro tabelas de dimensão. As tabelas de dimensão possuem uma *primary-*

key (PK), chave única de identificação para cada tupla. A tabela fato, por sua vez, está vinculada a todas as tabelas de dimensão por um relacionamento (1:n). Dessa forma, haverá uma *foreign-key* (FK) correspondente a cada dimensão. A tabela fato *Chuvas* contém os atributos *id_indice_pluviometrico* (PK), *indice_pluviometrico* (quantidade de interesse para análise), *id_fonte* (FK), *id_tempo* (FK), *id_localidade* (FK) e *id_estacao* (FK), onde todas as *foreign keys* referenciam as tabelas de dimensão associadas. A dimensão *Localidade* representa os locais onde as estações pluviométricas se encontram (onde as medições foram realizadas), e contém os atributos: *id_localidade* (PK), *latitude*, *longitude*, *rua*, *bairro*, *municipio* e *uf*. A dimensão *Fonte* representa a proveniência do dado, *i.e.*, de qual fonte esse dado foi obtido. Essa tabela contém os atributos: *id_fonte* (PK), *nome_fonte* e *url*. A dimensão *Tempo* representa os possíveis intervalos de tempo para consulta, e contém os atributos: *id_tempo* (PK), *minuto*, *hora*, *dia*, *mes*, *semestre* e *ano*. Por fim, a dimensão *Estação* representa as estações que foram consideradas para análise, e contém os atributos: *id_estacao* (PK), *cod_estacao_original* e *nome_estacao_original*. O *dump* do DW-TEMPO contendo os dados utilizados nesse artigo podem ser obtidos em <https://bit.ly/3dcnApY>.

Finalmente, na Camada de Análise, o Portal TEMPO fornece mecanismos de visualização de dados espaciais para o usuário. Apesar do usuário poder submeter consultas diretamente ao DW-TEMPO, muitas vezes a representação tabular retornada pela consulta pode dificultar a análise. Por exemplo, consideremos que um usuário deseja comparar a evolução dos índices pluviométricos em um período de tempo de interesse. Se o usuário puder, em vez de inspecionar uma planilha contendo uma série de atributos (*e.g.*, índice pluviométrico, a localização da estação e a data), visualizar um conjunto de mapas com valores acumulados de chuva na região de interesse, conseguirá identificar padrões de precipitação potencialmente relevantes de forma mais efetiva.

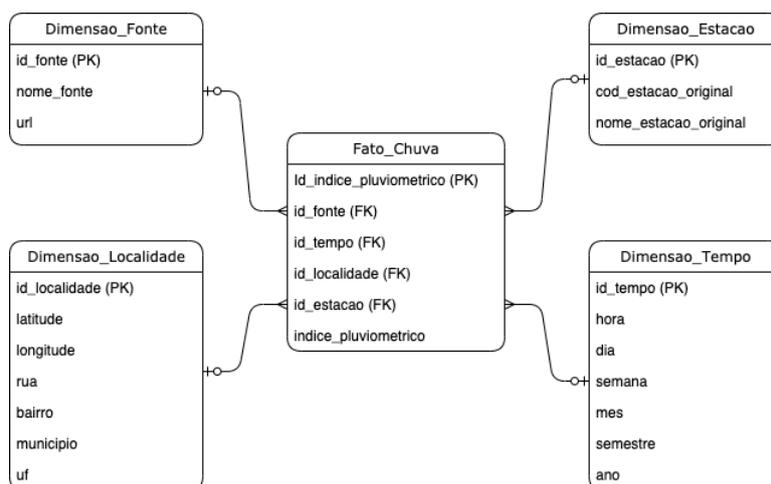


Figura 2. *Schema do DW-TEMPO*

4. Estudo de Caso

Nesta seção, apresentamos um estudo de caso realizado por especialistas da área de Defesa Civil, co-autores deste artigo. Com esse estudo de caso, os especialistas mostram como a TEMPO pode os ajudar em suas tarefas diárias de monitoramento e análise de acumulados de chuvas.

4.1. Fontes de Dados Utilizadas

As fontes de dados utilizadas para popular o DW-TEMPO foram (i) o Sistema Alerta Rio da Prefeitura do Rio de Janeiro² e, (ii) o CEMADEN. O Alerta Rio é o sistema de alerta de chuvas intensas e de deslizamentos em encostas da cidade do Rio de Janeiro com uma rede de 33 estações pluviométricas. Para popular o DW-TEMPO, consideramos apenas os dados de 1997 a 2019. Todos os arquivos obtidos pelo *crawler* se encontram em formato de texto plano (.txt). Cada arquivo está associado a um ano específico e possui 8 atributos. O CEMADEN realiza o monitoramento de 957 municípios classificados como vulneráveis a desastres naturais, 24 horas por dia. O CEMADEN fornece os dados pluviométricos de 77 municípios do estado do Rio de Janeiro. Para essa avaliação, consideramos os dados de 2014 a 2019. Todos os arquivos são do tipo CSV e se encontram separados por município, mês e ano. Cada arquivo possui 8 atributos.

4.2. Avaliação

De forma a avaliar a TEMPO, foram realizados dois tipos de análises: (i) análise espaço-temporal dos índices pluviométricos acumulados no município de Niterói (em múltiplas janelas temporais - 15 min, 30 min, 1 hora e 24 horas) e (ii) comparação dos índices pluviométricos em uma mesma região obtidos a partir de estações de múltiplas fontes. Em ambos os casos, uma série de consultas foi submetida ao DW-TEMPO. Para a primeira análise, consideramos o dia 07/06/2016³ e tomamos como horário-base as 10:00. A partir do horário-base, calculamos o índice acumulado nos últimos 15 min, 30 min, 1 hora e 24 horas. Como exemplo, a consulta submetida para o cálculo do índice acumulado na última hora é apresentada na Figura 3. Na Figura 3 apresentamos duas variações da consulta em SQL: **c1** que realiza a seleção dos dados de acordo com o nome da estação e **c2** que realiza a seleção dos dados pelo nome do município. Apesar de ambas as consultas retornarem os mesmos dados, elas apresentam desempenhos diferentes conforme apresentado na Figura 6, que apresenta os *box-plots* com os tempos de execução em (milissegundos) das consultas. É importante ressaltar que o tempo de execução médio (20 execuções) das consultas **c1** e **c2** foi de 0,25 segundos e 0,24 segundos, respectivamente. Segundo [Liu and Heer 2014], tempos de execução de consultas altos podem inviabilizar a construção de sistemas de exploração e análise visual de dados. De fato, os autores mostram que consultas com tempo de resposta de 0,5 segundos já podem diminuir a capacidade do usuário de construir e validar hipóteses sobre os dados que estão sendo visualizados. Sendo assim, os tempos de execução obtidos pelo DW-TEMPO permitem que o mesmo seja usado de forma efetiva como infraestrutura de gerenciamento de dados de sistemas de visualização. A Figura 5 apresenta um exemplo de visualização de índices pluviométricos, disponível no Portal TEMPO. Na Figura 5 podemos observar os acumulados de chuva dos últimos 15 min (janela **a**), 30 min (janela **b**), 1 hora (janela **c**) e 24 horas (janela **d**). Apesar do DW-TEMPO fornecer somente índices pluviométricos em pontos determinados do espaço (onde existem estações), utilizando técnicas de interpolação baseadas em distância [Lu and Wong 2008], podemos construir uma grade com valores acumulados de chuva cobrindo toda a região geográfica de interesse. Utilizando estes valores interpolados, podemos visualizar índices pluviométricos através de mapas de calor, como os exibidos na Figura 5.

²<http://alertario.rio.rj.gov.br/download/dados-pluviometricos/>

³Nessa data, chuvas fortes atingiram as cidades do Rio de Janeiro e Niterói - shorturl.at/tGTU9

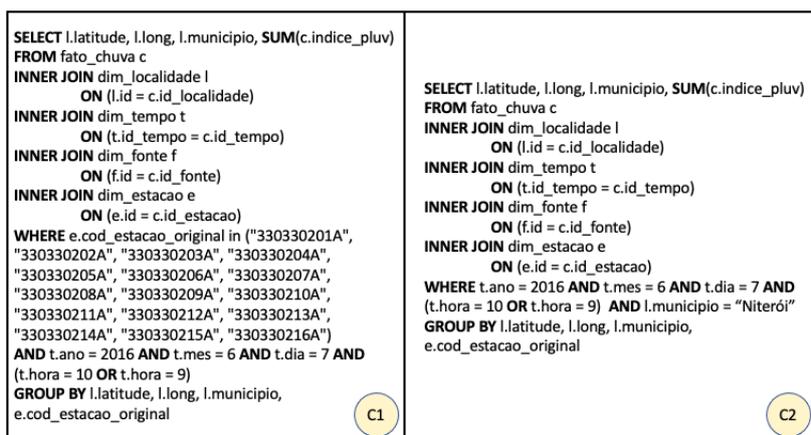


Figura 3. Estrutura das Consultas C1 e C2

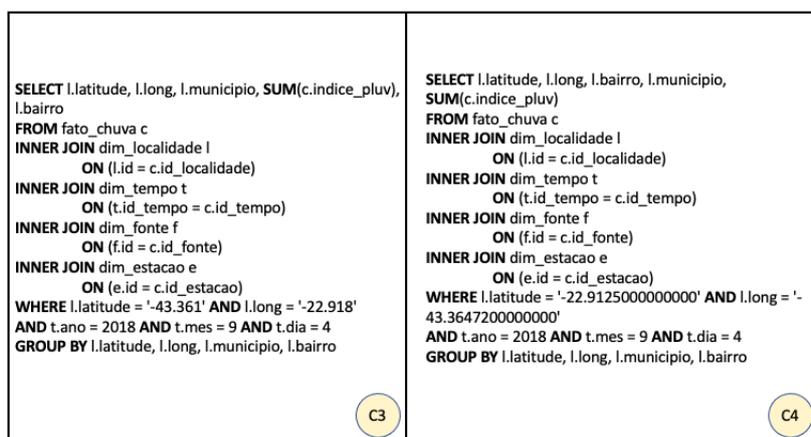


Figura 4. Estrutura das Consultas C3 e C4



Figura 5. Análise dos índices pluviométricos na região de Niterói em múltiplas janelas de tempo: (a) 15 minutos, (b) 30 minutos, (c) 1 hora, (d) 24 horas.

Para a segunda análise, consideramos o dia 04/09/2018 e o índice acumulado du-

rante todo o dia. Como o objetivo dessa análise é comparar os índices obtidos em uma região a partir de dados de múltiplas fontes, consideramos a região do bairro do Tanque (Rio de Janeiro) e duas estações distantes 1 Km uma da outra: (E1) -22.912,-43.3647 do CEMADEN e (E2) -22.918,-44.361 do Alerta Rio. Na Figura 4 apresentamos as consultas: C3 que realiza a seleção dos dados da estação E1 e C4 que realiza a seleção dos dados da estação E2. O tempo de execução das consultas C3 e C4 também é apresentado na Figura 6. Assim como nos casos das consultas C1 e C2, o tempo de execução médio (20 execuções) das consultas C3 e C4 foi de 0,27 segundos e 0,28 segundos, respectivamente, o que é aceitável. Além disso, pudemos observar uma diferença de 1,9 mm entre as leituras das estações. Esse tipo de diferença é esperada de acordo com trabalhos anteriores [Milanesi et al. 2017], principalmente considerando a distância entre as estações e diferenças na leitura entre os equipamentos. As avaliações anteriormente apresentadas demonstram o potencial da TEMPO no apoio às tarefas diárias dos especialistas.

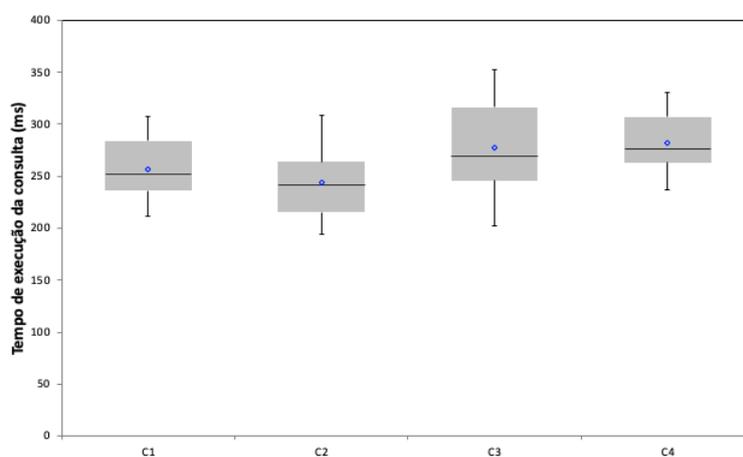


Figura 6. Tempos de Execução (ms) das Consultas C1, C2, C3 e C4

5. Conclusão

Estudos climáticos têm ganhado relevância nos últimos anos. Tais estudos já se encontram na era do *big data*, em que especialistas têm fácil acesso a um grande volume de dados. Tais dados podem ser oriundos de sensores ou simulações, e na grande maioria das vezes sua integração não é trivial. Um dos tipos de dados relacionados com análises climáticas são os dados pluviométricos. Esses dados são capturados em estações pluviométricas e comumente se encontram em diferentes resoluções e granularidades. Este artigo apresenta a ferramenta TEMPO para análise de dados pluviométricos multi-fonte. A TEMPO integra os dados obtidos em múltiplas fontes com múltiplas resoluções em um DW, que é capaz de representar dados pré-agregados de forma a otimizar consultas analíticas e a posterior visualização dos dados. O DW chamado DW-TEMPO segue o esquema estrela e foi implementado em um sistema de gerência de banco de dados relacional (PostgreSQL). De forma a avaliar a TEMPO, executamos um estudo de caso que integrou dados das bases do CEMADEN e do Alerta Rio, entre os anos de 1997 e 2019. As consultas executadas e as visualizações geradas mostram o potencial da TEMPO. Nosso plano de trabalho futuro inclui adicionar mais fontes de dados externas à TEMPO, como conjuntos de dados hidrometeorológicos oferecidos pela NASA e dados provenientes de simulações. Além disso,

pretendemos desenvolver um mecanismo para detectar falhas nos dados de origem, como por exemplo identificar possíveis *outliers*, a fim de garantir a qualidade dos dados.

Referências

- CEMADEN (2021). Centro nacional de monitoramento e alertas de desastres naturais - <http://www.cemaden.gov.br/apresentacao/>.
- 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.
- Cuzzocrea, A. (2015). Data warehousing and OLAP over big data: a survey of the state-of-the-art, open problems and future challenges. *Int. J. Bus. Process. Integr. Manag.*, 7(4):372–377.
- 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).
- Han, J., Kamber, M., and Pei, J. (2012). *Data mining concepts and techniques*. Elsevier.
- Inmon, W., Welch, J., and Glassey, K. (2005). *Building the Data Warehouse*. Sons Inc, New York.
- Kimball, R. and Ross, M. (2002). *The Data Warehouse Toolkit: The complete guide to dimensional modeling*. Wiley, New York.
- Liu, Z. and Heer, J. (2014). The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics*, 20(12):2122–2131.
- Lu, G. Y. and Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, 34(9):1044–1055.
- Milanesi, M., Rozolen, R., and Galvani, E. (2017). Comparativo entre instrumentos pluviométricos experimentais e automáticos. In *XVII Simpósio Brasileiro de Geografia Física Aplicada*, pages 2251–2261.
- 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 Eletrônicos*, 17.
- 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.