



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma Arquitetura para Dados Geoespaciais Massivos de Internet das Coisas no Contexto de Objetos com Diferentes Níveis de Mobilidade

Dennis Sávio Martins da Silva

Tese apresentada como requisito parcial para
conclusão do Doutorado em Informática

Orientadora
Prof.a Dr.a Maristela Terto de Holanda

Brasília
2025

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SS586a Silva, Dennis Sávio Martins da
Uma Arquitetura para Dados Geoespaciais Massivos de
Internet das Coisas no Contexto de Objetos com Diferentes
Níveis de Mobilidade / Dennis Sávio Martins da Silva;
orientador Maristela Terto de Holanda. Brasília, 2025.
124 p.

Tese(Doutorado em Informática) Universidade de Brasília,
2025.

1. Big Data. 2. Dados Geoespaciais. 3. Internet das
Coisas. I. Holanda, Maristela Terto de, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma Arquitetura para Dados Geoespaciais Massivos de Internet das Coisas no Contexto de Objetos com Diferentes Níveis de Mobilidade

Dennis Sávio Martins da Silva

Tese apresentada como requisito parcial para
conclusão do Doutorado em Informática

Prof.a Dr.a Maristela Terto de Holanda (Orientadora)
CIC/UnB

Prof. Dr. Angelo Roncalli Alencar Brayner Prof. Dr. Daniel Cardoso Moraes de Oliveira
Universidade Federal do Ceará (UFC) Universidade Federal Fluminense (UFF)

Prof.a Dr.a Célia Ghedini Ralha
Universidade de Brasília (UnB)

Prof.a Dr.a Claudia Nalon
Coordenadora do Programa de Pós-graduação em Informática

Brasília, 07 de outubro de 2025

Dedicatória

A meus pais, Rosineide e Domingos, por tudo. E à memória de meus avós, Julieta Anísia, Sebastião Leonel, Rita Francisca e José Martins, e da minha madrinha, Maria das Graças.

Agradecimentos

A realização desta tese foi possível graças ao apoio e incentivo de diversas pessoas e instituições, às quais expresso meus mais profundos agradecimentos.

Agradeço primeiramente à minha família, que me ofereceu toda a base e sempre me incentivou, acolheu e acreditou no sucesso deste projeto. Aos meus pais, Rosineide Martins Santana Silva e Domingos Sávio da Silva, pelo apoio incondicional, e às minhas irmãs, Deny Sávia e Sarah Martins. Estendo meus agradecimentos a todos os meus tios, tias, primos e primas, que torceram por mim ao longo desta caminhada.

Expresso minha profunda gratidão à minha orientadora, Maristela Terto de Holanda, pela oportunidade e pela constante prestatividade desde o primeiro contato, assim como pelos ensinamentos, pela gentileza, paciência, confiança e pela liberdade acadêmica que me foram concedidos ao longo de todo o trabalho.

Agradeço aos membros das bancas avaliadoras do exame de qualificação e da defesa da tese, os professores Ângelo Brayner, Daniel Oliveira e Célia Ghedini, pelas valiosas contribuições e pelos direcionamentos que foram fundamentais para o aprimoramento desta pesquisa.

Registro meus agradecimentos a todo o dedicado corpo docente do PPGI, em especial aos professores Li Weigang, Alba Melo, Maria Emília, e também ao professor Waldeyr Mendes, do IFG, pelo conhecimento e pelas experiências compartilhadas ao longo da minha formação.

Minha gratidão aos colegas do curso, em especial aos colegas de orientação Harley Vera, Aline Valente e Gustavo Van Erven, bem como aos amigos Teo Augusto, Henrique Jung, José Siqueira e Jefferson Viana, pela convivência e apoio ao longo desta jornada.

Agradeço aos colegas da UFPI – CSHNB, em especial aos coordenadores do curso de Sistemas de Informação, pelo apoio nos processos relacionados ao afastamento; aos colegas professores, especialmente à chefe Patrícia Medyna, que ingressou no PPGI na mesma época que eu; aos amigos Jonnison, Nonato e João Ramyllo, do CTI-CSHNB, pelo suporte técnico na preparação e execução dos experimentos; ao time de vigilância do campus, pelo apoio durante as longas horas extras que passei no laboratório; e aos discentes, pela compreensão enquanto precisei equilibrar as atividades docentes e de pesquisa.

Agradeço a todos que, durante a pandemia de COVID-19, fizeram o possível para manter o distanciamento social. Embora não tenham cumprido mais do que sua obrigação, esse esforço fez diferença neste período tão desafiador.

Por fim, fica meu agradecimento a todos os grandes amigos que tem me acompanhado e apoiado ao longo da vida, especialmente os integrantes do Time Dennis Sávio: Renan Vieira e Kennedy Rodrigues. Sem dúvidas, eu não teria ido tão longe sem o impulso de cada um de vocês!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

A Internet das Coisas (IoT, *Internet of Things*) tem sido crescentemente empregada em diferentes domínios de aplicação, e seus sistemas podem gerenciar objetos com variados níveis de mobilidade. Esse cenário resulta em um crescimento contínuo do volume de dados com natureza geoespacial. Em algumas aplicações, os dados gerados são principalmente leituras realizadas por dispositivos estacionários; em outras, podem também ser geradas localizações de dispositivos móveis. Considerar o nível de mobilidade dos dispositivos IoT possibilita a escolha de ferramentas com funcionalidades e desempenho mais adequados às necessidades do ambiente IoT em desenvolvimento. Entretanto, a literatura ainda é limitada no que diz respeito à relação entre mobilidade dos dispositivos e desempenho dos mecanismos de armazenamento e consulta de dados geoespaciais, distinguindo a manipulação de informações provenientes de dispositivos estacionários daquela originada de dispositivos móveis. Esta tese aborda *big data* geoespacial de dispositivos IoT com níveis heterogêneos de mobilidade, com as seguintes contribuições: i) um mapeamento sistemático de literatura referente à aplicação de *big data* geoespacial de IoT, que selecionou 196 trabalhos a partir de quatro bases acadêmicas (Scopus, IEEE Xplore, ACM, e Web of Science); ii) um levantamento dos recursos geoespaciais oferecidos pelas ferramentas mais presentes na literatura relacionada ao gerenciamento e análise de *big data* geoespacial de IoT; iii) a proposição e implementação de uma arquitetura em duas camadas – Armazenamento e Análise de Dados, e Aplicação – para o gerenciamento e análise de *big data* geoespacial oriundo de dispositivos com diferentes níveis de mobilidade; e iv) realização de dois experimentos para avaliar o desempenho dos ambientes computacionais propostos para a camada de Armazenamento e Análise de Dados — o primeiro voltado ao gerenciamento de grandes volumes de dados e o segundo à aplicabilidade de mecanismos de indexação. Foram utilizados *datasets* que simulam diferentes níveis de mobilidade, avaliando-se o desempenho geral dos ambientes, a variabilidade dos resultados, o consumo de armazenamento, a relação entre desempenho e mobilidade e os benefícios do uso de indexação.

Palavras-chave: Big Data, Dados Geoespaciais, Internet das Coisas

Abstract

The Internet of Things (IoT) has been increasingly employed in different application domains, and its systems can manage objects with varying levels of mobility. This scenario results in a continuous growth in the volume of geospatial data. In some applications, the data generated are mainly readings collected by stationary devices; in others, they may also include the locations of mobile devices. Considering the level of mobility of IoT devices enables the selection of tools with functionalities and performance more suited to the needs of the IoT environment under development. However, the literature remains limited regarding the relationship between device mobility and the performance of mechanisms for storing and querying geospatial data, distinguishing the handling of information originating from stationary devices from that produced by mobile ones. This thesis addresses geospatial big data from IoT devices with heterogeneous levels of mobility, with the following contributions: i) a systematic mapping of the literature on geospatial big data in IoT, in which 196 works were selected from four academic databases (Scopus, IEEE Xplore, ACM, and Web of Science); ii) a survey of the geospatial capabilities provided by the tools most frequently cited in the literature related to the management and analysis of geospatial big data from IoT; iii) the design and implementation of a two-layer architecture — Data Storage and Analysis, and Application — for the management and analysis of geospatial big data originating from devices with different levels of mobility; and iv) the execution of two experiments to evaluate the performance of the proposed computational environments for the Data Storage and Analysis layer — the first focusing on large-scale data management and the second on the applicability of indexing mechanisms. Simulated datasets representing different mobility levels were used to assess overall system performance, result variability, storage consumption, the relationship between performance and mobility, and the benefits of index utilization.

Keywords: Big Data, Geospatial Data, Internet of Things

Sumário

1	Introdução	1
1.1	Problema	2
1.2	Objetivos	4
1.3	Metodologia	4
1.4	Estrutura do Documento	5
2	Referencial Teórico	7
2.1	Dados Geoespaciais	7
2.2	<i>Big Data</i> Geoespacial	8
2.3	Internet das Coisas	11
2.4	Considerações Finais	13
3	Mapeamento da Literatura	14
3.1	Mapeamento Sistemático de Literatura	14
3.2	Etapa 1: Busca	15
3.3	Etapa 2: Seleção e Leitura	16
3.4	Resultados	18
3.5	Considerações Finais	29
4	Levantamento de Ferramentas	30
4.1	Contextualização	30
4.2	Armazenamento de <i>Big Data</i> Geoespacial	33
4.3	Processamento de <i>Big Data</i> Geoespacial	33
4.4	Considerações Finais	37
5	Proposta	39
5.1	Arquitetura Proposta	39
5.2	Ferramentas Utilizadas	41
5.2.1	Postgres-XL e PostGIS	41
5.2.2	Apache Spark	42

5.2.3	Apache Sedona	43
5.2.4	<i>Hadoop Distributed File System</i> (HDFS)	43
5.3	Implementação	44
5.3.1	Armazenamento e Análise de Dados	44
5.3.2	Camada de Aplicação	50
5.4	Considerações Finais	51
6	Experimentos	53
6.1	Contextualização	53
6.2	Preparação da Infraestrutura Computacional	54
6.3	Geração dos Dados	55
6.4	Elaboração das Consultas	57
6.5	Avaliação dos Ambientes	59
6.5.1	Experimento 1	61
6.5.2	Experimento 2	64
6.5.3	Análise dos Resultados	73
6.6	Limitações	78
6.7	Ameaças à Validade	79
6.8	Resultados Acadêmicos	80
6.9	Considerações Finais	80
7	Conclusão	82
7.1	Trabalhos Futuros	83
	Referências	84

Lista de Figuras

1.1	Exemplos de contextos de mobilidade.	3
1.2	Exemplo de contexto com dispositivos apresentando diferentes padrões de dados de localização e observação.	3
3.1	Metodologia do mapeamento de literatura.	16
3.2	Coocorrência de palavras-chave. Fonte: [55]	18
3.3	Domínios de aplicação.	19
3.4	Origem dos <i>datasets</i>	21
4.1	Plataformas para <i>big data</i> geoespacial.	32
5.1	Arquitetura proposta.	40
5.2	Instanciação da arquitetura proposta.	45
5.3	Modelo conceitual de dados (DER) adotado na arquitetura proposta. . . .	47
5.4	Ambiente 1 – Postgres-XL e PostGIS.	49
5.5	Ambiente 2 – Apache Sedona e HDFS.	50
5.6	Visualização dos resultados de operação espacial.	51
6.1	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 1.	65
6.2	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 1 (continuação).	66
6.3	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2.	69
6.4	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).	70
6.5	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).	71
6.6	Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).	72

Lista de Tabelas

3.1	Resultados da seleção.	17
3.2	Classificação dos estudos.	23
3.3	Análise dos trabalhos relacionados.	28
4.1	SGBDs com suporte a <i>big data</i> e seus recursos geoespaciais.	34
4.2	SGBDs populares com suporte a <i>big data</i> , suas geometrias e operações geoespaciais.	35
4.3	Ferramentas para análise de <i>big data</i> geoespacial.	36
5.1	Dicionário de dados da entidade Dispositivo (<i>Entity</i>).	46
5.2	Dicionário de dados da entidade Observação (<i>Observation</i>).	47
5.3	Dicionário de dados da entidade Localização (<i>Location</i>).	47
5.4	Dicionário de dados da entidade PropriedadeObservada (<i>ObservedProperty</i>).	47
5.5	Modelo de dados desnormalizado da arquitetura.	48
6.1	Sensores do projeto <i>Array of Things</i> utilizados na simulação.	56
6.2	Formato final dos arquivos .csv.	57
6.3	Consultas espaciais.	58
6.4	Consultas de observações.	59
6.5	Consultas mistas.	60
6.6	Características dos <i>datasets</i> simulados - Experimento 1.	61
6.7	Experimento 1: Resultados (tempo em milissegundos) para o <i>dataset</i> de mobilidade alta.	62
6.8	Experimento 1: Resultados (tempo em milissegundos) para o <i>dataset</i> de mobilidade intermediária.	62
6.9	Experimento 1: Resultados (tempo em milissegundos) para o <i>dataset</i> de mobilidade baixa.	63
6.10	Características dos <i>datasets</i> simulados - Experimento 2.	67
6.11	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de alta mobilidade (Postgres-XL).	67

6.12	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de alta mobilidade (Apache Sedona).	68
6.13	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de mobilidade média (Postgres-XL).	68
6.14	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de mobilidade intermediária (Apache Sedona).	73
6.15	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de baixa mobilidade (Postgres-XL).	73
6.16	Experimento 2: Resultados (tempo em milissegundos) para o <i>dataset</i> de baixa mobilidade (Apache Sedona).	74

Lista de Abreviaturas e Siglas

AoT *Array of Things.*

FCD *Floating Car Data.*

GIS *Sistemas de Informação Geográfica.*

GML *Geography Markup Language.*

GPS *Sistemas de Posicionamento Global.*

HDFS *Hadoop Distributed File System.*

IA *Inteligência Artificial.*

IDW *Inverse Distance Weighting.*

IoT *Internet das Coisas.*

JSON *JavaScript Object Notation.*

KML *Keyhole Markup Language.*

LIDAR *Laser Imaging Detection and Ranging.*

ML *Aprendizagem de Máquina.*

MMDB *Main Memory Database.*

OBD *On-Board Diagnosis.*

OGC *Open Geospatial Consortium.*

RFID *Radio Frequency Identification.*

SGBDs *Sistemas Gerenciadores de Banco de Dados.*

SOS *Sensor Observation Service.*

SWE *Sensor Web Enablement.*

VGI *Volunteered Geographic Information.*

VR *Realidade Virtual.*

XFCD *Extended Floating Car Data.*

Capítulo 1

Introdução

O volume de dados geoespaciais tem crescido ao longo dos anos, em decorrência da popularização das tecnologias de informação e comunicação com geolocalização, incluindo aplicações para dispositivos móveis e mídias sociais [1, 2, 3]. Entre os paradigmas de comunicação, a Internet das Coisas (IoT, *Internet of Things*) tem se destacado como uma das principais fontes de dados geoespaciais [4]. Nesse contexto, dispositivos com presença pervasiva desempenham diversas funções no ambiente em que estão inseridos e apresentam a capacidade de se conectar e trocar informações, tanto sobre si próprios quanto sobre o meio ao seu redor, por meio da internet [5, 6].

Parte significativa dos dados gerados por dispositivos de IoT possui aspectos geoespaciais como localização, dimensão e orientação (direção). A análise desses dados permite ressignificar a relação existente entre os dispositivos e o ambiente em que estão inseridos [7, 8]. Exemplos de domínios de aplicação incluem cidades inteligentes, planejamento urbano, monitoramento ambiental, automação industrial e residencial, *e-health* e *e-commerce* [8, 5, 9, 10].

A manipulação e análise desses dados geoespaciais de IoT é desafiadora devido ao seu grande volume, à alta velocidade de geração e à heterogeneidade em tipos, fontes, formatos e padrões geoespaciais [8]. Esses desafios são potencializados pelo aumento na velocidade da conexão entre dispositivos, sensores e sistemas, que amplia o fluxo de dados e a necessidade de processamento em tempo quase real. Tais características aproximam esse contexto do fenômeno de *big data*, o que reforça a necessidade de mecanismos robustos para gerenciar fluxos expressivos de dados de forma eficiente [6, 8, 11, 12, 13].

Big data é o termo utilizado para designar grandes volumes de dados, estruturados ou não, gerados em alta velocidade e variedade, e que exigem tecnologias avançadas para seu armazenamento, manipulação e análise [5, 14, 15]. Algumas dessas tecnologias são bancos de dados NoSQL, sistemas de arquivos distribuídos e *frameworks* de processamento de dados em larga escala. Quando tais dados apresentam características geoespaciais,

surtem demandas adicionais, como operações espaciais, mecanismos especializados de indexação e o emprego de tipos de dados adequados.

1.1 Problema

Aplicações de IoT podem conter registros de dispositivos diversos, com diferentes níveis de mobilidade [16]. Nesta tese, o nível de mobilidade é definido como a frequência com que os dispositivos geram novos registros de localização ao longo do tempo – ou seja, quanto mais registros de alterações na localização geográfica, maior a mobilidade [17]. Alguns dispositivos podem ser totalmente estacionários ou apresentar poucos registros de mudança de localização. Outros podem apresentar uma grande quantidade de registros devido à movimentação frequente.

Como exemplos, podem ser considerados diversos cenários de mobilidade para uma mesma aplicação composta por dispositivos heterogêneos. Em um cenário (Figura 1.1a), esta aplicação poderia gerenciar dispositivos móveis, como veículos – contendo sensores diversos e registrando localizações – e *smartphones* – executando aplicações com registro de localização dos usuários. Em outro cenário (Figura 1.1b), a aplicação poderia ser utilizada somente no gerenciamento de dispositivos estacionários ou cuja localização dificilmente é alterada ao longo da utilização da aplicação, como sensores de estacionamento – contendo o registro de utilização de uma determinada vaga –, sensores de temperatura e umidade, ou sensores de tráfego. Em um terceiro cenário (Figura 1.1c), poderiam ser realizadas operações envolvendo dispositivos apresentando diferentes níveis de mobilidade.

Adicionalmente, os dispositivos não produzem apenas dados geoespaciais, mas também grandes volumes de informações referentes a observações em fenômenos do ambiente. A relação entre os dados de localização e os dados de observação varia em função de fatores como a frequência de mobilidade do dispositivo, a quantidade de sensores que o compõem e a regularidade de coleta de cada tipo de dado. A Figura 1.2 ilustra um cenário composto por dispositivos heterogêneos, cujos níveis de mobilidade variam desde o estado estacionário — exemplificado por dispositivos de *smart home*, que tendem a gerar elevado volume de dados de observação e reduzido volume de dados de localização — até dispositivos móveis, como veículos, que produzem registros com maior variação espacial.

Observa-se que ainda há pouca literatura que relacione o nível de mobilidade dos dispositivos ao desempenho dos mecanismos de armazenamento e consulta a dados geoespaciais, diferenciando a manipulação de dados provenientes de dispositivos estacionários daquela de dispositivos móveis.

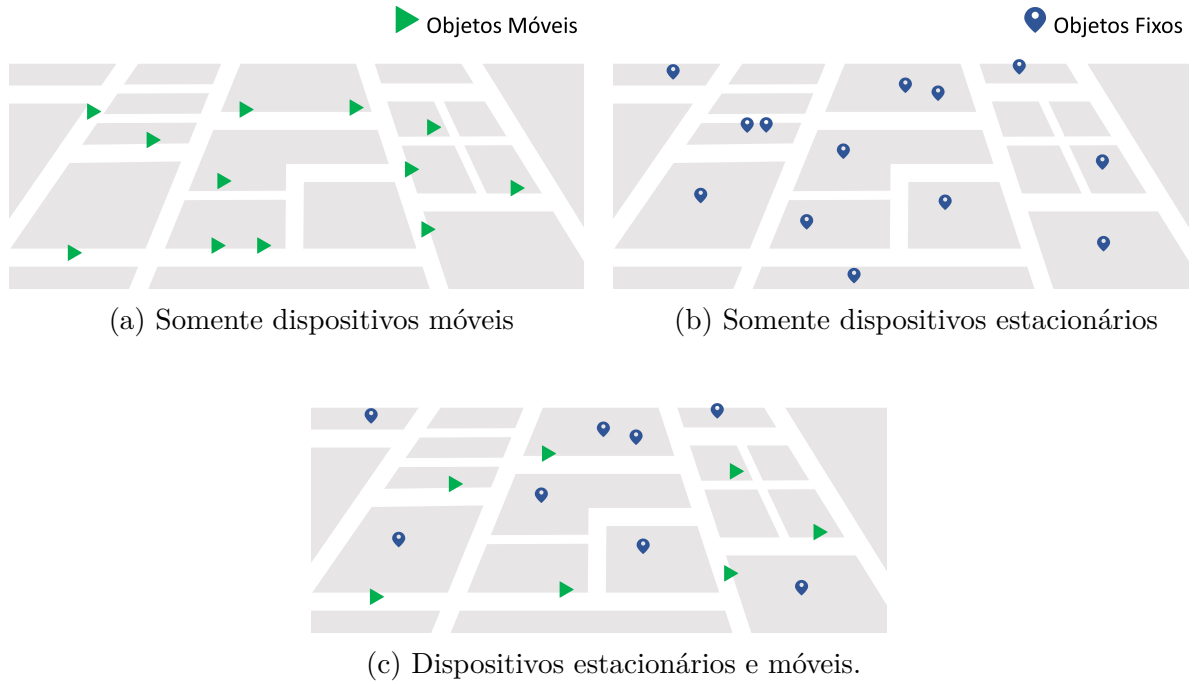


Figura 1.1: Exemplos de contextos de mobilidade.

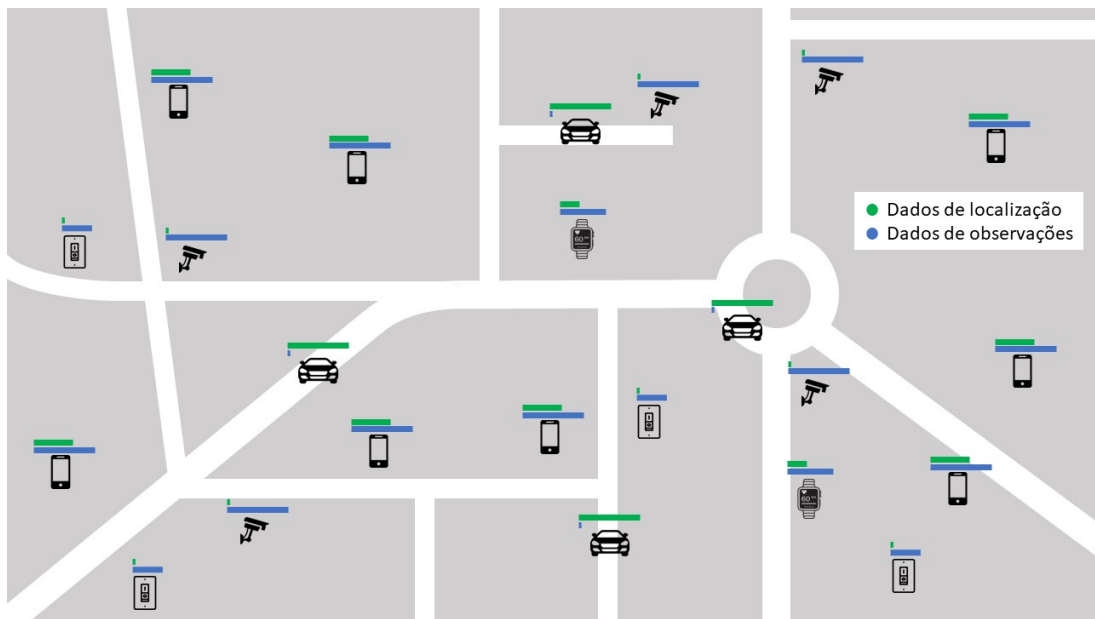


Figura 1.2: Exemplo de contexto com dispositivos apresentando diferentes padrões de dados de localização e observação.

Esta tese busca então responder a seguinte questão de pesquisa: *Como implementar uma arquitetura para armazenamento e análise de dados geoespaciais massivos de IoT com suporte a diferentes níveis de mobilidade dos dispositivos?*

1.2 Objetivos

Esta tese tem como objetivo propor e validar uma arquitetura para *big data* geoespacial oriundo de dispositivos IoT com diferentes níveis de mobilidade. Parte-se do pressuposto de que uma aplicação para gerenciamento e análise de *big data* geoespacial de IoT que considere a mobilidade dos dispositivos possui o potencial de viabilizar operações espaciais com melhor desempenho.

Diante desse contexto, para o alcance do objetivo geral foram estabelecidos os seguintes objetivos específicos:

- Analisar os principais estudos sobre a aplicação de *big data* geoespacial em IoT e identificar as soluções de software mais recorrentes na literatura para armazenamento e análise desses dados;
- Implementar uma arquitetura para armazenar e visualizar dados geoespaciais oriundos de dispositivos IoT com diferentes níveis de mobilidade; e
- Realizar experimentos para avaliar o desempenho a arquitetura proposta, empregando grandes volumes de dados de dispositivos IoT simulados em cenários caracterizados por diferentes padrões de mobilidade.

1.3 Metodologia

Esta pesquisa caracteriza-se como de natureza aplicada, por propor uma solução para o armazenamento e recuperação de *big data* geoespacial em ambientes de Internet das Coisas (IoT) que envolvem dispositivos com diferentes níveis de mobilidade. Quanto à abordagem, adota-se uma perspectiva quantitativa, uma vez que a avaliação da arquitetura proposta baseia-se em métricas formais de desempenho obtidas por meio de experimentos controlados. Sob o enfoque metodológico, o estudo classifica-se como experimental, pois compreende a proposição, implementação e análise de um artefato computacional desenvolvido e avaliado segundo os princípios da pesquisa em Engenharia e Ciência da Computação [18, 19, 20]. A metodologia adotada foi composta pelas seguintes fases:

- i) Com o objetivo de investigar a presença e as práticas de aplicação de *big data* geoespacial de IoT, foi realizado um mapeamento sistemático da literatura, visando identificar tendências de pesquisa e classificar os estudos de acordo com suas principais contribuições. A busca foi conduzida em quatro bases acadêmicas (Scopus, IEEE Xplore, ACM e Web of Science), e o processo de seleção dos trabalhos ocorreu em três etapas sequenciais. O mapeamento contemplou, entre outros aspectos, os

principais domínios de aplicação, as fontes e formatos de dados utilizados, as ferramentas de software adotadas e as abordagens mais recorrentes para armazenamento e análise dos dados geoespaciais de IoT.

- ii) Com base no mapeamento da literatura, realizou-se uma análise dos recursos oferecidos pelas principais ferramentas de armazenamento e análise de *big data* geoespacial aplicadas a IoT. No armazenamento, foram considerados diferentes modelos de bancos de dados, relacionais e não relacionais, além de sistemas de arquivos distribuídos. Na análise, foram examinados *frameworks* voltados a dados geoespaciais e os recursos disponíveis para geovisualização. Entre as funcionalidades identificadas nessas ferramentas destacam-se geometrias, operadores espaciais, índices, formatos de dados e linguagens de consulta.
- iii) Propôs-se e implementou-se o protótipo de uma arquitetura em duas camadas para o gerenciamento e análise de *big data* geoespacial proveniente de dispositivos com diferentes níveis de mobilidade. A primeira, denominada Camada de Armazenamento e Análise de Dados, é responsável por armazenar, gerenciar e dar suporte às análises, tendo sido implementada em dois ambientes distintos com diferentes ferramentas de armazenamento e análise geoespacial identificadas no mapeamento da literatura. A segunda, a Camada de Aplicação, disponibiliza ao usuário interfaces que permitem a realização de consultas aos dados dos dispositivos e a geovisualização dos resultados.
- iv) Por fim, para a validação da camada de Armazenamento e Análise de Dados da arquitetura proposta, foram conduzidos dois experimentos com o objetivo de avaliar o desempenho dos dois ambientes propostos. O primeiro experimento avaliou a performance das ferramentas na manipulação de grandes volumes de dados, enquanto o segundo investigou o impacto da utilização de índices na camada. Ambos os experimentos utilizaram *datasets* simulando três diferentes contextos de uso, cada um representando um nível de mobilidade dos dispositivos. A camada de aplicação não foi submetida a experimentos de validação, pois aspectos relacionados a usabilidade e interação com o usuário estão fora do escopo desta pesquisa.

1.4 Estrutura do Documento

O restante deste trabalho está estruturado da seguinte forma:

- O Capítulo 2 apresenta o referencial teórico sobre dados geoespaciais, incluindo: tipos de geometrias e operadores espaciais; tecnologias, operações e desafios associados ao processamento de *big data* geoespacial; e conceitos de IoT, destacando sua relação com a geração e análise de dados geoespaciais em ambientes de *big data*.

- O Capítulo 3 descreve o processo de condução do mapeamento sistemático da literatura sobre *big data* geoespacial em IoT, apresentando inicialmente a metodologia adotada para a realização do levantamento. Em seguida, expõe a estratégia de busca utilizada nas bases acadêmicas, detalha os procedimentos de seleção dos trabalhos e discute os resultados do mapeamento, enfatizando os aspectos mais relevantes para a proposta desta tese.
- O Capítulo 4 apresenta o levantamento das ferramentas para manipulação de dados geoespaciais, tanto sob uma perspectiva de armazenamento quanto de processamento de *big data* geoespacial.
- O Capítulo 5 detalha a arquitetura proposta, explicando as funções e integração entre suas camadas, descreve as ferramentas para armazenamento e análise de *big data* utilizadas, e detalha a implementação do protótipo empregado na experimentação.
- O Capítulo 6 apresenta a etapa de experimentação, contemplando a geração de *datasets* que simulam diferentes níveis de mobilidade, a preparação do ambiente utilizado, e a realização de experimentos para validação da camada de Armazenamento e Análise de Dados, concluindo com uma discussão acerca de suas limitações e das ameaças à validade.
- O Capítulo 7 contempla as conclusões do trabalho e sugestões para pesquisas futuras.

Capítulo 2

Referencial Teórico

Neste capítulo, apresenta-se uma visão geral dos principais conceitos que fundamentam o desenvolvimento desta tese. A organização deste capítulo segue a seguinte disposição: Na Seção 2.1, são apresentados os conceitos básicos relacionados a dados geoespaciais, bem como as principais geometrias e operadores espaciais; a Seção 2.2 introduz a definição de *big data* geoespacial, assim como algumas das tecnologias e operações ligadas ao conceito; a Seção 2.3 apresenta conceitos básicos relacionados a IoT e sua ligação com a pesquisa relacionada a *big data* geoespacial; e, por fim, a Seção 2.4 aborda as conclusões do capítulo.

2.1 Dados Geoespaciais

Dados geoespaciais consistem de dados georreferenciados – ou seja, representados em um sistema de coordenadas geográficas – que descrevem fenômenos dispostos sobre a superfície do planeta [2, 3]. Os principais tipos de dados geoespaciais são *raster* e vetor [11]. No tipo *raster*, o espaço geográfico é representado por uma grade composta por células (ou *pixels*), cada uma armazenando informações relativas a uma variável específica, como temperatura, umidade ou pressão atmosférica. Esse tipo é particularmente adequado para aplicações que envolvem a análise contínua de superfícies ou a sumarização de informações em nível celular, sendo amplamente utilizado em sensoriamento remoto e modelagem ambiental. Já no tipo vetorial, o espaço é representado por entidades geométricas discretas — pontos (**Point**), linhas (**Line**) e polígonos (**Polygon**) — que permitem a descrição detalhada de objetos e fenômenos geográficos, como localização de sensores, redes de transporte e delimitação de áreas. Esse tipo mostra-se especialmente adequado para o armazenamento e análise de trajetórias, rotas, fronteiras e outras estruturas espaciais que exigem precisão geométrica [21].

No contexto de sua representação e análise, os dados geoespaciais são manipulados por meio de geometrias e operadores espaciais. Entre as geometrias mais comuns, **Point**,

Line e **Polygon** constituem as formas elementares a partir das quais se definem versões compostas, como **MultiPoint**, **LineString**, **MultiLineString** e **MultiPolygon** [11].

Para analisar essas geometrias, diversos operadores espaciais são utilizados, permitindo avaliar relações topológicas, medir distâncias e identificar padrões de proximidade entre os objetos geográficos [22, 23, 24]:

- **Bounding Box** – retorna geometrias localizadas dentro de um retângulo;
- **Distance** – calcula a distância entre geometrias;
- **Intersects** – verifica se duas geometrias possuem pelo menos um ponto em comum;
- **Overlap** (sobreposição) – verifica quando duas geometrias intersectam, mas nenhuma está totalmente contida na outra;
- **Contains** – verifica quando uma geometria contém outra totalmente;
- **Disjoint** – negação lógica de **Intersects**, verifica se duas geometrias não possuem pontos em comum;
- **Within** – o contrário de **Contains**, verifica quando uma geometria está totalmente contida em outra);
- **WithinDistance** (**DWithin**) – encontra geometrias localizadas dentro de um determinado raio de distância de um ponto;
- **Area** – área de um polígono;
- **Touch** – verifica se duas geometrias intersectam, mas seus pontos em comum ficam em seus limites; e
- **Equals** – verifica se duas geometrias ocupam exatamente o mesmo espaço no mapa.

2.2 *Big Data* Geoespacial

Big data é o termo utilizado para designar grandes volumes de dados cujo armazenamento, recuperação e análise demandam tempo ou custo elevados para realização via ferramentas tradicionais de gerenciamento de dados, como sistemas centralizados baseados em bancos de dados relacionais [15].

A aplicação de *big data* envolve a utilização de bancos de dados relacionais e não relacionais – também conhecidos como NoSQL (*Not Only SQL*) –, plataformas de *cloud computing*, mineração de dados em ambientes de *grid*, sistemas de arquivos distribuídos, entre outras tecnologias [1, 4, 6, 25, 26, 27].

Tradicionalmente, a aquisição de dados geoespaciais estava majoritariamente restrita a governos e grandes corporações, sendo viabilizada por tecnologias de alto custo e de difícil acesso ao público em geral, tais como satélites, telescópios espaciais, técnicas de sensoramento remoto e observações ambientais [28]. Esses dados, de caráter predominantemente oficial e elevado grau de confiabilidade, eram compostos sobretudo por mapas, imagens aéreas e de satélite, bem como por registros provenientes de estudos que descreviam de forma estática a superfície terrestre [3].

Nas últimas décadas, em decorrência da ampla disseminação das tecnologias de informação e comunicação, observou-se um crescimento significativo no volume e na velocidade de geração de dados, incluindo aqueles de natureza geoespacial. Novas fontes de dados emergiram e tornaram-se comercialmente acessíveis ao público em geral, como sensores, dispositivos móveis com acesso à localização por Sistemas de Posicionamento Global (GPS, *Global Positioning Systems*) e aplicações que exploram esse recurso. Outras fontes relevantes incluem leitores de *Radio Frequency Identification* (RFID) e redes de sensores sem fio [29].

As operações geoespaciais, cujo custo computacional e complexidade já são intrinsecamente mais elevados que a dos operadores relacionais convencionais, têm seu impacto ampliado em cenários com grande volume de dados geoespaciais. As demandas de *data-intensive* GIS têm desafiado progressivamente as capacidades de sistemas *single-processor*, inserindo as aplicações de dados geoespaciais no contexto de *big data* [3].

O termo *big data* geoespacial descreve conjuntos de dados com atributos geoespaciais, que podem demandar elevada robustez computacional para um processamento eficiente devido ao seu grande volume e à complexidade das operações geoespaciais [25, 30]. O crescimento no volume, complexidade e velocidade de geração dos dados geoespaciais demanda procedimentos especiais de preparação e processamento, dificultando sua manipulação [2]. Por exemplo, o processamento de coordenadas envolve operações complexas, incluindo a conversão em geometria, a transformação entre padrões, formatos e sistemas de coordenadas, e a avaliação das relações espaciais entre objetos, tornando a manipulação e análise desses dados mais desafiadora [2, 31].

As primeiras definições de *big data* abrangiam três “V’s” - Volume, Variedade, e Velocidade - que representam suas características. Mais tarde, outros “V’s” – como por exemplo Veracidade, Valor e Visualização – foram adicionados à definição [4, 32]. Em se tratando de dados geoespaciais, estas características se apresentam da seguinte forma: [33, 34, 6, 8, 25, 35, 21, 36]:

- **Volume:** Refere-se à grande quantidade de dados que necessitam ser armazenados e processados. No caso dos dados geoespaciais, o volume cresce rapidamente em

função da diversidade de fontes, como RFID, mídias sociais com funcionalidades de geolocalização, redes de sensores, IoT e VGI (*Volunteered Geographic Information*).

- **Variedade:** Dados geoespaciais possuem tipos, estruturas, semânticas e formatos heterogêneos devido às diferenças entre as múltiplas fontes de dados, propósitos, tecnologias e abordagens de coleta.
- **Velocidade:** Se refere à elevada taxa em que os dados geoespaciais são criados e precisam ser ingeridos e processados para gerar respostas em tempo real. Exemplos incluem fluxos de dados de GPS de frotas de veículos, transmissões de coordenadas de dispositivos móveis, e leituras em tempo real de sensores de *IoT* espalhados em áreas urbanas.
- **Veracidade:** Indicativo da confiabilidade, qualidade e autenticidade dos dados, ligada a características como corretude, integridade, consistência e validade. Dados geoespaciais podem possuir alta imprecisão dependendo da aplicação e técnica de coleta. Outras fontes de incerteza são a heterogeneidade dos dispositivos de geração de dados, ruído, inconsistências e ambiguidades.
- **Valor:** O que os dados agregam ao domínio de aplicação, às vezes na forma de conhecimento relevante. A análise de dados espaciais permite a obtenção de *insights* que auxiliam o planejamento e tomada de decisões em diversos domínios. Esse processo não se limita a saber a localização de fenômenos espaciais, permitindo também compreender suas causas e condições de ocorrência, e inferir suas consequências e relacionamento com os dispositivos que compõem o sistema.
- **Geovisualização:** Se refere à representação dos dados, buscando simplificar a interpretação dos mesmos e a identificação de informação relevante. Essa característica advém da adequação dos dados geoespaciais a uma apresentação interativa em alto nível na forma de mapas interativos 2D, 3D e 4D, infográficos, globos virtuais e *dashboards*, dentre outras formas de representação que enriquecem a apresentação dos resultados analíticos e possibilitam novas formas de interpretação da informação geoespacial.

Os Sistemas Gerenciadores de Banco de Dados (SGBDs) relacionais dispõem de suporte consolidado para dados espaciais, exemplificado pelo PostgreSQL em conjunto com sua extensão PostGIS. A pesquisa geoespacial tem também direcionado esforços a soluções de computação distribuída [4, 37, 38]. A escalabilidade alcançada por meio da distribuição e replicação de dados, bem como de serviços, contribui para aumentar a confiabilidade e a elasticidade, disponibilizando recursos computacionais de armazenamento e processamento de *big data* geoespacial com melhor relação custo-eficácia [37]. Contudo,

o suporte de bancos de dados distribuídos a dados geoespaciais ainda apresenta diversas oportunidades e desafios. Por exemplo, a implementação de operações complexas, como junções e agregações, assim como a criação de índices, pode ser permeada de desafios referentes à distribuição dos dados em múltiplos servidores [12].

Além dos operadores apresentados na Seção 2.1, Eldawy [39] identifica cinco categorias de operações amplamente empregadas no processamento de dados geoespaciais em larga escala: operações básicas, junções, geometria computacional, mineração de dados e operações do tipo *raster*. Entre essas categorias, este trabalho concentrou-se nas operações básicas, que compreendem:

- ***Range queries***: a entrada é um conjunto de registros R e uma área A , geralmente retangular. A saída é cada registro $r \in R$ que sobrepõe a área A .
- ***Point queries***: consultas em que a região de interesse é reduzida a um único ponto, retornando todos os registros que coincidem exatamente com essa localização.
- ***Nearest Neighbor (NN) queries***: geralmente é uma consulta *top-k*, que usa distâncias para classificar os registros.

2.3 Internet das Coisas

Internet das coisas é um paradigma que envolve a comunicação, manipulação e percepção remota de “coisas”, dispositivos inteligentes e auto-configuráveis, interconectados direta ou indiretamente via internet, com presença pervasiva, que interagem com o ambiente em tempo real e são capazes de descrever essa interação [6, 10, 40, 41]. Coisas podem ser objetos físicos, como *smartphones*, eletrodomésticos, veículos e edificações, dotados de sensores que monitoram propriedades deles próprios – tais como velocidade ou níveis de carga – ou do ambiente onde estão – como temperatura, intensidade sonora e pressão do ar. Coisas podem também ser objetos virtuais, como aplicativos [9, 10]. Dentre as capacidades dos dispositivos IoT, podem ser citadas [42]:

- Comunicação e cooperação: habilidade de se conectar a recursos da internet, como dados e serviços, e atualizar seus estados;
- Endereçamento: Capacidade de ser encontrados remotamente e endereçados por meio de serviços de descoberta e busca;
- Sensoriamento (*Sensing*): capacidade de coletar informação do ambiente, que pode ser registrada, encaminhada, ou disparar uma ação/reação;
- Manipulação do ambiente: Além de sensores, os dispositivos IoT também podem ser dotados de atuadores, que são componentes digitais, elétricos ou mecânicos que

recebem sinais para desempenhar funções no ambiente, como mover objetos ou acionar mecanismos;

- Processamento incorporado de informações: processadores ou microcontroladores, que permitem aos dispositivos processar, interpretar e armazenar as informações dos sensores;
- Localização: ciência da localização espacial e capacidade de ser localizado; e
- Interfaces de usuário: capacidade de se comunicar com indivíduos, direta ou indiretamente (por exemplo, via *smartphone*).

Para fazer uso dessas capacidades, protocolos com diversos propósitos são organizados em arquiteturas IoT. Geralmente, essas arquiteturas possuem três camadas [10, 43]:

- Camada física, compreendendo o hardware – por exemplo, sensores e atuadores – instalado nos sistemas IoT;
- Camada de rede, que oferece a comunicação entre dispositivos IoT e um *gateway* para conectar esses dispositivos a serviços e aplicações na internet; e
- Camada de aplicação, composta do software que oferece aos usuários interfaces de acesso aos dados.

Os benefícios de se conectar dispositivos à internet têm se tornado cada vez mais evidentes, e IoT reúne tecnologias que aumentam a imersividade e ubiquidade da internet. Assim, a discussão não se limita à conectividade de dispositivos, mas também abrange o relacionamento entre mundo real e virtual [6]. O aumento do volume e da complexidade dos dados de IoT abre oportunidades significativas, desde que os obstáculos inerentes ao seu gerenciamento sejam adequadamente tratados [10, 44]. O ambiente IoT tem se tornado uma fonte representativa de dados, cujas características remetem aos desafios de *big data* [45]. Objetos IoT podem gerar rapidamente um grande *volume* de dados, que devem ser processados, analisados e utilizados para disparar diferentes operações. A *velocidade* de geração dos dados exige recursos de gerenciamento capazes de lidar com dados em diferentes frequências de coleta, desde leituras esporádicas de dispositivos com conectividade oportunística a objetos que enviam dados em tempo real. A alta *variedade* de dados é um fator relevante, devido à diversidade de padrões e representações adotados pelos fabricantes, e aos dispositivos com naturezas e propósitos distintos.

A interação entre IoT e *big data* impulsiona o avanço de ambas as áreas [32]. Ferramentas de armazenamento de *big data*, aliadas a tecnologias de processamento de *streams* em tempo real, ampliam as possibilidades de inovação em IoT, possibilitando soluções transformadoras em novos contextos e modelos de negócios [32, 46]. Dentre os requisitos

de bancos de dados para aplicações IoT podem ser citados: custo de armazenamento, velocidade das consultas, flexibilidade de mudança de esquema, escalabilidade, capacidade de manipular rapidamente grandes volumes de dados, e integração com ferramentas analíticas [45]. A natureza interconectada dos dados de IoT permite que eles sejam integrados a outras fontes de dados [47], oferecendo possibilidades relacionadas à extração de conhecimento. Para permitir que essa análise seja realizada com eficiência em domínios com fluxo crescente de dados, como *smart cities* e serviços de transporte, há uma crescente aplicação de computação em nuvem, oferecendo o uso sob demanda de recursos computacionais [48, 49]. Para complementar o potencial da nuvem e a robustez, escalabilidade e qualidade de serviço das aplicações IoT, os paradigmas de computação de borda e em névoa tem sido cada vez mais utilizados [50, 51]. Ge, Bangui e Buhnova [8] mencionam a grande diversidade de domínios de IoT, cada um aplicando as tecnologias de *big data* mais adequadas às suas próprias finalidades e tipos de dados. Essas especificidades dificultam o compartilhamento de conhecimento entre diferentes domínios e apontam para desafios que permanecem em aberto na comunidade científica.

Aplicações IoT também geram um grande volume de dados potencialmente espaço-temporais sobre o ambiente em que estão inseridos [10, 52, 53]. Processos de IoT envolvem dados de diversas fontes e formatos, e podem demandar mecanismos para consulta e análise dos dados geoespaciais [5], visando agregar valor, reconhecer padrões e relacionamentos entre os dados, e gerar informação [54]. Neste contexto, métodos analíticos espaciais são bastante utilizados, incrementando o valor da informação.

2.4 Considerações Finais

O presente capítulo abordou os conceitos fundamentais que sustentam a tese, incluindo os princípios básicos dos dados geoespaciais, suas principais geometrias e operadores espaciais. Também foram explorados os conceitos de *big data* geoespacial, com destaque para as tecnologias e operações associadas ao seu processamento. Por fim, foram apresentados os fundamentos de IoT, enfatizando sua relevância como fonte de dados e sua integração com aplicações de *big data* geoespacial, estabelecendo, assim, a base conceitual necessária para o desenvolvimento dos capítulos subsequentes.

Capítulo 3

Mapeamento da Literatura

Diversos trabalhos da literatura investigam aplicações de IoT que envolvem tanto dados geoespaciais oriundos de dispositivos estacionários quanto informações referentes às trajetórias de dispositivos móveis. Neste capítulo, é apresentado um levantamento dos principais estudos relacionados à aplicação de *big data* geoespacial em IoT, conduzido em quatro bases acadêmicas a partir de um mapeamento sistemático da literatura, originalmente publicado em [55] e atualizado em outubro de 2024. O capítulo encontra-se organizado conforme descrito a seguir: A Seção 3.1 descreve o processo de mapeamento aplicado; na sequência, a Seção 3.2 expõe a estratégia de busca aplicada às bases acadêmicas; a Seção 3.3 detalha os procedimentos de seleção dos trabalhos; a Seção 3.4 discute os resultados do mapeamento em aspectos pertinentes à proposta desta tese; e, por fim, a Seção 3.5 aborda as conclusões do capítulo.

3.1 Mapeamento Sistemático de Literatura

O método empregado para a análise da literatura foi o mapeamento sistemático de literatura (MSL), adaptado de Kitchenham e Charters [56] e Petersen *et al.* [57], que categoriza e sumariza de forma imparcial a informação sobre determinada área de pesquisa.

O objetivo principal do mapeamento foi investigar a presença e práticas de aplicação de *big data* geoespacial em IoT. Outros objetivos foram identificar tendências de pesquisa e classificar os estudos de acordo com suas principais contribuições. Para isso, as seguintes questões de pesquisa foram formuladas:

- *RQ1 - Em quais domínios big data geoespacial tem sido aplicado em IoT?* – O propósito desta questão é identificar os setores e áreas de aplicação de *big data* geoespacial no contexto de IoT.

- *RQ2 - Qual a fonte dos datasets empregados nos experimentos realizados nos estudos?* – A segunda questão busca determinar a origem (proprietária ou não) e a natureza (real ou simulada) dos dados utilizados nos experimentos para avaliar sua disponibilidade e representatividade.
- *RQ3 - Quais os principais formatos de dados armazenados?* – O propósito desta questão é identificar os formatos de dados e tipos de arquivos em que os dados geoespaciais são armazenados nas aplicações IoT.
- *RQ4 - Quais as ferramentas mais utilizadas para lidar com big data geoespacial em IoT?* – O intuito desta questão é levantar os softwares e *frameworks* empregados para armazenamento, processamento e análise de *big data* geoespacial em aplicações IoT.
- *RQ5 - Como os trabalhos empregando big data geoespacial em IoT podem ser classificados?* – A última questão teve por propósito classificar os trabalhos empregando *big data* com dimensão geoespacial em IoT.

A análise da literatura selecionou 196 estudos primários, e a Figura 3.1 ilustra o processo adotado. A primeira etapa consiste de uma busca abrangendo quatro bases acadêmicas. Os resultados foram unificados, e na segunda etapa foram submetidos a um processo de seleção dividido em três passos. No primeiro passo, os trabalhos foram selecionados com base na leitura dos resumos. No segundo passo, a seleção dos textos remanescentes foi realizada com base na introdução e conclusão dos trabalhos. Os textos restantes foram inteiramente lidos no terceiro passo e, no caso de responderem alguma pergunta de pesquisa, tiveram as informações relevantes registradas. Após a seleção foi realizado um processo de *snowballing*, no qual referências relevantes encontradas nos estudos selecionados foram lidas e tiveram sua informação extraída. O processo descrito pode ser replicado periodicamente com o objetivo de manter a atualização das referências, sendo que a última atualização ocorreu em outubro de 2024. Complementarmente, com a finalidade de incorporar referências mais recentes, realizou-se um *forward snowballing*, isto é, a identificação de estudos que citaram os trabalhos incluídos, abrangendo publicações até agosto de 2025.

3.2 Etapa 1: Busca

A busca por estudos mencionando *big data* contendo aspectos geoespaciais em aplicações IoT compreendeu estudos primários publicados em periódicos e conferências indexados

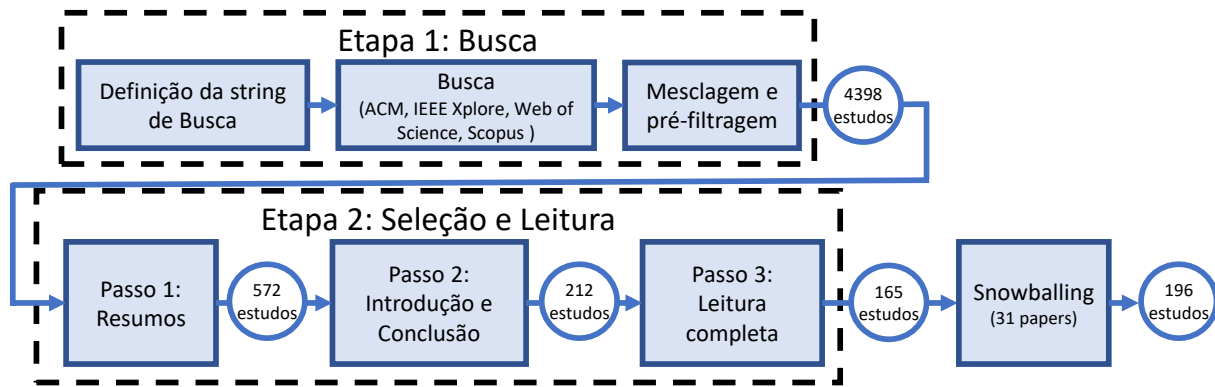


Figura 3.1: Metodologia do mapeamento de literatura.

nas bases acadêmicas Scopus¹, IEEE Xplore², ACM³, e Web of Science⁴. A *string* de busca utilizada passou por sucessivos refinamentos, com base em testes nos mecanismos de busca e na opinião de especialistas. Após um refinamento iterativo, a *string* empregada foi:

((("Big Data" OR "data analytics" OR "Big Dataset" OR "massive data" OR "database" OR "nosql" OR "data science" OR "data-intensive") AND ("geographic" OR "spatial" OR "geospatial" OR "gis" OR "geodatabase" OR "cyberGIS" OR "Spatiotemporal")) OR "big spatial data" OR "big geospatial data" OR "big spatiotemporal data") AND ("internet of things" OR "iot"))

A *string* foi calibrada e aplicada nas bases acadêmicas selecionadas. Os resultados das diferentes bases foram unificados e as duplicatas foram removidas, assim como prefácios, índices, *keynotes*, atas de conferências e resumos estendidos. Ao fim dessa etapa, 4398 estudos foram compilados para seleção.

3.3 Etapa 2: Seleção e Leitura

A seleção foi realizada em três passos. A cada passo, os textos eram lidos e os considerados como não-relevantes eram removidos do conjunto, não sendo considerados no passo seguinte. O critério de inclusão dos trabalhos era escolher estudos primários apresentando *big data* em IoT com dimensões geoespaciais. Os critérios de exclusão adotados foram:

- EC1 - O estudo não aborda uma das áreas de interesse do mapeamento (*big data* geoespacial ou IoT) ou não responde a nenhuma das perguntas de pesquisa;

¹<https://www.scopus.com/>

²<https://ieeexplore.ieee.org/>

³<https://dl.acm.org/>

⁴<https://www.webofknowledge.com/>

- EC2 - O estudo não está escrito em inglês;
- EC3 - Não foi possível obter acesso ao trabalho;
- EC4 - O estudo não é um trabalho completo (por exemplo: resumos estendidos e atas de conferências não detectados na pré-filtragem.)
- EC5 - Os resultados do estudo são melhor descritos em outro trabalho já incluído no mapeamento.
- EC6 - O estudo não é revisado por pares.

No primeiro passo, os artigos foram selecionados com base na leitura do resumo, resultando em 572 papers. No segundo passo, a introdução e a conclusão dos textos foram consideradas, e 212 artigos permaneceram. No terceiro passo, os textos remanescentes foram selecionados com base na leitura completa, resultando em um total de 165 artigos (Figura 3.1).

Ao final da Etapa 2, um processo de *snowballing* foi realizado, onde as referências dos trabalhos selecionados foram analisadas em busca de outros estudos que também pudessem responder questões de pesquisa. Além disso, foi conduzida uma busca manual complementar, incluindo um processo de *forward snowballing* (rastreamento das citações aos trabalhos mapeados), visando abranger na busca também o primeiro semestre de 2025. Nestas atividades, 31 artigos foram adicionados ao mapeamento. Ao final do processo, 196 estudos primários foram considerados aptos a responder alguma das questões de pesquisa. A Tabela 3.1 descreve os resultados da seleção, detalhando o número de trabalhos incluídos/excluídos em cada etapa.

Tabela 3.1: Resultados da seleção.

Decisão	Passo 1	Passo 2	Passo 3	Snowballing
Excluído - EC1	3807	230	39	-
Excluído - EC2	4	14	-	-
Excluído - EC3	3	94	-	-
Excluído - EC4	11	22	3	-
Excluído - EC5	1	-	3	-
Excluído - EC6	-	-	2	-
Excluído - Total	863	184	36	-
Incluído	572	212	165	31
Total				196

A Figura 3.2 apresenta o grafo de coocorrência de palavras-chave, construído no software VOSviewer [58], a partir das palavras-chave extraídas durante a leitura integral dos

artigos selecionados. Cada nó representa uma palavra-chave, cujo tamanho é proporcional à sua frequência de ocorrência nos trabalhos. As conexões indicam a coocorrência entre os termos, enquanto as cores agrupam as palavras-chave mais fortemente relacionadas. Essa visualização permite identificar com maior clareza os principais temas, domínios e ferramentas abordados na literatura analisada.

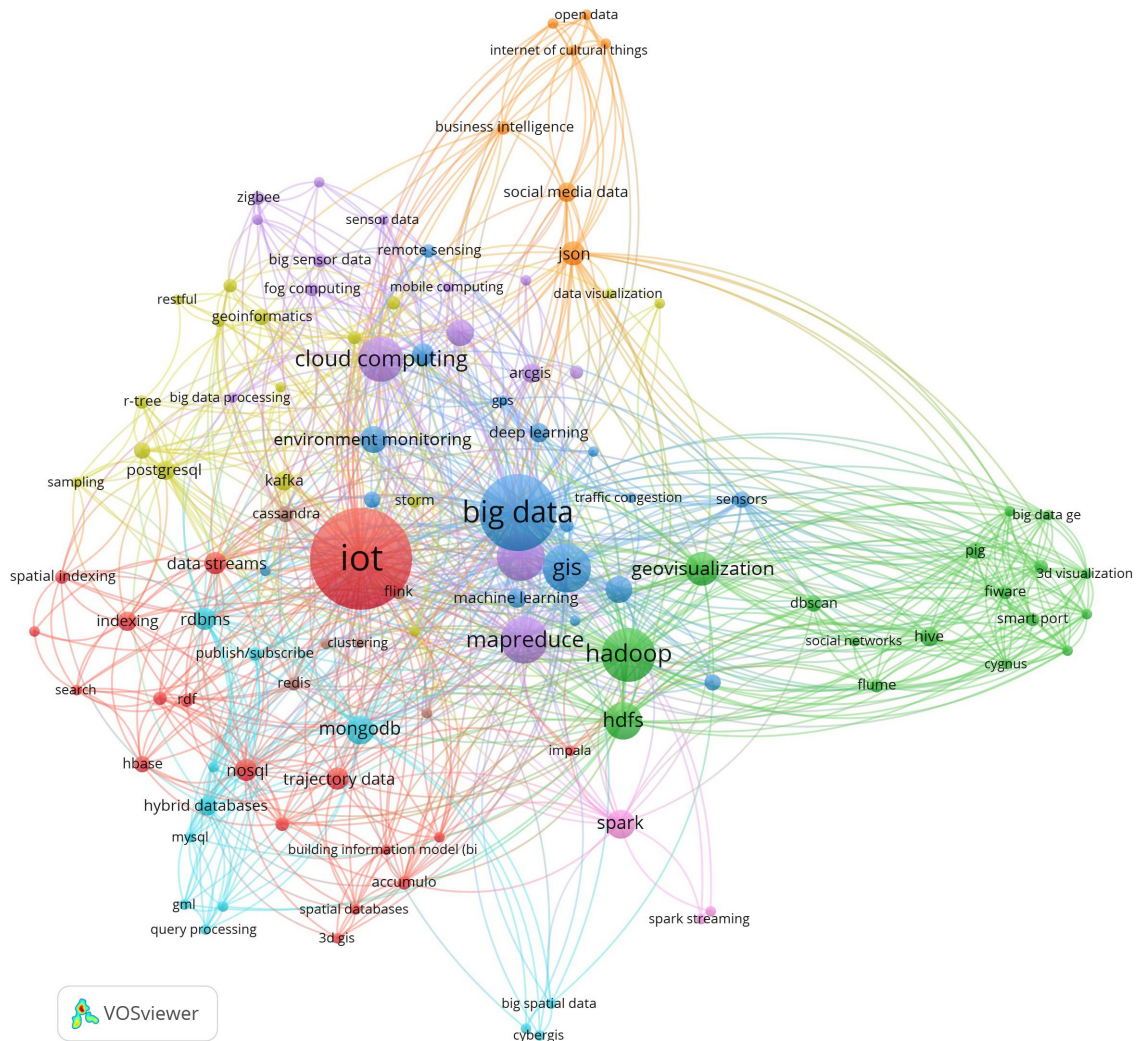


Figura 3.2: Coocorrência de palavras-chave. Fonte: [55]

3.4 Resultados

RQ1 - Em quais domínios *big data* geoespacial tem sido aplicados em IoT?

A diversidade de dispositivos IoT tem guiado o desenvolvimento de novas aplicações, permitindo novos serviços ubíquos e pervasivos em diversos domínios, como *smart cities* e planejamento urbano, monitoramento ambiental, automação industrial e residencial,

saúde, *e-commerce*, dentre outros [8, 5, 9, 10]. A Figura 3.3 demonstra a presença de *big data* geoespacial em domínios IoT nos trabalhos analisados.

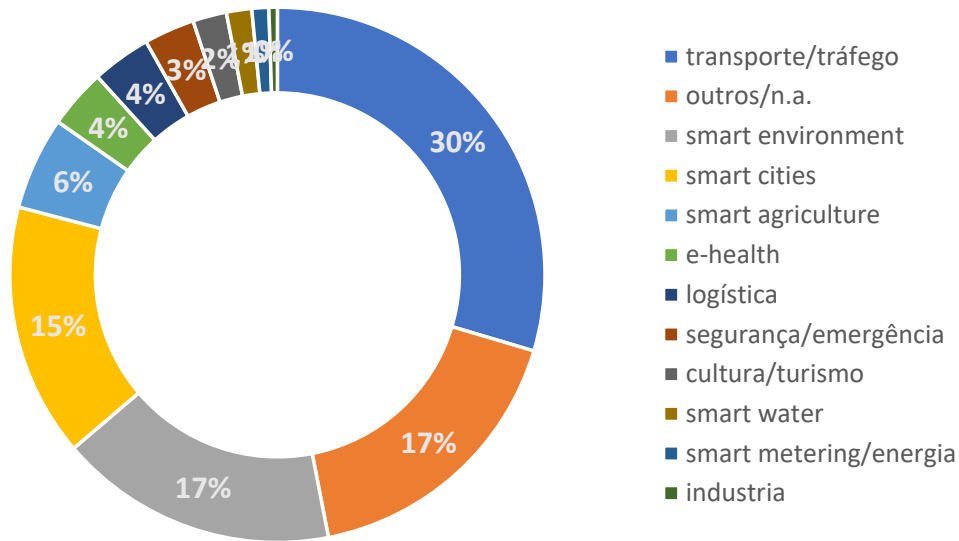


Figura 3.3: Domínios de aplicação.

Um número significativo de estudos foi encontrado descrevendo a aplicação de *big data* geoespacial em *smart cities*, monitorando diversos aspectos do planejamento urbano, como gerenciamento de distribuição de energia [59, 60], planejamento urbano [61, 62], e gerenciamento de distribuição de energia, água e esgoto [63, 64, 65].

Alvarez, Morales e Kraak [66] propõem uma abordagem para adicionar capacidades espaço-temporais a serviços de informação, e uma arquitetura de referência para aplicações orientadas a eventos em *smart cities*. Ela foi implementada na SmartSantander, uma rede de sensores instalada em Santander, Espanha, que monitora aspectos da vida urbana como tráfego, estacionamento e irrigação.

Um aspecto do desenvolvimento urbano que se destacou foi o controle de tráfego e aplicações relacionadas a transportes [40]. Sensores embutidos em veículos geram grandes quantidades de *Floating Car Data* (FCD) – registros de localização e *timestamps* durante os trajetos – e *Extended Floating Car Data* (XFCD) – complemento do FCD com dados de sensores veiculares – obtidos via interfaces de *On-Board Diagnosis* (OBD) [67, 68]. Isso pode estimular o desenvolvimento de veículos inteligentes, capazes de transmitir dados para análise em tempo real [69] em grandes aplicações de IoV (*Internet of Vehicles*) [70]. Os dados veiculares podem ser utilizados para *crowdsensing*, com aplicabilidade em estimativas de velocidade de tráfego e controle de fluxo [71]. Finogeev et al. [40] descrevem um modelo de processamento de *big sensor data* para um sistema de *smart roads* que monitora incidentes e o status das vias. Unal, Kocak e Donmez [69] apresentam um projeto para um serviço de *smart winter* que auxilia em atividades como remoção de

neve e aplicação de materiais de degelo. Dados veiculares são coletados e processados na nuvem, alimentando serviços de monitoramento e planejamento de rotas.

Além disso, há uma presença significativa de *big data* geoespacial em aplicações de *smart environment*, como monitoramento de desastres [72], gerenciamento de qualidade do ar [27, 73, 74, 75], e gerenciamento de água e esgotos [76].

Outros domínios também foram mencionados, como logística [77, 78, 79, 80, 40, 81, 82], cultura e turismo [83, 84, 85, 86], e *Smart Water* [76, 87].

RQ2 - Qual a fonte dos *datasets* empregados nos experimentos realizados nos estudos?

A disponibilidade dos dados pode influenciar a replicação dos experimentos e a implementação de novos trabalhos, razão pela qual é relevante verificar se os dados utilizados advêm de aplicações proprietárias, bem como se são reais ou sintéticos. Ademais, é importante identificar quais *datasets* foram empregados na literatura de maneira mais recorrente.

Conforme evidenciado na RQ1, aplicações IoT podem agregar dados geoespaciais originários de diversas fontes. Para aplicações com geovisualização, imagens e vetores georreferenciados podem ser extraídos de aplicações geoespaciais como o OpenStreetMap, Google Maps e Earth, Baidu Map e Ali CCloud Server [88]. Além dessas fontes, os estudos empregam dados de diversos tipos de sensores embutidos em dispositivos equipados com sistemas de navegação [66, 89, 90]. Dispositivos móveis como *smartphones* também são fontes de dados amplamente utilizadas em aplicações IoT [91, 92, 93, 94].

Também foi investigado se os *datasets* empregados no estudo são reais ou sintéticos (Figura 3.4). Dentre os 196 estudos, observou-se que 162 (81.9%) empregam dados reais, e 11 (5.8%) usam dados sintéticos. Em 18 estudos (9.6%), a origem dos *datasets* não foi detalhada, e em 5 estudos (2.7%), foram empregados tanto dados reais quanto simulados. Os projetos reais de IoT abrangiam diversas fontes e domínios. Alguns trabalhos usam *datasets* abertos e disponíveis na web, dentre os quais podem ser mencionados o Graph of Things [95], o registro de viagens da *NYC Taxi and Limousine Commission*⁵ e o BerlinMOD⁶.

RQ3 - Quais os principais formatos de dados armazenados?

Em IoT, dados espaciais podem aparecer em diferentes formatos. Aplicações com sensores móveis armazenam dados relacionados a trajetórias [96, 25, 97, 98, 85, 99] ou a localização de sensores [10], construções [100] e dispositivos móveis [86, 101]. Além disso, localizações

⁵<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

⁶<http://dna.fernuni-hagen.de/secondo/BerlinMOD/BerlinMOD.html>

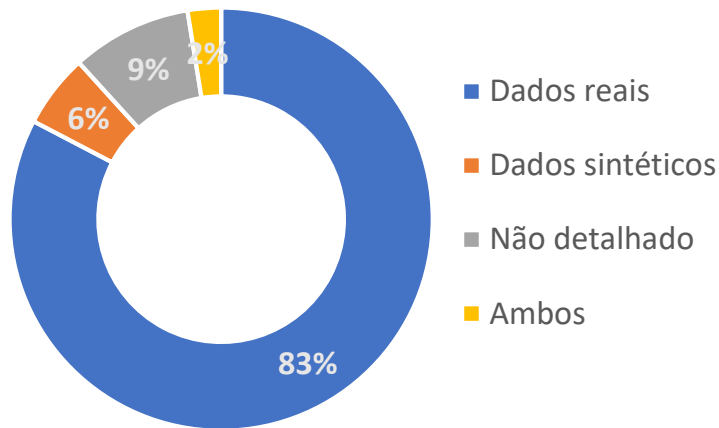


Figura 3.4: Origem dos *datasets*.

relacionadas a elementos abstratos podem ser armazenadas, como eventos [40] e conteúdo de mídias sociais [102].

Algumas aplicações, principalmente relacionadas a geo-visualização, também lidam com mapas digitalizados em diversos formatos [27], *shapefiles* [27, 103, 104, 105], imagens geo-referenciadas de sensoriamento remoto [106], informação de terreno extraída por escaneamento LIDAR (*Laser Imaging Detection and Ranging*) [80], Geovídeo [107], dentre outros tipos de dados.

Alguns formatos de arquivo usados no armazenamento de informação espacial encontrados foram o KML (*Keyhole Markup Language*) [74, 108] e o KMZ (*Zipped KML*) [60]. O OGC (*Open Geospatial Consortium*) propôs a GML (*Geography Markup Language*) [104, 105, 109]. Há também aplicabilidade para JSON (*JavaScript Object Notation*) [77, 82, 83, 84, 102, 110, 111, 53], um formato para armazenamento e troca de dados; e GeoJSON [10, 94, 112], uma variação do JSON orientada à codificação de estruturas geográficas.

RQ4 - Quais as ferramentas mais utilizadas para lidar com *big data* geoespacial em IoT?

Essa questão visa investigar o suporte de software à manipulação de *big data* geoespacial de IoT. Há ferramentas de *big data* com diferentes propósitos e características, mas também há espaço para melhoria no suporte a dados geoespaciais. Alguns exemplos de bancos NoSQL utilizados nos trabalhos mapeados são Neo4j [10, 113], KairosDB [76], HBase [97, 114, 81], MongoDB [97, 78, 94, 105, 104, 107, 111, 115] e LevelDB [102].

Vale mencionar que, apesar da disponibilidade de ferramentas para lidar com *big data*, muitos trabalhos ainda empregaram bancos relacionais, principalmente o PostgreSQL e sua extensão espacial, PostGIS [9, 52, 62, 67]. Outros bancos relacionais foram aplicados,

como o Microsoft SQL Server [64, 74], VoltDB [61] e Informix [53]. Junto dessas extensões geoespaciais, alguns Sistemas de Informação Geográfica (GIS, *Geographic Information Systems*) como o ArcGIS [74, 116, 117, 118] e o QGIS [10, 119] também são utilizados, principalmente para consulta e visualização de dados geoespaciais.

Alguns trabalhos implementam abordagens híbridas, com dados alocados em diferentes bancos. Ding, Xu e Yang [52] combinam bancos relacionais e chave-valor. Cai et al. [104] propõem um esquema de banco de dados multicamadas para aplicações IoT *Online-to-Offline* (O2O) e divide os dados - inclusive os geográficos - nos bancos, de acordo com suas características. Seus algoritmos de busca consideram informação geográfica - que usa um índice espacial - e de contexto. A implementação combina bancos relacionais e NoSQL.

O ecossistema Apache oferece diversos recursos para *big data* geoespacial. O Apache Hadoop [120], implementação *open-source* do MapReduce, possui grande aplicabilidade e aparece em diversos trabalhos [121, 80, 2]. Jo e Lee [2] empregam Hadoop/MapReduce em uma abordagem D_ELT (*delayed extracting-loading-transforming*) para análise de *big data* geoespacial de sensores. Fang et al. [122] aplicam MapReduce em um modelo geoespacial para detectar incidentes em recursos terrestres. MapReduce também foi utilizado para processar *streams* de dados de sensores [90, 123], e Xia et al. [116] empregaram-no na predição de fluxo de tráfego em tempo real. Sambrekar e Rajpurohit [121] aplicam Hadoop MapReduce na classificação de dados agrícolas, como umidade e temperatura, coletados por sensores geograficamente dispersos. O *Hadoop Distributed File System* (HDFS) [124] também foi aplicado em diversos trabalhos, como [10, 2, 80]. Spark [96, 125] oferece uma *engine* para processamento distribuído de *big data*. Storm [126, 99, 61] é empregado na computação distribuída de *data streams*. Kafka [127, 94, 99] é uma plataforma distribuída empregada em *streaming*.

RQ5 - Como os trabalhos empregando *big data* geoespacial em IoT podem ser classificados?

Um dos objetivos do mapeamento é classificar os trabalhos empregando *big data* com dimensão geoespacial em IoT. Na etapa de leitura, os artigos foram classificados em seis grupos, com base em seus objetivos e contribuições (Tabela 3.2):

G01 - Aplicações IoT empregando *big data* geoespacial: O primeiro grupo descreve aplicações IoT contendo *big data* com aspectos geoespaciais. Os trabalhos nesse grupo englobaram vários dos domínios listados na RQ1, como transporte e tráfego [118, 119, 154, 158, 170, 188, 192, 193, 247, 248], *smart cities* [61, 64, 69, 62, 60, 159, 160], logística [40, 81, 77, 80] e *smart environment* [72, 73, 74, 92, 122, 145, 148, 171, 173].

Tabela 3.2: Classificação dos estudos.

Grupo	Estudos
G01 (70 trabalhos)	[128] [129] [130] [131] [132] [133] [134] [135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [80] [92] [88] [64] [69] [83] [72] [122] [145] [146] [27] [94] [74] [147] [40] [81] [77] [148] [117] [86] [149] [62] [76] [79] [150] [151] [73] [152] [153] [154] [59] [155] [156] [157] [87] [158] [159] [160] [161] [162] [163] [164] [165] [93] [60] [166] [167] [168] [61] [169] [170] [171] [172] [173]
G02 (40 trabalhos)	[174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [101] [114] [10] [185] [102] [66] [9] [186] [187] [125] [1] [53] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [89] [99] [198] [78] [199] [200] [52]
G03 (60 trabalhos)	[201] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211] [212] [213] [214] [215] [216] [217] [218] [219] [65] [111] [84] [220] [118] [63] [221] [123] [222] [223] [71] [82] [224] [90] [225] [2] [226] [106] [91] [227] [228] [229] [230] [231] [232] [233] [234] [100] [75] [235] [236] [237] [116] [238] [119] [239] [240] [241] [242] [243] [244]
G04 (14 trabalhos)	[105] [104] [107] [97] [103] [121] [85] [245] [246] [247] [248] [249] [250] [251]
G05 (6 trabalhos)	[96] [115] [252] [70] [253] [254]
G06 (6 trabalhos)	[255] [98] [67] [256] [257] [258]

G02 - Artefatos de suporte à implementação de aplicações IoT empregando *big data* geoespacial: Os artigos neste grupo descrevem a criação de ferramentas, modelos, arquiteturas, *frameworks*, etc., que auxiliam o projeto de aplicações IoT com grandes volumes de dados contendo alguma dimensão geoespacial. Alguns exemplos são Liu et al. [189], com um esquema de *Accurate Range Query* (ARQ) para serviços IoT baseados em localização em servidores na nuvem, que preservam a privacidade dos dados e das consultas do usuário; Isikdag [10], que testa vários métodos de aquisição, armazenamento, transmissão e visualização, e propõe uma arquitetura para integração em tempo real de geoinformação IoT; Ding, Xu e Yang [52], que propõem o SeaCloudDM, um mecanismo de gerenciamento de *big data* para sensores heterogêneos; e You et al. [101] e Alvarez, Morales e Kraak [66] que propõem arquiteturas aplicáveis na criação de serviços em *Smart Cities*.

G03 - Análise de dados: Neste grupo, os estudos focaram na análise de dados. Há uma demanda significativa para processamento em tempo real de *streams* de *big data* geoespacial [96]. Sensores móveis localizados em veículos [224, 67], *wearables* e dispositivos móveis transmitem informação em tempo real, que pode ser utilizada em serviços de localização, detecção e predição de fluxo e velocidade do tráfego [116, 71, 118, 239, 242, 244, 243], planejamento de rotas [69], soluções em transporte e mobilidade urbana, utilização de recursos terrestres [221], dentre outros. Leung, Braun e

Cuzzocrea [111] usam *deep supervised learning* em dados de transporte para compreender a mobilidade dos cidadãos e desenvolver melhores modelos de transporte. Algoritmos de Inteligência Artificial (IA) e Aprendizagem de Máquina (ML, *Machine Learning*) podem ser empregados na detecção de anomalias [123] e monitoramento de eventos [90] em dados de sensores. Qiu et al. [219] empregam IA na análise de informação geoespacial de sensores, objetivando melhorar a infraestrutura da rede IoT. Em monitoramento ambiental, ML também é aplicável. Chiwewe e Ditsela [225] trabalham com a predição do nível de ozônio ao nível do solo, baseada em *cross-correlation* e na correlação espacial das mensurações de poluentes em estações de monitoramento distribuídas. Sambrekar e Rajpurohit [121] propõe um modelo para classificação em tempo real de dados agrícolas multi-dimensionais, empregando Hadoop para lidar com o grande volume de dados. Jiang et al. [230] realizam predição do fluxo de pessoas para sistemas de social IoT (SIoT) em espaço urbano, baseada em *usage detail records* (UDRs) extraída de *big data* de redes móveis. Shuja et al. [233] descreve um *framework* para clusterização hierárquica geo-textual para aplicações de Social IoT, que alcança alta eficiência no uso de tempo e memória, sem comprometer significativamente a qualidade da clusterização em comparação a um *framework* híbrido padrão com a mesma finalidade.

G04 - Organização de dados: Os artigos nesta categoria relatam experimentos envolvendo organização de dados espaciais. Alguns trabalhos descrevem a viabilidade de armazenar os dados geoespaciais de forma distinta dos demais dados de aplicação, combinando sistemas de gerenciadores de bancos de dados e outras estruturas de armazenamento para melhorar a manipulação e consulta aos dados. Luan et al. [105] propõem um *framework* para dados espaciais IoT que separa a informação espacial dos demais conteúdos, e combina bancos relacionais, NoSQL e sistemas de arquivo para uma maior eficiência. Wu et al. [107] apresenta uma abordagem híbrida para gerenciar e organizar dados de sensores, contendo um *Main Memory Database* (MMDB), um banco relacional, um sistema de arquivos distribuído para lidar com *big data* geoespacial, e um banco NoSQL como estrutura de armazenamento extensível. A abordagem é demonstrada empregando geovideo de segurança pública. Zhou, Li e Tu [97] propõem uma estratégia de *path-divided data blocking* (PDDB) para otimizar o acesso a *hyper massive spatio-temporal trajectory data* (HMSTD), atualmente limitado às técnicas vigentes de indexação espaço-temporal. Al Jawarneh et al. [245] desenvolveram um protótipo baseado no Apache Spark para uma *engine* de processamento *batch* espacial que dissemina lotes de dados espaciais em *clusters* de *cloud computing*, aplicando um método de particionamento baseado na localização dos dados espaciais. Como uma resposta ao enriquecimento dos dados de mobilidade em contextos de *big data* de IoT, Mello et al. [85] propõem a MASTER, uma abordagem para modelar trajetórias de múltiplos aspectos – informações semânticas representando fatos

do mundo real relevantes para a análise de dados – e armazená-las em uma *triplestore* RDF baseada em NoSQL.

G05 - Indexação de dados: A indexação de dados influencia o desempenho da busca e recuperação de dados, especialmente em aplicações em tempo real [9]. Alguns trabalhos descrevem experiências envolvendo indexação de dados baseada em informação de contexto espacial ou temporal [252]. Zhong, Fang e Zhao [70] propõem um esquema de indexação distribuída para otimizar consultas em dados de sensores espaço-temporais. R-Trees se destacam dentre os métodos de indexação, devido à sua eficiência no gerenciamento e processamento de dados [97, 96]. Limkar e Jha [96] investigam a relação entre a forma de geração das R-Trees e o tempo necessário para indexar os dados geoespaciais. Jamjoum et al. [254] apresentam o GeoApproxJoin, um *framework* para a execução eficiente de junções *Point-in-Polygon*, que utiliza duas técnicas de indexação espacial: Geohash e H3.

G06 - Geovisualização: Geospatial *big data* permite a utilização de diversas formas de geovisualização, e alguns estudos enfatizam esse aspecto. Em *smart cities*, foram encontrados estudos sobre visualização em sistemas de monitoramento de distribuição de energia [60]. Visualização 3D e Realidade Virtual (VR, *Virtual Reality*) apresentam diversas possibilidades práticas [185]. Yu et al. [256] apresentam uma plataforma de visualização 3D para cabos de transmissão baseada em *big data* espaço-temporal, permitindo o uso de óculos VR. A implementação do SmartPort, um sistema para gerenciamento e visualização 3D de dados de sensores heterogêneos de sensores do porto de Las Palmas em Gran Canaria, é descrita em alguns dos trabalhos mapeados [77, 79, 80]. Lwin et al. [98] explora a criação de *dashboards* geoespaciais para coletar, manipular, compartilhar e visualizar dados de objetos geograficamente dispersos de *smart cities*. Hossam et al. [255] propõe uma solução para ajudar o governo egípcio no monitoramento da expansão da COVID-19, contendo um *dashboard* interativo que declara o número e localização de casos confirmados e mortes pela doença.

A análise da literatura ofereceu uma perspectiva da pesquisa aplicando *big data* geoespacial em IoT, que tem avançado sob múltiplas frentes e revela cada vez mais oportunidades e desafios. A pesquisa na área geoespacial já é consolidada, e as linhas de *big data* e IoT estão em crescentes níveis de interesse da comunidade. É possível observar a intersecção entre esses três temas, dado o número elevado de publicações sobre dados geoespaciais em IoT, *big data* geoespacial, e *big data* em IoT. A disponibilidade de soluções em IoT tem crescido ao longo dos anos, exigindo formas de armazenamento e realização de operações nos dados dessas aplicações. Esse desafio requer o engajamento de desenvolvedores e pesquisadores ligados a múltiplas especialidades.

Na literatura analisada, o volume dos dados não foi a única característica de *big data*

considerada. A velocidade é um fator relevante, dada a importância de estruturas robustas de ingestão e análise de dados em tempo real. A variedade também é presente, devido à pluralidade de domínios de aplicação, que dificulta a criação de soluções IoT genéricas. Há iniciativas para interoperabilidade de geodados em IoT, como o *Sensor Observation Service* (SOS), parte do *Open Geospatial Consortium* (OGC) *Sensor Web Enablement* (SWE), mas ainda com impacto limitado [5]. Este mapeamento identificou algumas iniciativas nesse sentido, como Laska et al. [99], que propuseram uma arquitetura de integração de fluxos de dados IoT espaço-temporais; e Isikdag [10], que apresenta uma arquitetura para integração de sensores heterogêneos instalados em ambientes IoT, capaz de adquirir, armazenar, trocar e visualizar geoinformação. A visualização é uma característica importante relacionada a *big data* geoespacial. A comunidade pode se beneficiar de interfaces web ricas e amigáveis, e *dashboards* para permitir a visualização, análise e consulta de *big data* espacial de IoT em representações 2D/3D/4D/VR [80, 98, 185, 256, 255]. Um desafio apontado por Kamilaris e Osterman [5] é a combinação de *datasets* heterogêneos no desenvolvimento de modelos de visualização 4D, permitindo rastrear a evolução dos dados para detectar e mensurar mudanças em diferentes domínios, como construções, logística e agricultura.

A adoção e popularização de padrões de dados geoespaciais nas ferramentas de gerenciamento de dados IoT deve ser estimulada. Existem propostas bem definidas para a representação de dados geoespaciais e sensores, especialmente iniciativas de interoperabilidade OGC, como a SWE e SensorThings. Essas propostas de padronização podem ser mais extensivamente adotadas pela comunidade, especialmente em contextos de *big data*.

Iyer e Stoica [115] levantam alguns desafios na manipulação de *big data* geoespacial em IoT. A natureza das consultas evoluiu da busca em dados predominantemente estáticos para a análise de grandes volumes de dados heterogêneos e com elevada taxa de ingestão. As aplicações IoT começaram a demandar mecanismos robustos de injeção e busca de dados. Técnicas mais robustas de indexação espacial, como R-trees, quadrees, octrees, geohashes, e indexação baseada em grids, ainda podem melhorar o desempenho de aplicações em dados espaciais dinâmicos e distribuídos [259]. O balanceamento dos objetos indexados em ambientes distribuídos é também complicado em aplicações IoT, com dados imprevisíveis e com veracidade variável.

Ao lidar com grandes volumes de dados, estudos sobre o desempenho de bancos de dados espaciais de IoT em relação às características de mobilidade podem beneficiar buscas e análise espaciais. Alguns trabalhos se aprofundam na diferenciação entre dados de sensores e de localização, e propõem abordagens híbridas para o manuseio de *big data* de IoT. Um exemplo é descrito por Junior et al. [260], com uma arquitetura híbrida para dados de IoT onde dados escalares e posicionais foram separados e avaliados em três

bancos: Redis, MongoDB e Cassandra. Observou-se que, em geral, o Redis apresentou melhor desempenho geral para ambos os tipos de dados, apesar de não terem sido propostos experimentos relacionados a operações espaciais com os dados posicionais. Quoc e Le Phuoc [114] apresentam um mecanismo de consulta espaço-temporal para dados de sensores baseado em *linked data*. Para o armazenamento de dados climáticos (de natureza temporal) é empregado o OpenTSDB, ao passo que o Elasticsearch é empregado no armazenamento de trajetórias.

A Tabela 3.3 indica, dentre os trabalhos encontrados no mapeamento da literatura, os relacionados com a arquitetura a ser proposta. A coluna “Mob” indica se o trabalho descreve a utilização de dispositivos fixos (F), móveis (M) ou um misto de ambos os tipos (X). Por fim, a coluna “Suporte” enumera os recursos de software empregadas pelos autores no suporte ao armazenamento e análise dos dados geoespaciais. Os trabalhos que descrevem atividades de processamento e análise, contém funcionalidades de geovisualização e empregam dispositivos fixos e móveis estão destacados em negrito.

Na literatura referente à aplicação de *big data* geoespacial em IoT é possível encontrar diversos estudos descrevendo o gerenciamento de dados geoespaciais em aplicações contemplando dispositivos com diferentes padrões de mobilidade [16]. Em alguns estudos, o volume de dados gerado consiste majoritariamente de séries temporais de observações produzidas por dispositivos total ou predominantemente estacionários [232, 231, 148, 187]. Outros trabalhos lidam tanto com trajetórias de dispositivos móveis quanto suas observações que variam ao longo do tempo [9, 91, 116, 88, 97]. Por fim, alguns estudos descrevem aplicações que gerenciam tanto dispositivos móveis quanto estacionários em diversos domínios, como *smart cities* [66, 61, 117], *smart ports* [79, 80, 77], logística [40], controle de tráfego [154, 157], dentre outros [52, 114, 191]. Fernández et al. [79, 80] e Suarez et al. [77] descrevem uma implementação de *smart port* que manipula dados heterogêneos, oriundos tanto de sensores estacionários quanto presentes nas embarcações. Alvarez, Morales e Kraak [66], Sukmaningsih et al. [117] e Rathore et al. [61] apresentam soluções para *smart cities* dotadas de sensores móveis e estacionários. Finogeev et al. [40] realizam monitoramento de incidentes de tráfego, utilizando tanto dados de foto-radares quanto de dispositivos móveis. Baig et al. [191] apresenta o SPEAR, um *framework* para processamento de streams espaço-temporais, e propõe um conjunto generalizado de tipos de dados espaço-temporais estacionários e móveis, e funções de *range* e *nearest neighbor* com suporte para ajuste dinâmico para objetos móveis. Wang et al. apresentam em [139] um sistema para reduzir a sobrecarga na alocação de recursos de recarga para uma frota de mais de 13.000 táxis elétricos (ETaxi) na cidade de Shenzhen, na China. Além dos registros geolocalizados dos ETaxis, são armazenados metadados de 117 estações de recarga. Wang et al. propõem em [179] um método de preenchimento de dados espaço-temporais

Tabela 3.3: Análise dos trabalhos relacionados.

Trabalho	Mob.	Suporte
Artigo [10]	X	Neo4j
Artigo [40]	X	
Artigo [59]	F	
Artigo [61]	X	Hadoop, Spark, MapReduce, VoltDB, Storm, S4
Artigo [62]	X	PostgreSQL, PostGIS
Artigo [66]	X	
Artigo [67]	M	PostgreSQL, PostGIS, leaflet.
Artigo [69]	M	Kafka, Flink
Artigo [76]	F	Jena, KairosDB
Artigo [77]	X	HDFS, MapReduce
Artigo [79]	X	Hadoop, MapReduce, HDFS
Artigo [80]	X	Hadoop, MapReduce, HDFS, Hive, Pig
Artigo [84]	F	
Artigo [92]	M	
Artigo [94]	M	Kafka, MongoDB
Artigo [99]	M	Storm, Kafka
Artigo [89]	X	Cassandra, Oracle
Artigo [93]	M	MongoDB, MapReduce
Artigo [97]	M	
Artigo [98]	M	
Artigo [101]	X	
Artigo [111]	M	MongoDB
Artigo [114]	X	OpenTSDB, ElasticSearch
Artigo [117]	X	
Artigo [119]	M	PostgreSQL, MongoDB, QGIS
Artigo [122]	X	MapReduce
Artigo [125]	M	Spark
Artigo [151]	F	Cassandra, Redis, Kafka, Flink
Artigo [158]	M	SQL Server
Artigo [166]	F	Storm, Redis, Cassandra, HDFS
Artigo [167]	X	Flume, HDFS, Spark, Impala
Artigo [170]	M	PostgreSQL, MapReduce
Artigo [183]	X	
Artigo [184]	X	
Artigo [186]	M	Hadoop, Spark, Kafka
Artigo [190]	M	Hbase
Artigo [185]	X	Accumulo
Artigo [187]	F	HDFS, Spark, Accumulo, Geoserver, GeoMesa
Artigo [191]	X	Apache Flink
Artigo [192]	F	
Artigo [193]	M	Apache Spark; GeoTools; HDFS;
Artigo [197]	F	
Artigo [245]	M	Spark
Artigo [255]	X	
Artigo [257]	M	Elasticsearch

de *Mobile CrowdSensing* em grande escala, realizando análises em dois *datasets*: o U-Air, com dados de qualidade do ar oriundos de dispositivos móveis geolocalizados, e o Sensor-Scope, com dados de sensores estacionários.

Entretanto, mesmo em trabalhos reportando aplicações voltadas ao gerenciamento simultâneo de dispositivos fixos e móveis, não foram encontradas abordagens que considerem explicitamente o impacto de diferentes padrões de mobilidade sobre o armazenamento dos dados, e tampouco análises sobre o desempenho de ferramentas para *big data* geoespacial quando alimentadas para esses cenários. Em contraposição, a arquitetura proposta nesta tese, detalhada no Capítulo 5, contempla tanto dispositivos fixos quanto móveis, e a experimentação conduzida no Capítulo 6 avalia aspectos relacionados ao armazenamento — como o consumo de espaço — e à análise — como o tempo de recuperação nas consultas — em ambientes de *big data* geoespacial, considerando contextos com diferentes níveis de mobilidade.

3.5 Considerações Finais

Este capítulo apresentou um levantamento dos principais estudos relacionados à aplicação de *big data* geoespacial em IoT, conduzido por meio de um mapeamento sistemático da literatura. Discussões adicionais sobre as atividades conduzidas e os resultados alcançados nesse mapeamento encontram-se descritos na publicação [55]. A investigação realizada na RQ4 foi considerada na elaboração do Capítulo 4, no qual se apresenta um levantamento das principais soluções de software referenciadas na literatura para o processamento e a análise de *big data* geoespacial, abrangendo também a descrição de seus recursos geoespaciais e dos meios mais comuns documentados de integração entre elas.

Capítulo 4

Levantamento de Ferramentas

Este capítulo apresenta um levantamento das principais ferramentas referenciadas na literatura para o processamento e a análise de *big data* geoespacial. A seguir, apresenta-se a estrutura do capítulo: A Seção 4.1 fornece uma visão introdutória sobre o tema; a Seção 4.2 descreve as soluções voltadas ao armazenamento de dados geoespaciais; a Seção 4.3 discute as ferramentas consolidadas destinadas ao processamento e à análise desses dados; e, por fim, a Seção 4.4 aborda as considerações finais do capítulo.

4.1 Contextualização

Conforme discutido no Capítulo 2, o interesse por soluções de armazenamento e processamento de dados geoespaciais acompanha a evolução das tecnologias da informação. Embora SGBDs relacionais, como o PostgreSQL com a extensão PostGIS, já ofereçam suporte consolidado a dados espaciais, pesquisas recentes têm despertado crescente interesse em alternativas paralelas e distribuídas, como bancos NoSQL, sistemas de arquivos distribuídos e *frameworks* de *big data* [1, 37, 38]. A distribuição e replicação de dados e serviços possibilitam escalabilidade, maior disponibilidade e elasticidade, além de viabilizar o processamento de grandes volumes de dados geoespaciais com melhor relação custo-benefício [37].

Na escolha de plataformas para o desenvolvimento de aplicações baseadas em *big data* geoespacial, alguns aspectos devem ser cuidadosamente considerados:

- **Volume de dados:** a plataforma deve prover mecanismos de armazenamento escalável e capacidades de processamento adequadas para lidar com volumes de dados em constante crescimento.
- **Operações espaciais:** a plataforma deve disponibilizar um conjunto de operações espaciais abrangente e compatível com o nível de complexidade exigido pela apli-

cação. Entre as funcionalidades usualmente oferecidas por ferramentas de análise geoespacial, destacam-se o cálculo de distâncias e áreas, junções, interseções, uniões entre polígonos e a conversão entre diferentes sistemas de coordenadas.

- **Suporte a múltiplos formatos de dados:** entre os diversos formatos utilizados por sistemas de informação geoespacial, destacam-se KML (Keyhole Markup Language), GML (Geography Markup Language), shapefiles e GeoJSON. Plataformas voltadas a dados geoespaciais devem oferecer suporte a geometrias como **Point**, **Line** e **Polygon**. Além disso, mecanismos de indexação espacial — como R-Trees, Quad-Trees e HashMaps — impactam diretamente o desempenho das consultas.
- **Interconexão com outras plataformas:** a plataforma deve possibilitar integração com softwares de *data science*, *machine learning*, geovisualização e outros recursos relacionados a dados geoespaciais.
- **Funcionalidades de geovisualização:** a plataforma deve oferecer recursos para a geração de mapas e outros formatos de visualização de dados geoespaciais.
- **Facilidade de uso:** além de questões de usabilidade, as plataformas devem ser fáceis de instalar, configurar e utilizar, possuir manuais e tutoriais atualizados, e contar com uma comunidade de usuários ativa.
- **Custo:** recursos necessários para obter, instalar e manter a infraestrutura de software.

A Figura 4.1 apresenta algumas das plataformas *open-source* identificadas na RQ4 do mapeamento de literatura realizado no Capítulo 3 e que oferecem funcionalidades de armazenamento, consulta e geovisualização de *big data* geoespacial. Em aplicações que apresentam requisitos mais complexos, é possível adotar abordagens híbridas, nas quais os dados são distribuídos em diferentes plataformas de acordo com suas características específicas. Dessa forma, informações geoespaciais por exemplo podem ser armazenadas de maneira distinta dos demais dados da aplicação, de modo a explorar mecanismos especializados de manipulação e recuperação de dados. Essa estratégia viabiliza a integração de estruturas de armazenamento e tecnologias de banco de dados, com a finalidade de aprimorar os processos de manipulação, otimizar o desempenho operacional e aumentar a eficiência na recuperação das informações.

Além disso, a integração entre diferentes *frameworks* contribui para o aumento e a diversificação das funcionalidades disponíveis. Há diversas possibilidades de interação entre essas ferramentas, sendo que a figura ilustra as formas mais recorrentes, conforme descrito em suas documentações oficiais. As linhas contínuas indicam ferramentas que compõem

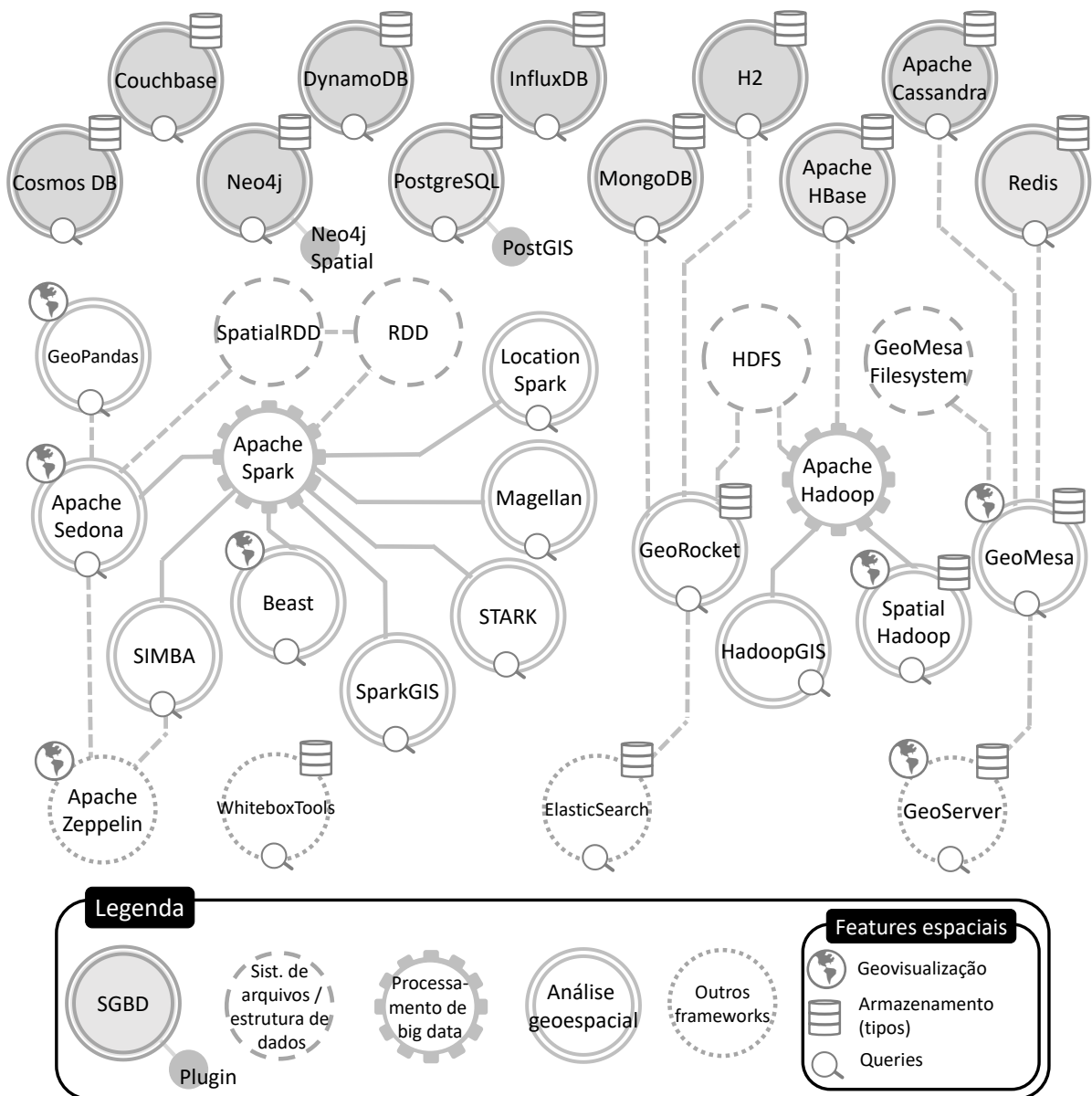


Figura 4.1: Plataformas para *big data* geoespacial.

ou servem de base para o desenvolvimento de outras, enquanto as linhas tracejadas representam plataformas que se relacionam por meio de *plugins* ou de outros mecanismos de integração. Diversos bancos como MongoDB, Cassandra, e Redis também oferecem algum tipo de suporte a dados espaciais [31]. *Frameworks* voltados à manipulação de *big data*, como Apache Hadoop e Apache Spark, ainda apresentam limitações em relação ao suporte a operações geométricas quando comparados a bancos de dados tradicionais, demandando, em muitos casos, extensões específicas para o tratamento de tipos de dados e operações espaciais [261]. O Hadoop conta com extensões como SpatialHadoop e Hadoop-GIS, enquanto o Spark é complementado por ferramentas como Apache Sedona, SpatialSpark, LocationSpark, SIMBA e Magellan. Entre as tecnologias baseadas em nu-

vem voltadas ao gerenciamento de dados geoespaciais, destacam-se GeoRocket e Google Earth Engine. Nas próximas subseções, serão detalhadas as funcionalidades fornecidas por essas plataformas para o armazenamento e a manipulação de grandes volumes de dados geoespaciais.

4.2 Armazenamento de *Big Data* Geoespacial

Dentre os principais requisitos para SGBDs voltados a aplicações geoespaciais, destacam-se o custo de armazenamento, o desempenho das consultas, a flexibilidade na troca de esquemas, a capacidade de manipular grandes volumes de dados e a integração com ferramentas de análise [45]. Essas características colocam os bancos NoSQL em evidência, uma vez que são capazes de armazenar dados estruturados, semi-estruturados e não estruturados, além de utilizarem linguagens de consulta distintas do SQL tradicional [1]. Além disso, sua eficiência os torna adequados para o suporte a *streams* de *big data* [157]. No caso de demanda por maior capacidade de processamento ou armazenamento, bancos NoSQL podem escalar horizontalmente a um custo relativamente baixo, fazendo *clusters* de alta disponibilidade e tolerância a falhas. A Tabela 4.1 apresenta alguns dos SGBDs mais citados em trabalhos descrevendo aplicações que armazenam e analisam *big data* geoespacial, juntamente com o seu modelo de dados predominante e seus principais recursos geoespaciais, como disponibilidade de indexação, linguagens de consulta e formatos de dados suportados.

O PostgreSQL – que serve de base ao Postgres-XL – e seu *plugin* PostGIS são consolidados e apresentam amplo suporte a aspectos geoespaciais. Outros bancos populares são o MongoDB, Couchbase, Redis, Cassandra, Neo4j, InfluxDB, Cosmos DB, e DynamoDB. Os bancos NoSQL em geral contam com um rol mais reduzido de opções geoespaciais, visto que muitas operações complexas – como junções e agregações – e estruturas de indexação não foram originalmente projetadas para lidar com os desafios referentes à distribuição dos dados ao longo de múltiplos servidores [12]. Portanto, a pesquisa e desenvolvimento nessa área ainda reserva diversas oportunidades e desafios.

A disponibilidade de opções de geometrias e tipos de operações espaciais se dá de diferentes formas de acordo com o sistema utilizado. A Tabela 4.2 enumera as principais geometrias e operações espaciais dos bancos analisados.

4.3 Processamento de *Big Data* Geoespacial

Além dos SGBDs, diversas ferramentas oferecem funcionalidades de gerenciamento e análise de *big data* geoespacial. A Tabela 4.3 apresenta alguns dos *frameworks*, bibliotecas e

Tabela 4.1: SGBDs com suporte a *big data* e seus recursos geoespaciais.

SGBD	Modelo e licença	Indexação	Linguagem de consulta	Formatos
PostgreSQL e PostGIS	Relacional, <i>open source</i>	✓	SQL	(PostGIS) Importação: KML, GML, GeoJSON, GeoHash, WKT; Renderização: GeoTIFF, PNG, JPG, NetCDF.
MongoDB	Documentos, <i>open source</i>	✓	MQL (MongoDB Query Language)	GeoJSON
Redis	Chave-valor, <i>open source</i>	✓	Comandos Redis	GeoJSON
Cassandra	Colunas, <i>open source</i>	✓	Cassandra Query Language (CQL)	WKT
DynamoDB	Multi-modelos, comercial	✓	Queries DynamoDB	GeoJSON
Neo4j e Neo4j Spatial	Grafos, <i>open source</i>	✓	Cypher	Neo4j: Graph Model, RDF Neo4j Spatial: Shapefile, GeoJSON, OSM
Cosmos DB	Multi-modelos, comercial	✓	Azure Cosmos DB SQL API	GeoJSON
InfluxDB	Séries temporais, <i>open source</i>	✓	InfluxQL	InfluxDB Line Protocol, JSON, Graphite
Couchbase	Documentos, <i>open source</i>	✓	Couchbase REST API, N1QL	GeoJSON

ferramentas mais utilizados em aplicações que armazenam e analisam dados geoespaciais, acompanhados de um resumo de suas principais características, como o ano de criação (+) e da última atualização (*) do projeto, geometrias compatíveis, principais operadores e funções geoespaciais disponíveis, e recursos relacionados à geovisualização.

O GeoRocket é um *data store* de alto desempenho voltado para formatos de arquivo geoespaciais. Com ênfase em escalabilidade, ele suporta diferentes formas de armazenamento de dados, como sistemas de arquivos locais ou distribuídos, os bancos H2 e MongoDB, e o serviço de armazenamento em nuvem Amazon S3.

Os principais *frameworks* voltados ao processamento de *big data* contam com extensões para suportar funções e tipos de dados espaciais. O Apache Hadoop¹, utilizado em conjunto com o HDFS (*Hadoop Distributed File System*), serve como base para o desenvolvimento de diversos *frameworks*, como o Hadoop-GIS², um sistema de *data warehousing*

¹<https://hadoop.apache.org/>

²<https://bmidb.cs.stonybrook.edu/hadoopgis/index>

Tabela 4.2: SGBDs populares com suporte a *big data*, suas geometrias e operações geo-espaciais.

Banco	Geometrias			Operações Espaciais									
	Point	Line/LineString	Polygon	Bounding Box	Distance	Overlaps	Intersects	Contains	Disjount	Within	WithinDistance	Area	Touch
PostgreSQL / PostGIS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MongoDB	✓	✓	✓		✓		✓			✓			
Redis	✓				✓								
Cassandra	✓	✓	✓				✓	✓		✓			
DynamoDB	✓									✓			
Neo4j e Neo4j Spatial	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cosmos DB	✓	✓	✓		✓		✓			✓			
InfluxDB	✓	✓	✓		✓		✓	✓			✓		
Couchbase	✓	✓	✓	✓							✓		

para dados espaciais que estende o Apache Hive; e o SpatialHadoop³ (descontinuado).

Frameworks baseados no Apache Spark⁴ para processamento de *big data* geoespacial tem sido crescentemente utilizados, visto que eles tem apresentado desempenho superior aos baseados em Hadoop [262]. Alguns deles são:

- Magellan⁵, a primeira biblioteca a estender o Spark SQL oferecendo uma abstração relacional para análise geoespacial;
- STARK⁶, que adiciona suporte a tipos de dados e operações espaciais;
- SparkGIS⁷, que é capaz de reescrever dinamicamente consultas para gerenciar cargas de trabalho que excedam os recursos distribuídos disponíveis;
- SIMBA⁸ (*Spatial In-Memory Big data Analytics*), que incorpora índices espaciais em contextos de RDDs (*Resilient Distributed Datasets*) e SQL, além de possuir otimizadores físicos e lógicos para selecionar os melhores planos de execução de consultas;

³<https://github.com/aseldawy/spatialhadoop2>

⁴<https://spark.apache.org/>

⁵<https://github.com/harsha2010/magellan>

⁶<https://github.com/dbis-ilm/stark>

⁷<https://bmidb.cs.stonybrook.edu/sparkgis/index>

⁸<https://github.com/InitialDLab/Simba>

Tabela 4.3: Ferramentas para análise de *big data* geoespacial.

Ferramenta	Geometrias	Principais operadores e funções	Geovisualização
Apache Sedona +2016 *2025	Point, MultiPoint, LineString, Polygon, MultiLineString, MultiPolygon, GeometryCollection	Range, Range join, Distance join, KNN, Conversão entre Sistemas de Coordenadas de Referência	Geração de mapas em alta resolução via SedonaViz, com operadores para visualização de mapas e <i>map tiles</i>
Beast +2021 *2024	Point, MultiPoint, LineString, Polygon, MultiLineString, MultiPolygon, GeometryCollection	Spatial join	Visualização <i>single image</i> e multilevel para exploração interativa.
GeoRocket +2020 *2024	Point, MultiPoint, LineString, Polygon, MultiLineString, MultiPolygon, GeometryCollection	Via Elasticsearch	N/A
GeoMesa +2015 *2025	Point, MultiPoint, LineString, Polygon, MultiLineString, MultiPolygon, GeometryCollection	Via SparkSQL ou API GeoJSON	Integração com Leaflet e GeoServer.
HadoopGIS +2013 *2013	Point, Box, Polygon, LineString	Intersects, KNN, Contains, Touches, Area, Distance, funções agregadas, dentre outras	N/A
Spatial Hadoop +2015 *2018	Point, Rectangle, Polygon	Range queries, KNN, Spatial joins, dentre outras.	Duas operações: <i>plot</i> (gera mapas a partir de arquivos) e <i>plotp</i> (gera <i>map tiles</i>)
STARK +2017 *2022	Point, MultiPoint, LineString, Polygon, MultiLineString, MultiPolygon, GeometryCollection	Filter (intersects, contains, containedBy, withinDistance), joins, clustering (DBSCAN), kNN	N/A
SparkGIS +2015 *2016	Point, Line, MultiPoint, MultiLine, Polygon	Dimension, Area, Equals, Envelope, Length, Crosses, Centroid, Contains, Disjoint, Distance, Intersects, Within, Overlaps, Touches, Union, Boundary	N/A
Magellan +2015 *2017	Point, MultiPoint, Polygon, MultiPolygon, LineString	range queries, spatial joins	N/A
SIMBA +2016 *2017	Point, LineString, Polygon, Circle	range, kNN, Distance join, kNN join	Visualização interativa via Zeppelin (modificação)
Location Spark +2016 *2017	Point, Rectangle	Range, kNN, Spatial join, distance join, kNN join	N/A
Whitebox Tools +2017 *2022	Point, MultiPoint, PolyLine, Polygon, PolyLineZ, PolygonZ, MultiPointZ	Análise GIS, process. de imagens, análise hidrológica espacial, análise geomorfométrica e process. de dados LiDAR.	N/A

- LocationSpark⁹, que utiliza filtros em seus índices para evitar comunicação desnecessária na rede durante o processamento de dados espaciais;
- Beast¹⁰ (*Big Exploratory Analytics for Spatio-temporal data*), projeto dos criadores do SpatialHadoop; e

⁹<https://github.com/purduedb/LocationSpark>

¹⁰<https://bitbucket.org/bdlabucr/beast/src/master/>

- Apache Sedona¹¹ (anteriormente conhecido como GeoSpark), que permite o processamento em *cluster* de *big data* geoespacial em memória, utilizando SRDDs (SpatialRDDs, uma extensão geoespacial dos RDDs do Spark), para particionamento espacial dos dados no *cluster* com base na proximidade geográfica.

Outro destaque é o GeoMesa¹², um pacote de ferramentas voltado ao armazenamento, busca e análise de dados geoespaciais que utiliza como *backend* bancos como Accumulo, HBase, Google Bigtable ou Cassandra.

WhiteboxTools¹³ é um pacote de análise de dados contendo mais de 450 ferramentas de processamento de dados geoespaciais. Sua instalação e interface de *scripting* é simples, e muitas ferramentas podem operar em paralelo em processadores *multicore*. Além disso, a plataforma pode servir como *backend* de análise para outras ferramentas de GIS, sensoriamento remoto e softwares geoespaciais (ex.: *plugins* do QGIS e ArcGIS).

GeoPandas [263] é uma biblioteca de código aberto para análise de dados espaciais em Python, que adiciona às funcionalidades do Pandas a manipulação, análise e visualização de dados geoespaciais, permitindo lidar com geometrias como pontos, linhas e polígonos. Oferece suporte a formatos espaciais como Shapefile, GeoJSON e PostGIS; operações espaciais (ex.: interseção, união); e visualização de mapas com integração ao matplotlib.

Visualização de dados é um processo que permite a obtenção de *insights* a partir de conjuntos de dados complexos e com volume expressivo. No que diz respeito a dados geoespaciais, a geovisualização se beneficia da adequabilidade de dados geoespaciais à apresentação interativa em um contexto visual na forma de mapas 2D, 3D e 4D, *dashboards* e realidade virtual, dentre outros métodos de representação [33, 80, 98]. Algumas plataformas de *big data* geoespacial, como Spatial Hadoop, SIMBA, Beast, GeoMesa, e Apache Sedona, possuem funcionalidades de geovisualização. SIGs tradicionais, como ArcGIS e QGIS, ainda são aplicados comumente, principalmente para consulta e visualização de dados. Há também serviços comerciais de mapas que permitem visualização de dados geoespaciais, como Google Maps e MapBox, e sistemas escaláveis de visualização de mapas, como HadoopViz, SHAHED, e GeoTrellis.

4.4 Considerações Finais

Este capítulo apresentou um levantamento das principais ferramentas referenciadas na literatura para o processamento e a análise de *big data* geoespacial. Foram abordadas soluções voltadas ao armazenamento de dados geoespaciais, bem como ferramentas con-

¹¹<https://sedona.apache.org/>

¹²<https://www.geomesa.org/>

¹³<https://www.whiteboxgeo.com/geospatial-software/>

solidadas para o processamento e a análise desses dados. As tecnologias apresentadas serviram como base para a seleção das ferramentas utilizadas na implementação da arquitetura proposta no Capítulo 5, garantindo que a escolha considerasse tanto o suporte a grandes volumes de dados quanto a capacidade de manipulação e análise de informações geoespaciais de forma eficiente.

Capítulo 5

Proposta

Este capítulo apresenta uma proposta de arquitetura para o gerenciamento e análise de *big data* geoespacial proveniente de dispositivos IoT com diferentes níveis de mobilidade, e está organizado da seguinte forma: A Seção 5.1 apresenta a arquitetura proposta e suas respectivas camadas; a Seção 5.2 descreve as principais ferramentas empregadas na arquitetura, incluindo Postgres-XL com PostGIS, Apache Spark, Apache Sedona e HDFS; a Seção 5.3 apresenta a implementação do protótipo da arquitetura, utilizado como base para a experimentação conduzida nesta tese; e, por fim, a Seção 5.4 aborda as considerações finais do capítulo.

5.1 Arquitetura Proposta

Esta tese apresenta uma arquitetura para o gerenciamento e análise de *big data* geoespacial proveniente de dispositivos IoT com diferentes níveis de mobilidade. Por gerenciamento, nesta proposta compreende-se o armazenamento dos dados gerados pelos dispositivos IoT, incluindo tanto as localizações quanto as observações realizadas, considerando seus diferentes padrões de mobilidade. Por análise, consideram-se as consultas efetuadas sobre esses dados. Os dispositivos IoT gerenciados pela arquitetura possuem informações de geolocalização e, no caso dos dispositivos móveis, podem registrar múltiplas localizações ao longo do tempo. Além disso, eles são capazes de capturar diferentes observações do ambiente por meio de sensores. A Figura 5.1 ilustra de forma simplificada a arquitetura proposta, que é formada por duas camadas, descritas a seguir: Armazenamento e Análise de Dados e Aplicação.

A primeira camada, denominada **Armazenamento e Análise de Dados**, tem como finalidade armazenar, gerenciar e fornecer suporte à análise das informações provenientes dos dispositivos IoT. Para esse propósito, contempla duas funcionalidades centrais: o armazenamento de dados e o suporte a operações geoespaciais. Ambas exigem mecanismos

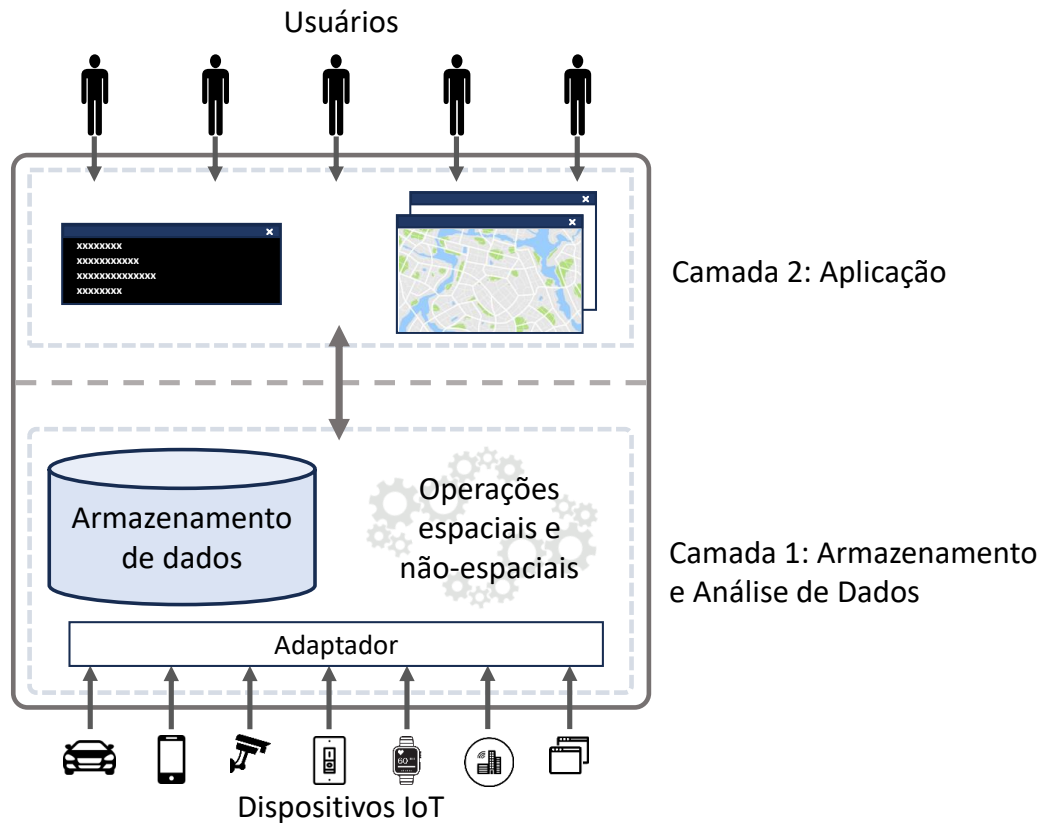


Figura 5.1: Arquitetura proposta.

distribuídos capazes de lidar com formatos de dados heterogêneos, bem como de realizar a ingestão, em tempo real, de grandes volumes de dados oriundos de dispositivos com diferentes níveis de mobilidade.

A segunda camada, de **Aplicação**, se situa no nível superior da arquitetura e atua como interface entre o usuário e os dados armazenados, permitindo a realização de consultas aos dados armazenados por meio de formulários e a visualização dos resultados obtidos.

Um protótipo da arquitetura foi desenvolvido com o objetivo de viabilizar a validação e a condução de experimentos na camada de Armazenamento e Análise de Dados. Nenhum experimento de validação foi realizado sobre a camada de aplicação, uma vez que questões de usabilidade e interação com o usuário estão além do escopo desta pesquisa. A implementação detalhada de cada camada da arquitetura é apresentada na Seção 5.3.

5.2 Ferramentas Utilizadas

5.2.1 Postgres-XL e PostGIS

Postgres-XL [264] é uma extensão que oferece escalabilidade ao PostgreSQL, sem prejuízo de suas funcionalidades, compatibilidades e consistência ACID. Pode ser configurado com múltiplos servidores, que oferecem às transações uma visão unificada do banco – cada atualização em um servidor é imediatamente visível a todos os demais. Tabelas são armazenadas no Postgres-XL de forma distribuída, sejam replicadas ou particionadas, de forma transparente às aplicações. O Postgres-XL é composto dos seguintes componentes:

- GTM (*Global Transaction Manager*, ou Gerenciador Global de Transações): responsável pelo gerenciamento consistente de transações. Assim como o PostgreSQL, também mantém a consistência dos dados empregando o modelo MVCC (*Multi-Version Concurrency Control*) [265]. Os coordenadores e *datanodes* podem se conectar diretamente a ele, mas uma vez que cada transação realiza solicitações ao GTM com frequência, podem ser utilizados *proxies* GTM para lidar com as demandas dos coordenadores e *datanodes*.
- *Coordinator* (Coordenador): interface do banco para aplicações que funciona como o processo tradicional de *backend* do PostgreSQL. O *cluster* pode ter diversos coordenadores, todos com uma visão consistente do banco, e todos podem receber comandos SQL, os examinar, criar sub-planos, determinar os *datanodes* envolvidos e solicitar que eles executem o comando (ou partes dele), respeitando as propriedades MVCC. Todas as atualizações feitas por um coordenador tornam-se visíveis imediatamente para todos os demais.
- *Datanode*: onde os dados são armazenados nas tabelas distribuídas/replicadas. Não possui uma visão global do banco, apenas gerencia os dados armazenados localmente. O *cluster* pode conter diversos *datanodes*, que por sua vez podem receber solicitações de diferentes coordenadores. Para oferecer paralelismo, quando um coordenador gera um plano de execução ele é enviado aos *datanodes*, que se comunicam diretamente entre si.

Para equilibrar a carga e aumentar a eficiência das consultas, o Postgres-XL distribui os dados entre os nós do *cluster* por meio de um mecanismo de fragmentação baseada em *hash*. Cada linha da tabela é associada a um valor de *hash* calculado sobre uma coluna de distribuição, e esse valor determina o nó em que a linha será armazenada. Caso a coluna de distribuição não seja especificada durante a criação da tabela, o sistema utiliza, por padrão, a chave primária como base para o cálculo do *hash*.

O PostGIS [266] é uma extensão do banco PostgreSQL que permite armazenar, indexar e consultar dados geoespaciais. Assim como o PostgreSQL, o Postgres-XL também é compatível com o PostGIS. Dentre suas principais funcionalidades, podem ser citadas:

- Armazenamento de diversos tipos de dados espaciais, como pontos, linhas, polígonos e multi-geometrias.
- Indexação espacial para rápida busca e recuperação dos dados com base em localização.
- Funções para filtragem e análise espacial, medição de distâncias e área, e intersecção de geometrias.
- Processamento e manipulação de geometrias, como simplificação, conversão e generalização.
- Processamento e armazenamento de dados *raster*.
- Integração com ferramentas como QGIS, GeoServer, MapServer e ArcGIS.

5.2.2 Apache Spark

O Apache Spark¹ é uma *engine* de processamento paralelo que permite processar prioritariamente em memória cargas de trabalho de *big data* [12]. Os principais componentes do Apache Spark são [96]:

- *Spark context*: é o ponto de entrada principal da aplicação Spark, que especifica como ocorrerá a interação com o gerenciador de *cluster*.
- *Cluster manager*: o nó principal do Spark, também chamado de *spark master*. Administra os recursos disponíveis no *cluster*, como número de CPUs, memória disponível e quantidade de executores.
- *Worker node*: local onde os executores trabalham. Em cada executor, múltiplas tarefas – unidades básicas de execução – podem estar sendo processadas. Também possuem cache, para manter os dados que serão utilizados na computação em memória.

O Spark lida com três tipos de objetos:

- *Resilient Distributed Dataset* (RDD), a estrutura de dados básica do Spark, consistindo de uma coleção distribuída de registros processada em paralelo pelo *cluster*. RDDs não possuem esquema e permitem maior controle sobre as operações, sendo indicados para dados não estruturados (ex.: *logs*, texto bruto).

¹<https://spark.apache.org/>

- *Dataframe*, uma coleção distribuída de dados estruturados em colunas com esquema definido. *Dataframes* são construídos sobre RDDs, mas são mais eficientes devido a otimizações automáticas. Podem ser manipulados com SQL ou APIs Spark. Recomendados para contextos em que os dados são estruturados ou semi-estruturados (ex.: JSON, CSV, Parquet).
- *Dataset*, um conjunto distribuído de dados que combina os benefícios dos RDDs (tipagem forte, possibilidade de utilizar funções lambda) aos dos *dataframes* (eficiência, plano de execução otimizado do Spark SQL). Recomendados para lidar com objetos complexos, combinando eficiência e controle.

5.2.3 Apache Sedona

O Apache Sedona [267] é um mecanismo para *big data* geoespacial projetado para sistemas de computação em *cluster*, como o Apache Spark e o Apache Flink. Permite aos desenvolvedores elaborar suas tarefas de processamento de dados espaciais em Spatial SQL, Spatial Python e Spatial R. Oferece funcionalidades de carga, indexação, particionamento e processamento/otimização de consultas.

Algumas de suas principais funcionalidades são:

- Suporte a diversos formatos de dados geoespaciais, incluindo GeoJSON, WKT e ESRI Shapefile.
- Processamento distribuído de grandes *datasets* de vetores e *raster*.
- Ferramentas para indexação, consultas e junções espaciais.
- Integração com ferramentas populares de *big data*, como Spark, Hadoop, Hive e Flink para armazenamento e consultas espaciais.
- API para dados geoespaciais em SQL, Python, Scala e Java.

5.2.4 Hadoop Distributed File System (HDFS)

HDFS é um sistema de arquivos distribuídos projetado para executar em *cluster*. Foi desenvolvido como um sistema de arquivos subjacente para o ecossistema Hadoop, e tem sido amplamente utilizado pelo Apache Spark [261]. HDFS oferece tolerância a falhas para cada bloco de dados por meio da replicação em múltiplos nós. Além disso, é escalável, permitindo que novos nós sejam adicionados e redistribuindo automaticamente os dados considerando os nós adicionados.

Um *cluster* HDFS típico segue uma arquitetura mestre-trabalhador, consistindo de um *namenode* ativo e múltiplos *datanodes*.

- O *namenode* gerencia os metadados que indicam em quais *datanodes* cada bloco de dados está armazenado e é responsável por enviar ao cliente a lista dos *datanodes* que contêm os blocos solicitados.
- Os *datanodes* armazenam os blocos de dados particionados, e enviam os dados solicitados ao cliente.

5.3 Implementação

Nesta seção são apresentados detalhes da implementação das duas camadas que compõem a arquitetura proposta (Figura 5.2): Armazenamento e Análise de Dados e Aplicação. Na subseção dedicada à primeira camada, apresentam-se o modelo de dados — que abrange informações sobre os dispositivos, suas localizações e as observações realizadas no ambiente — e a descrição dos dois ambientes empregados para o armazenamento e a análise desses dados. Na subseção correspondente à segunda camada, são abordados os recursos utilizados para a realização de consultas e para a geovisualização dos resultados, destacando as ferramentas que possibilitam a interação do usuário com os dados processados.

5.3.1 Armazenamento e Análise de Dados

Nesta camada, são realizadas as consultas aos dados de geolocalização dos dispositivos IoT e às leituras coletadas em seus sensores. Para os experimentos conduzidos nesta tese, os dados foram simulados a partir de *datasets* reais de trajetórias e observações de sensores. O ponto de entrada da camada de Armazenamento e Análise de Dados é um receptor, responsável pela coleta, padronização e organização dos seguintes tipos de dados oriundos dos dispositivos IoT:

- **Dados dos dispositivos:** Nome e descrição do dispositivo;
- **Dados de localização:** Coordenadas geoespaciais (latitude e longitude), momento do registro da localização (*timestamp*);
- **Dados de observação:** Momento da coleta do dado (*timestamp*), valor observado (dado coletado pelo sensor);
- **Dados das propriedades observadas:** Nome da propriedade (ex.: temperatura, umidade, pressão, etc.), descrição da propriedade e unidade de medida correspondente (ex.: °C para temperatura, hPa para pressão).

A arquitetura proposta foi testada em dois ambientes computacionais, cada um composto de uma combinação de ferramentas de armazenamento e processamento de dados geoespaciais. A escolha das ferramentas foi realizada com base nos seguintes critérios:

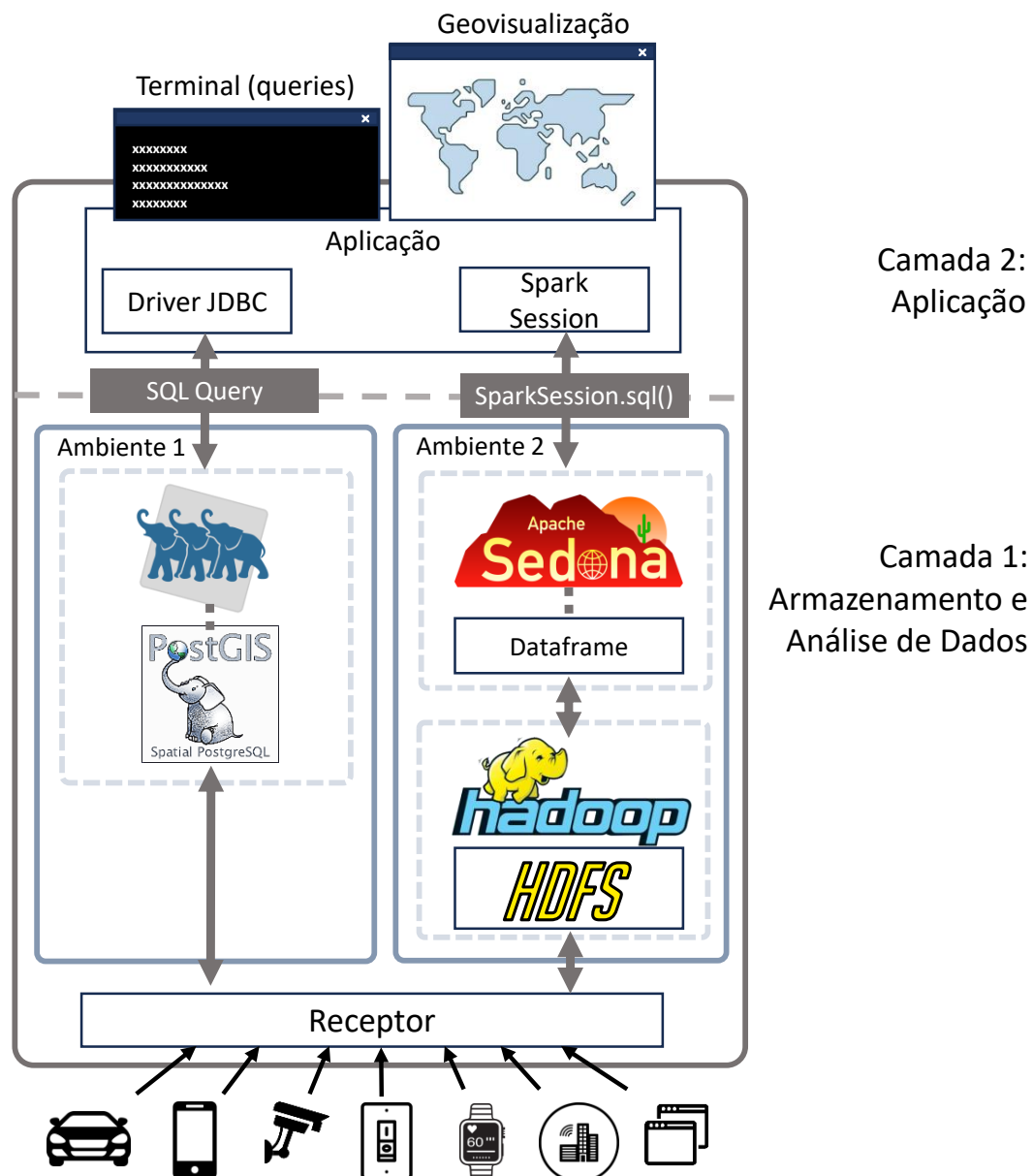


Figura 5.2: Instanciação da arquitetura proposta.

- Suporte ao armazenamento distribuído de grandes volumes de dados geoespaciais;
- Eficiência no uso do espaço de armazenamento dos dados;
- Disponibilidade de índices e formatos de dados geoespaciais;
- Suporte às linguagens de consulta, com possibilidade de realizar não apenas consultas espaciais, mas também consultas sobre as observações registradas pelos dispositivos;
- Diversidade e desempenho de operações geoespaciais;

- Compatibilidade com outros artefatos de software relacionados a armazenamento e análise de dados geoespaciais; e
- Interesse da comunidade de desenvolvedores, aplicação recorrente em pesquisas na área de *big data* geoespacial, e disponibilidade de documentação.

Dentre as possibilidades de ferramentas analisadas no Capítulo 4, o Ambiente 1 foi implementado utilizando o Postgres-XL (solução distribuída baseada no PostgreSQL) em conjunto com o PostGIS. O Ambiente 2 foi desenvolvido utilizando o HDFS integrado ao Apache Sedona. Em ambas, as funcionalidades de ingestão e consulta de dados geoespaciais foram implementadas em linguagem Java. As próximas subseções apresentam a modelagem conceitual e lógica dos dados utilizados na arquitetura proposta, e descreve os ambientes computacionais definidos para a implementação da camada de Armazenamento e Análise de Dados.

Modelagem Conceitual

A Figura 5.3 apresenta o modelo conceitual adotado na implementação da camada de Armazenamento e Análise de Dados, na forma de um Diagrama Entidade-Relacionamento (DER), e as Tabelas 5.1 a 5.4 apresentam os dicionários de dados das entidades modeladas, nos quais são descritos os atributos, seus tipos e respectivos significados. A modelagem adotada foi inspirada na especificação da API SensorThings [268] proposta pela OGC, porém com adaptações e simplificações para atender aos objetivos desta tese. Essa escolha buscou manter a consistência semântica com um padrão amplamente reconhecido para representação de dados de sensores e observações, ao mesmo tempo em que reduziu a complexidade do modelo para possibilitar sua implementação de forma mais direta nos ambientes considerados.

Tabela 5.1: Dicionário de dados da entidade Dispositivo (*Entity*).

Atributo	Tipo	Descrição
idEntity	Inteiro	Identificador único do dispositivo.
entityName	String	Nome do dispositivo.
entityDescription	String	Descrição do dispositivo.
entityLocation	Coordinate	Coordenadas geoespaciais (latitude e longitude) atuais do dispositivo.

Modelagem Lógica

O modelo conceitual descrito para a camada de Armazenamento e Análise de Dados foi desnormalizado na transição para o modelo lógico e posterior implementação. Essa deci-

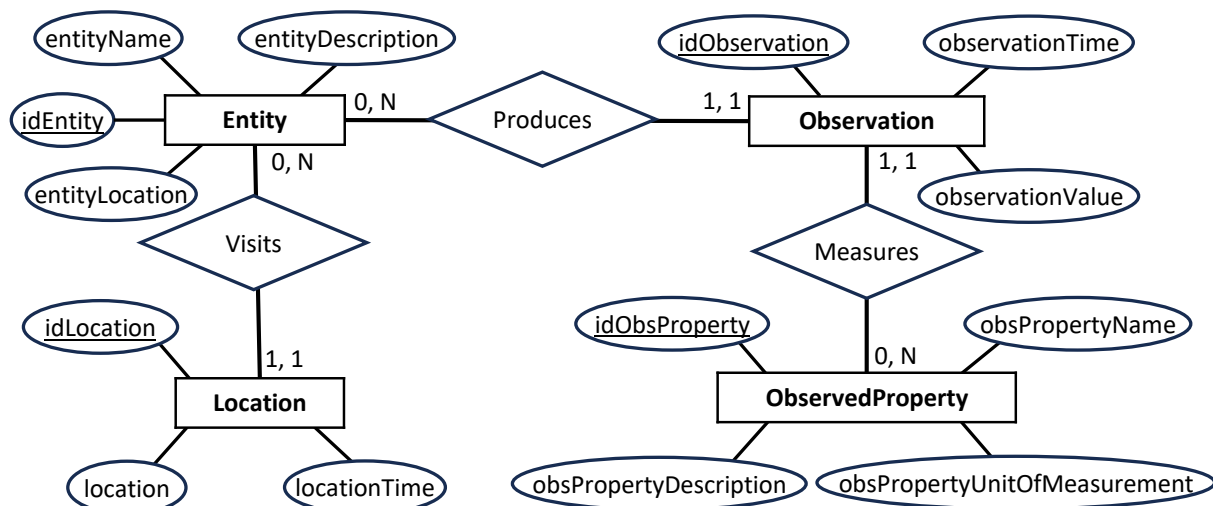


Figura 5.3: Modelo conceitual de dados (DER) adotado na arquitetura proposta.

Tabela 5.2: Dicionário de dados da entidade Observação (*Observation*).

Atributo	Tipo	Descrição
idObservation	Inteiro	Identificador único do registro de observação.
observationTime	DateTime	<i>Timestamp</i> da coleta do dado de observação.
observationValue	String	Valor do dado de observação coletado pelo sensor.

Tabela 5.3: Dicionário de dados da entidade Localização (*Location*).

Atributo	Tipo	Descrição
idLocation	Inteiro	Identificador único do local associado ao dispositivo.
location	Coordinate	Coordenadas geoespaciais (latitude e longitude) do dispositivo no momento da geração do registro.
locationTime	DateTime	<i>Timestamp</i> do registro da localização.

Tabela 5.4: Dicionário de dados da entidade PropriedadeObservada (*ObservedProperty*).

Atributo	Tipo	Descrição
idObsProperty	Inteiro	Identificador único da propriedade observada.
obsPropertyName	String	Nome da propriedade observada (ex.: temperatura, umidade, pressão).
obsPropertyDescription	String	Descrição da propriedade observada.
obsPropertyUnitOfMeasurement	String	Unidade de medida correspondente à propriedade observada (ex.: °C para temperatura, hPa para pressão).

são teve como objetivo reduzir a complexidade das consultas e minimizar as operações de junção, que apresentam custo computacional elevado, de modo a favorecer o desempenho na manipulação de grandes volumes de dados geoespaciais. Dessa forma, as entidades e

relacionamentos foram consolidados em uma única tabela (no Postgres-XL) ou em um *DataFrame* distribuído (no Apache Sedona), mantendo a consistência semântica do modelo original, mas privilegiando a eficiência operacional. A Tabela 5.5 apresenta o modelo desnormalizado da arquitetura, detalhando os atributos armazenados para cada registro, bem como exemplos de dados para cada campo.

Tabela 5.5: Modelo de dados desnormalizado da arquitetura.

Atributo	Tipo	Exemplo de Dado
idRecord	Inteiro	258
idEntity	Inteiro	8
entityName	String	ME08
entityDescription	String	Mobile Entity 08
idLocation	Inteiro	131
location	Coordinate	41.9434853,-87.6813467
locationTime	DateTime	2018-01-01T01:05:06
idObservation	Inteiro	258
observationTime	DateTime	2018-01-01T01:05:06
observationValue	String	1013,25
idObsProperty	Inteiro	2
obsPropertyName	String	pressure
obsPropertyDescription	String	Force exerted by the atmosphere or another fluid on a surface.
obsPropertyUnitOfMeasurement	String	hPa

Ambiente 1 – Postgres-XL e PostGIS

O ambiente baseado no Postgres-XL (versão 10r1.1), em conjunto com a extensão PostGIS (versão 2.5.0), foi estruturado a partir de dois tipos de nós, conforme ilustrado na Figura 5.4:

- i) Um nó mestre, responsável pela execução do serviço GTM *Master* que coordena as transações do *cluster*; e
- ii) nós de trabalho, cada um executando um serviço *datanode* e um serviço coordenador, ambos associados a *proxies* GTM próprios, que realizam a comunicação com o GTM *Master*. Qualquer coordenador pode responder a solicitações da aplicação e realizar consultas.

O modelo de dados foi implementado em uma tabela relacional, para a qual os registros provenientes dos *datasets* simulados foram inseridos. A aplicação cliente conecta-se a um dos coordenadores de maneira análoga ao processo utilizado no PostgreSQL, recorrendo a

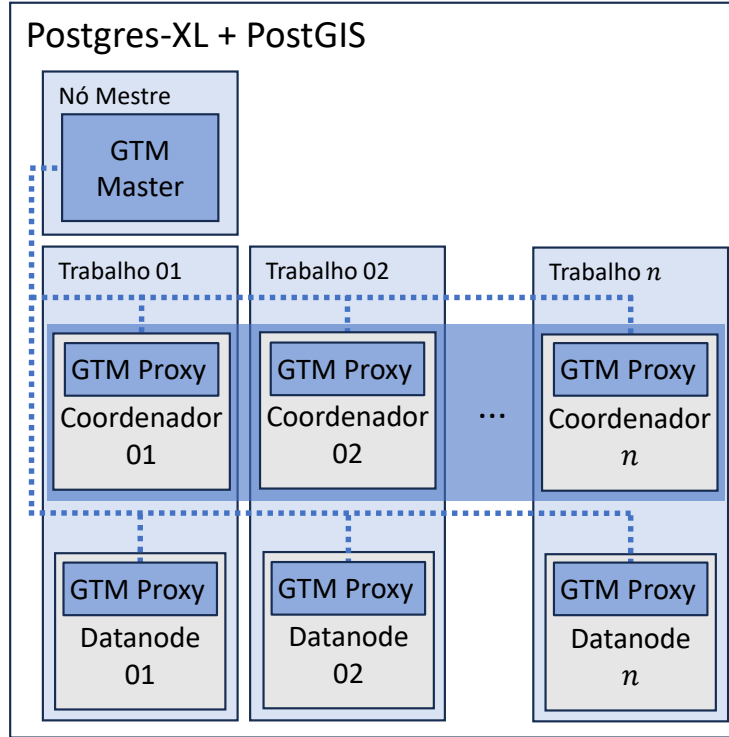


Figura 5.4: Ambiente 1 – Postgres-XL e PostGIS.

um *driver* JDBC convencional, por meio do qual são realizadas as operações de inserção e de consulta sobre o banco distribuído.

Ambiente 2 – Apache Sedona e HDFS

O ambiente baseado no Apache Sedona (versão 1.2.1), em conjunto com o HDFS (Hadoop 3.4.1) foi estruturado a partir de dois tipos de nós, conforme ilustrado na Figura 5.5:

- i) Um nó mestre, responsável pela execução do serviço Spark *Master* (responsável pelo gerenciamento dos recursos do *cluster* Spark), além dos serviços *Namenode* e *Resource Manager* do Apache Hadoop; e
- ii) nós de trabalho, cada um contendo um serviço *datanode* (responsável pelo armazenamento distribuído de dados) e um *node manager* (incumbido do gerenciamento de recursos locais do *datanode*) do Hadoop, ambos conectados ao *Resource Manager* e ao *Namenode*; e um serviço Spark *Worker* (onde são executadas as tarefas do *cluster* Spark), conectado ao Spark *Master*.

O modelo de dados foi implementado na forma de *dataframes*, carregados a partir de arquivos dos *datasets* simulados previamente armazenados no HDFS. Inicialmente, considerou-se o uso dos Spatial RDDs disponibilizados pelo Apache Sedona; contudo, essa

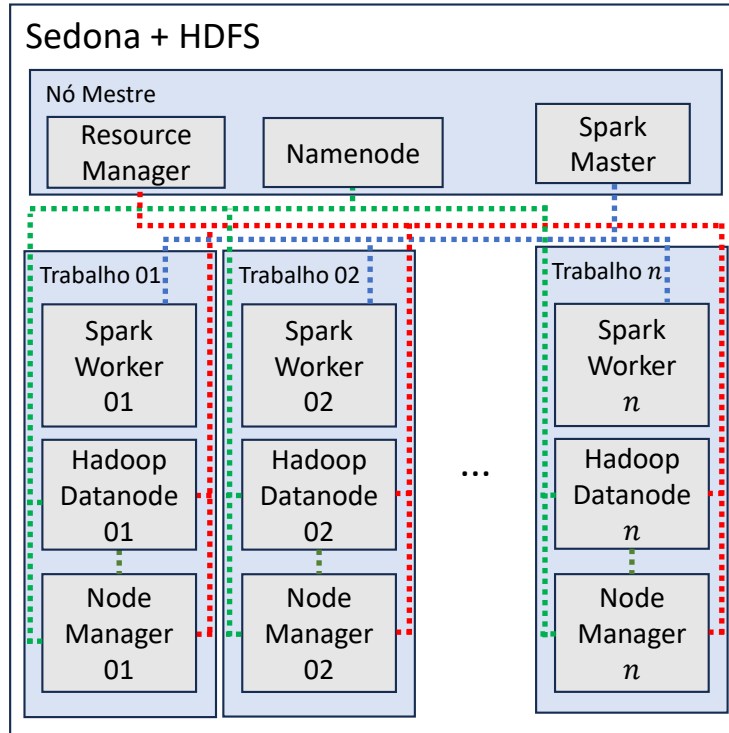


Figura 5.5: Ambiente 2 – Apache Sedona e HDFS.

abordagem limita-se à execução de operações espaciais específicas e não oferece suporte completo a consultas utilizando Spark SQL. Assim, optou-se pelos *dataframes* dentre os tipos de objetos gerenciados pelo Spark, em razão da natureza do *dataset* de entrada e da viabilidade de realizar consultas por meio do Spark SQL. Uma vez carregados, os registros são distribuídos entre os *hosts* do *cluster*, permitindo que as consultas sejam processadas por meio de varreduras completas nos *dataframes*, enquanto as operações nos dados são executadas diretamente na memória distribuída do Apache Spark.

5.3.2 Camada de Aplicação

Para a implementação da Camada de Aplicação, um protótipo foi desenvolvido com o intuito de visualizar os resultados das operações geoespaciais realizadas na arquitetura. A camada de Aplicação contempla dois aspectos principais:

- **Interface de consultas** – permite ao usuário inserir consultas diretamente ou carregar arquivos contendo conjuntos de consultas, oferecendo análise dos dados geoespaciais armazenados na arquitetura. No estado atual do protótipo, a Interface de consultas foi desenvolvida na linguagem de programação Java, e encontra-se implementada em interface de console.

- **Interface de geovisualização** – implementação realizada utilizando a GeoTools², uma biblioteca Java de código aberto para o processamento de dados geoespaciais. a aplicação apresenta visualmente os resultados das consultas, contendo as localizações de dispositivos móveis e estacionários, bem como outros elementos geoespaciais carregados a partir de *shapefiles*, como limites administrativos, ruas, avenidas e pontos de interesse. Todos os elementos são exibidos sobre o mapa com ícones personalizados, facilitando a identificação e interpretação dos dados. A Figura 5.6 ilustra um exemplo de geovisualização dos resultados de uma listagem de localizações de dispositivos móveis e estacionários em um determinado intervalo de tempo.

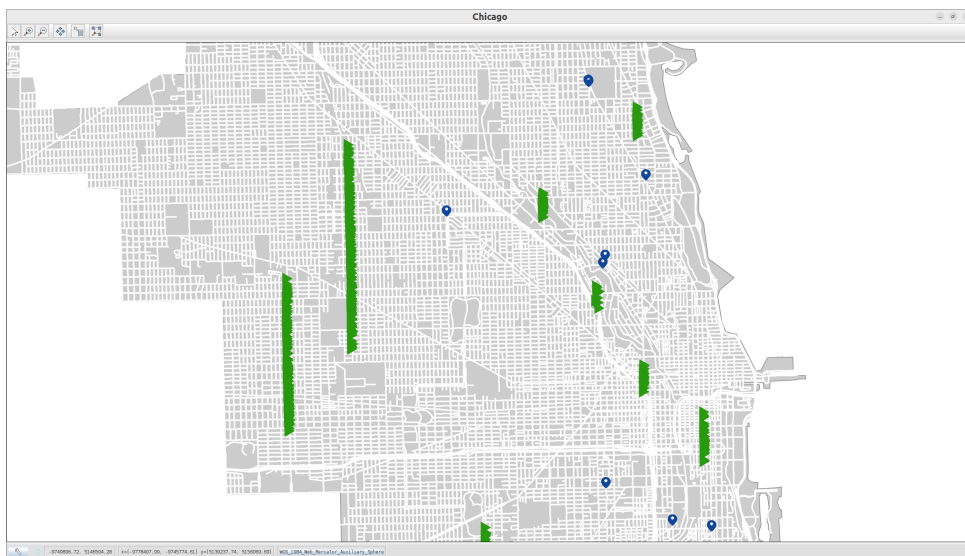


Figura 5.6: Visualização dos resultados de operação espacial.

A Camada de Aplicação conecta-se à camada de Armazenamento e Análise de Dados por meio de dois mecanismos distintos, de acordo com o ambiente considerado para o armazenamento e análise de dados. No Ambiente 1, as operações SQL são encaminhadas à camada por intermédio de um *driver* JDBC. No Ambiente 2, a integração ocorre por meio de uma *Spark Session*. Cabe destacar que ambas as interfaces encontram-se em estágio de protótipo, com funcionalidades básicas voltadas à validação do modelo e à análise experimental.

5.4 Considerações Finais

Este capítulo apresentou a arquitetura proposta para o gerenciamento e análise de *big data* geoespacial proveniente de dispositivos IoT com diferentes níveis de mobilidade. Ini-

²<https://geotools.org/>

cialmente, foram descritas as tecnologias utilizadas; em seguida, a arquitetura foi descrita, com a explicitação das funções e interações de suas camadas; por fim, discutiu-se a implementação, contemplando cada uma das camadas e destacando suas funcionalidades atuais.

Destaque-se que a arquitetura encontra-se em estágio de protótipo e que os ambientes apresentados para a camada de Armazenamento e Análise de Dados não interagem diretamente entre si. A implementação sustentou a fase experimental desta tese (Capítulo 6), permitindo avaliar o comportamento dos ambientes propostos em diferentes contextos de mobilidade, bem como em operações envolvendo dados geoespaciais e observações dos objetos.

Capítulo 6

Experimentos

Este capítulo apresenta o processo de experimentação conduzido nesta tese para validar a arquitetura proposta no Capítulo 5. Este capítulo encontra-se organizado conforme as seções a seguir: Na Seção 6.1 é apresentada uma contextualização inicial dos experimentos; na Seção 6.2, é descrita a preparação da infraestrutura empregada no estudo, abordando o processo de instalação e as configurações realizadas para adequá-la ao contexto do experimento; a Seção 6.3 descreve o processo de simulação utilizado na geração dos *datasets* com diferentes níveis de mobilidade empregados na etapa de experimentação; na Seção 6.4 são descritas as consultas utilizadas na experimentação; na Seção 6.5 são descritos os experimentos realizados, os resultados obtidos e considerações sobre suas implicações; a Seção 6.6 discute algumas limitações do trabalho, e as iniciativas tomadas para mitigá-las; a Seção 6.7 apresenta as ameaças à validade; a Seção 6.8 enumera as publicações realizadas ao longo da pesquisa; e, por fim, a Seção 6.9 aborda as considerações finais do capítulo.

6.1 Contextualização

Para a experimentação conduzida neste trabalho, foram implementados dois ambientes computacionais distintos para o tratamento de *big data* geoespacial na camada de Armazenamento e Análise de Dados da arquitetura proposta. O principal objetivo foi analisar e comparar o desempenho desses ambientes, considerando operações espaciais e não espaciais em diferentes circunstâncias de mobilidade dos dispositivos.

O desempenho dos ambientes foi avaliado sob diferentes aspectos. Primeiramente, analisou-se o uso do armazenamento, considerando a relação entre os distintos contextos de mobilidade dos *datasets* e o espaço ocupado em disco. Em seguida, investigou-se o desempenho das operações em cenários com diferentes padrões de mobilidade, bem como

a relação entre mobilidade e tipo de consulta, distinguindo entre consultas a dados de observação e, sobretudo, a consultas relacionadas à localização dos dispositivos.

Além disso, buscou-se avaliar o comportamento dos ambientes, com especial atenção ao comportamento diante de operações espaciais e não espaciais, de modo a verificar a capacidade de equilíbrio dos ambientes frente às diferentes naturezas dos dados processados. Outro aspecto avaliado foi o impacto do uso de índices sobre os resultados obtidos.

Para a experimentação, foram empregados *datasets* simulados — cujo método de geração é descrito na Seção 6.3 — baseados em dados reais de trajetórias e observações de sensores, representando diferentes contextos de mobilidade dos dispositivos.

6.2 Preparação da Infraestrutura Computacional

Os experimentos foram executados em um laboratório contendo 16 máquinas interconectadas em rede, com processadores Intel® Core™ i5-3470 CPU 2.90GHz (quad-core), 500 GB de armazenamento em disco rígido (HD), 4 GB de memória RAM e sistema operacional Ubuntu 22.04. Para adequar as máquinas às necessidades de cada experimento, foi utilizada a plataforma de containerização Docker¹. O ambiente de execução dos experimentos em *cluster* foi configurado por meio do Docker Swarm (ferramenta de orquestração que permite gerenciar e distribuir contêineres em um conjunto de máquinas), distribuído entre os 16 nós disponíveis. A organização da infraestrutura de hardware e software utilizada na avaliação das ferramentas envolveu as seguintes etapas:

- Geração das imagens Docker: Os serviços a serem instalados nas máquinas foram implementados em imagens Docker contendo todas as ferramentas e bibliotecas necessárias para atender as especificidades de cada experimento, permitindo a criação de *clusters* de Postgres-XL suporte à extensão PostGIS, Apache Hadoop (para aplicação do sistema de arquivos HDFS) e Apache Spark (para utilização do Apache Sedona).
- Instalação e configuração das ferramentas: configuração da infraestrutura necessária para a realização dos experimentos, incluindo preparações na rede e nas máquinas, com instalação do Docker para uso do Swarm e das imagens necessárias aos experimentos.

Os serviços empregados em cada ambiente computacional (Seção 5.3) foram organizados da seguinte forma:

¹<https://www.docker.com/>

- **Ambiente 1 – Postgres-XL e PostGIS:** composta de um nó mestre, para o serviço GTM Master, e quinze nós de trabalho, cada qual contendo um serviço *Datanode* e um serviço coordenador, sendo cada serviço associado a um *proxy* GTM.
- **Ambiente 2 – Apache Sedona e HDFS:** composta de um nó mestre, para o serviço Spark Master e o Hadoop Namenode, e quinze nós de trabalho, cada qual contendo um serviço de *Datanode* do Hadoop e um Spark *Worker*.

6.3 Geração dos Dados

Para representar os dados dos dispositivos IoT na experimentação, inicialmente procurou-se na literatura *datasets* abertos que representassem dispositivos móveis e estacionários dotados de sensores gerando observações do ambiente. Entretanto, não foi encontrado nenhum conjunto de dados com essas características, motivo pelo qual optou-se pela realização de uma simulação.

Inicialmente, foram consideradas algumas ferramentas para automatizar essa tarefa, como o SUMO² (*Simulation of Urban MObility*) e o Vissim³, ambos simuladores voltados para dados de tráfego. Contudo, nenhuma das ferramentas investigadas foi capaz de integrar, de forma conjunta, a simulação da movimentação dos dispositivos e das observações dos sensores ao longo do tempo.

Diante disso, foi desenvolvida uma ferramenta própria utilizando a linguagem de programação Java, permitindo simular tanto as localizações geográficas (latitude e longitude) quanto as observações realizadas no ambiente, considerando dispositivos IoT móveis e estacionários, com base em *datasets* reais de mobilidade e de dados de sensores.

A ferramenta gerou três *datasets*, cada um representando um cenário de mobilidade. O primeiro (mobilidade alta) envolve apenas dispositivos móveis, que geram novas informações de localização geoespacial em intervalos regulares. O segundo (mobilidade intermediária) combina tanto dispositivos móveis quanto estacionários. O terceiro (mobilidade baixa) é composto apenas de dispositivos estacionários, com localização geoespacial fixa. Nos três cenários, cada dispositivo simulado possui um sensor que gera observações em intervalos regulares.

A simulação baseou-se em *datasets* reais oriundos de projetos desenvolvidos na cidade de Chicago, Illinois, EUA. Para tal, foram empregados dois tipos de *datasets*:

- Arquivos de rotas: utilizados para representar as trajetórias dos dispositivos móveis. Foram empregados os *datasets* *Bike Routes* [269], que reúne diferentes rotas regis-

²<https://eclipse.dev/sumo/>

³<https://www.ptvgroup.com/pt-br/produtos/ptv-vissim>

tradas por bicicletas, e *Chicago Campus Bus GPS trajectory data* [270], contendo trajetórias de ônibus.

- Arquivos de observações: extraídos do projeto *Array of Things* (AoT) [271], composto por dados de sensores estacionários de múltiplos tipos. Dentre os arquivos disponibilizados, três foram empregados na simulação: *data.csv*, contendo os valores das observações; *nodes.csv*, a lista de nodes (nós registrados) do *dataset* e seus metadados; e *sensors.csv*, a lista de sensores e seus metadados. Para a experimentação, foram selecionados sensores de umidade, temperatura e pressão, cujas características estão descritas na Tabela 6.1.

Tabela 6.1: Sensores do projeto *Array of Things* utilizados na simulação.

Subsistema	Sensor	Parâmetro	Unidade	Valor min.	Valor max.
metsense	bmp180	temperatura	C	-40	85
metsense	htu21d	temperatura	C	-40	125
metsense	pr103j2	temperatura	C	-55	80
metsense	tmp112	temperatura	C	-40	125
metsense	tsys01	temperatura	C	-40	125
metsense	hih4030	umidade	RH	0	100
metsense	htu21d	umidade	RH	0	100
lightsense	hih6130	umidade	RH	0	100
wagman	htu21d	umidade	RH	0	100
chemsense	sht25	umidade	RH	0	100
chemsense	lps25h	pressão	hPa	260	1260
metsense	bmp180	pressão	hPa	300	1100

Conforme dito anteriormente, foram simulados dispositivos móveis e estacionários. Para a posição dos dispositivos estacionários, era escolhida uma localização adjacente a algum sensor do *dataset* de observações. Para cada dispositivo móvel, era atribuída uma trajetória dentre as registradas nos arquivos de observações. A localização inicial do dispositivo era o ponto inicial da rota, e era simulado um deslocamento ao longo da rota, com velocidade fixa pré-definida.

Para a simulação das observações, foi atribuído a cada dispositivo um tipo de sensor gerando observações em intervalos regulares, com base nos valores dos sensores do *dataset* AoT. Para o cálculo desse valor, foi empregado o algoritmo *Inverse Distance Weighting* (IDW) [272], um método de interpolação espacial que estima valores desconhecidos para um determinado ponto com base em valores da mesma variável observados em pontos vizinhos, atribuindo pesos inversamente proporcionais à distância. Para calcular o valor atual de um sensor, são utilizadas como entradas a posição do dispositivo e a localização

e o valor atual observado por sensores do mesmo tipo. A saída é o valor estimado para o sensor, na localização do dispositivo. Os *datasets* gerados simulam 3 anos de observações (janeiro de 2018 a janeiro de 2021). A Tabela 6.2 apresenta o esquema final dos arquivos, gerados no formato .csv, e detalha a origem de cada coluna do *dataset*. Para colunas extraídas de arquivos do AoT, a tabela especifica de quais arquivos (*nodes*, *sensors* ou *data*) os dados foram obtidos, e indica quais colunas desses arquivos foram utilizadas. Já as colunas que foram geradas na simulação estão indicadas com um asterisco (*) na coluna correspondente.

Tabela 6.2: Formato final dos arquivos .csv.

Atributo	Simul.	Arquivo de origem (AoT)			Tipo
		nodes	sensors	data	
idRecord	*				int
idEntity	*				int
entityName	*				String
entityDescription	*				String
idLocation	*				int
location		lat, lon			Coordinate
locationTime	*				LocalDateTime
idObservation	*				int
observationTime	*				LocalDateTime
observationValue				value_hrf	String
idObsProperty	*				int
obsPropertyName			parameter		String
obsPropertyDescription			datasheet		String
obsPropertyUnitOfMeasurement			hrf_unit		String

6.4 Elaboração das Consultas

Para avaliar o desempenho da arquitetura proposta, foram elaboradas doze consultas, organizadas em três categorias, de modo a contemplar tanto os dados de localização quanto as observações coletadas pelos dispositivos, permitindo uma análise abrangente do comportamento do sistema. As categorias definidas foram:

- i) Consultas espaciais (Consultas 1 a 4), com operadores espaciais básicos aplicados aos dados de localização (Tabela 6.3);
- ii) Consultas de observações (Consultas 5 a 8), envolvendo dados coletados pelos dispositivos no ambiente, com operações de filtragem por valor e agregações por dispositivos (Tabela 6.4); e

- iii) Consultas mistas (Consultas 9 a 12), que combinam os filtros de localização e de observações das duas primeiras categorias (Tabela 6.5).

Tabela 6.3: Consultas espaciais.

Consulta 1 - Localizações de um dispositivo entre os instantes y e z. <pre> SELECT ST_AsText(location) FROM tese WHERE entityId = ? AND observationTime BETWEEN 'yyyy-MM-dd HH:mm:ss.SSS' AND 'yyyy-MM-dd HH:mm:ss.SSS' </pre>
Consulta 2 - Dispositivos que estiveram na região delimitada por um retângulo. <pre> SELECT DISTINCT entityId FROM tese WHERE ST_Within(location, ST_MakeEnvelope(x1, y1, x2, y2, 4326)) </pre>
Consulta 3 - Observações realizadas a até 600 metros de um ponto. <pre> SELECT ST_AsText(location) AS location FROM tese WHERE ST_DWithin(location, ST_GeogFromText('SRID=4326;POINT(x, y)'), 600) </pre>
Consulta 4 - Três observações mais próximas de determinado ponto (<i>kNN query</i>). <pre> SELECT idObservation, ST_AsText(location) AS location, ST_Distance(location, ref_geom) AS distance FROM tese, LATERAL ST_GeogFromText('SRID=4326; POINT (x, y)') AS ref_geom ORDER BY location <-> ref_geom LIMIT 3 </pre>

As consultas espaciais implementadas são baseadas nas operações básicas propostas por Eldawy [39] (Seção 2.2), como *Range queries (bounding box)* nas Consultas 2 e 10, *proximity query* nas Consultas 3 e 11, e *Nearest Neighbor (NN) queries* nas Consultas 4 e 12. Nos dados de observações, além de filtragem por valor (Consultas 8, 10 e 11), também foram analisadas agregações (Consultas 5, 6, 7 e 9). Ao avaliar tanto consultas espaciais quanto de observações, busca-se analisar o equilíbrio das ferramentas, verificando o comportamento da arquitetura frente aos dados espaciais sem desconsiderar o desempenho com os dados de observação coletados pelos sensores.

Tabela 6.4: Consultas de observações.

Consulta 5 - Valores foram coletados por determinado dispositivo em um intervalo de tempo? <pre> SELECT COUNT(observationValue) FROM tese WHERE entityId = 0 AND observationTime BETWEEN 'yyyy-MM-dd HH:mm:ss.SSS' AND 'yyyy-MM-dd HH:mm:ss.SSS' </pre>
Consulta 6 - Média dos valores para uma propriedade em um intervalo de tempo? <pre> SELECT AVG(CAST(observationValue AS double precision)) AS media FROM tese WHERE entityId = 0 AND observationTime BETWEEN 'yyyy-MM-dd HH:mm:ss.SSS' AND 'yyyy-MM-dd HH:mm:ss.SSS' AND obsPropertyName = 'pressure' </pre>
Consulta 7 - Menor registro de umidade <pre> WITH min_humidity AS (SELECT MIN(CAST(observationValue AS double precision)) AS min_value FROM tese WHERE obsPropertyName = 'humidity') SELECT observationValue FROM tese, min_humidity WHERE obsPropertyName = 'humidity' AND ABS(CAST(observationValue AS double precision) - min_humidity.min_value) < 1e-8 </pre>
Consulta 8 - Registros de pressão superior a 2200 hPa <pre> SELECT idRecord, observationValue FROM tese WHERE obsPropertyName = 'pressure' AND CAST(observationValue AS double precision) >= 2200 </pre>

6.5 Avaliação dos Ambientes

A avaliação dos ambientes computacionais propostos para a camada de Armazenamento e Análise de Dados considerou dois fatores: o espaço de armazenamento utilizado, e o desempenho na recuperação dos dados (tempo de realização das consultas, em milissegundos). Os três *datasets* gerados foram analisados nas diversas combinações de ferramentas sob avaliação, conforme os passos descritos a seguir.

Tabela 6.5: Consultas mistas.

Consulta 9 - Em que localização determinado dispositivo encontrou o maior valor para determinada propriedade observada? <pre> WITH max_pressure AS (SELECT MAX(CAST(observationValue AS double precision)) AS max_value FROM tese WHERE entityId = 0 AND obsPropertyName = 'pressure') SELECT ST_AsText(location) AS location FROM tese, max_pressure WHERE entityId = 0 AND obsPropertyName = 'pressure' AND ABS(CAST(observationValue AS double precision) - max_pressure.max_value) < 1e-8 </pre>
Consulta 10 - Localização das observações de pressão que ultrapassaram 1020 hPa na região delimitada por um retângulo [x1,y1,x2,y2]. <pre> SELECT ST_AsText(location) AS geolocation, observationValue FROM tese WHERE entityId = 0 AND obsPropertyName = 'pressure' AND CAST(observationValue AS double precision) >= 1020 AND ST_Within(location, ST_MakeEnvelope(x1, y1, x2, y2, 4326)) </pre>
Consulta 11 - Registros de temperatura superior a 40°C realizados a até 20 metros do ponto [x, y] (<i>range query</i>). <pre> SELECT ST_AsText(location) AS location, observationValue FROM tese WHERE ST_DWithin(location, ST_GeogFromText('SRID=4326;POINT(x, y)'), 600) AND obsPropertyName = 'temperature' AND CAST(observationValue AS double precision) >= 40 </pre>
Consulta 12 - Três registros de umidade mais próximos a um determinado ponto (<i>kNN query</i>). <pre> SELECT idObservation, ST_AsText(location) AS location, ST_Distance(location, ref_geom) AS distance, observationValue FROM tese, LATERAL ST_GeogFromText('SRID=4326;POINT(x, y)') AS ref_geom WHERE obsPropertyName = 'humidity' AND CAST(observationValue AS double precision) >= 20 ORDER BY location <-> ref_geom LIMIT 3 </pre>

Para cada ambiente computacional avaliado, cada *dataset* foi carregado individualmente, registrando-se o espaço ocupado em disco. Em seguida, as consultas definidas foram executadas 10 vezes cada, registrando-se todos os tempos de execução, a partir dos quais foram calculados os valores máximo, mínimo, médio e desvio-padrão. Concluída essa etapa, o banco de dados era apagado, garantindo um ambiente limpo para a repetição do procedimento com o *dataset* seguinte, até que todos fossem processados.

Os resultados obtidos foram compilados e comparados, permitindo identificar quais ferramentas apresentaram os melhores resultados (menores médias dos tempos de execução) em diferentes cenários de mobilidade dos dispositivos, além de observar a relação entre registros de localização e de observações.

Os passos descritos foram adotados na condução de dois experimentos: o Experimento 1 teve como objetivo avaliar o desempenho da camada de Armazenamento e Análise de Dados frente a um grande volume de registros, enquanto o Experimento 2 buscou analisar os possíveis ganhos proporcionados pelo uso de índices espaciais. Os experimentos estão descritos nas subseções a seguir.

6.5.1 Experimento 1

No Experimento 1, verificou-se o funcionamento da camada de Armazenamento e Análise de Dados diante dos três *datasets* com diferentes níveis de mobilidade. A Tabela 6.6 apresenta as características dos *datasets* simulados utilizados no Experimento 1, com diferentes níveis de mobilidade (alta, intermediária e baixa). Para cada *dataset*, são mostrados o tamanho do arquivo original e o espaço ocupado nos dois ambientes avaliados: o Ambiente 1, baseado em Postgres-XL com PostGIS, e o Ambiente 2, utilizando Apache Sedona em conjunto com HDFS. Observa-se que, embora o tamanho original dos *datasets* seja aproximadamente o mesmo (≈ 200 GB), a ocupação de espaço nos ambientes distribuídos, após a alimentação dos dados, é significativamente maior, refletindo a sobrecarga de armazenamento, a replicação de dados e a organização interna de cada sistema.

Tabela 6.6: Características dos *datasets* simulados - Experimento 1.

Dataset	Tamanho		
	Arquivo original	Ambiente 1 (Postgres-XL e PostGIS)	Ambiente 2 (Apache Sedona e HDFS)
Mobilidade alta	≈ 200 GB	≈ 454 GB	≈ 618 GB
Mobilidade intermediária	≈ 200 GB	≈ 450 GB	≈ 617 GB
Mobilidade baixa	≈ 200 GB	≈ 447 GB	≈ 617 GB

Após a ingestão dos dados pelas ferramentas sob análise, as 12 consultas (C1 a C12) geradas foram executadas 10 vezes cada uma. Na análise dos resultados, foram descon-

siderados os valores extremos (maiores e menores tempos de execução observados para cada consulta), de modo a reduzir a influência de picos ocasionais. Com base nesses dados processados, os tempos (maior, menor e média) em milissegundos e o desvio padrão (DP) nos dois ambientes para os *datasets* de mobilidade alta, intermediária e baixa estão representados nas Tabelas 6.7, 6.8 e 6.9, respectivamente. As células em negrito indicam, para cada consulta, o menor tempo de execução e a menor média de tempo obtidos, considerando ambos os ambientes avaliados.

Tabela 6.7: Experimento 1: Resultados (tempo em milissegundos) para o *dataset* de mobilidade alta.

	Ambiente 1 (Postgres-XL e PostGIS)				Ambiente 2 (Apache Sedona e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	158.673	304.547	136.479	55.179	464.443	472.818	459.435	4.466
C2	751.020	827.336	584.352	95.449	509.346	513.989	505.248	2.722
C3	149.077	243.725	134.231	35.867	505.664	528.986	489.178	14.514
C4	1.749.431	2.100.064	1.562.442	193.106	491.248	508.219	480.127	8.538
C5	156.325	158.611	155.544	916	472.863	484.449	457.009	8.177
C6	155.751	156.275	155.455	243	469.867	476.360	460.128	5.909
C7	384.678	390.972	380.073	3.752	959.022	980.273	936.365	14.584
C8	155.906	157.075	155.421	473	473.515	485.410	459.248	8.833
C9	333.857	347.080	330.469	5.274	941.735	960.748	923.632	12.854
C10	155.999	156.283	155.725	188	468.792	476.945	457.616	6.346
C11	155.995	156.464	155.759	246	486.963	500.757	477.862	8.409
C12	506.983	528.285	490.611	13.916	490.351	499.589	484.376	5.496

Tabela 6.8: Experimento 1: Resultados (tempo em milissegundos) para o *dataset* de mobilidade intermediária.

	Ambiente 1 (Postgres-XL e PostGIS)				Ambiente 2 (Apache Sedona e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	276.720	283.326	268.814	4.928	474.256	492.355	465.792	9.422
C2	641.322	733.189	542.838	78.237	514.096	524.513	492.192	9.371
C3	161.401	162.638	159.759	942	501.818	511.737	482.025	8.259
C4	1.572.334	1.615.230	1.539.564	23.243	511.155	520.309	503.055	5.344
C5	146.653	147.281	146.257	312	475.635	491.970	466.330	8.874
C6	146.995	148.001	145.933	617	480.526	493.380	467.389	9.154
C7	376.463	386.092	373.266	3.867	962.256	982.866	942.614	10.912
C8	146.718	147.176	146.302	339	472.982	482.578	461.985	7.746
C9	317.005	321.690	315.464	1.992	953.571	979.263	928.780	16.214
C10	146.534	147.234	146.018	435	484.815	492.178	478.899	4.369
C11	146.437	146.906	146.000	260	496.218	519.108	479.576	9.372
C12	512.543	527.795	498.230	10.075	495.178	509.913	471.717	12.764

Tabela 6.9: Experimento 1: Resultados (tempo em milissegundos) para o *dataset* de mobilidade baixa.

	Ambiente 1 (Postgres-XL e PostGIS)				Ambiente 2 (Apache Sedona e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	303.772	333.832	293.525	13.532	479.424	496.062	468.318	9.645
C2	567.014	679.466	471.310	89.887	519.075	533.454	504.071	9.328
C3	145.502	151.702	144.267	2.362	511.329	522.192	503.783	5.803
C4	1.571.155	1.591.108	1.556.300	11.603	519.337	532.866	510.894	7.806
C5	144.245	144.637	143.782	249	478.001	492.401	460.458	10.443
C6	143.942	144.108	143.739	129	478.951	493.972	458.338	10.306
C7	883.862	928.826	843.294	31.241	955.476	977.058	943.004	9.782
C8	153.510	154.370	153.168	384	481.575	495.143	469.332	10.024
C9	308.822	312.383	307.649	1.467	958.502	972.549	939.255	12.969
C10	144.821	145.540	144.183	409	476.827	487.245	464.748	8.344
C11	148.355	172.338	144.263	9.072	489.712	512.173	465.805	14.030
C12	530.289	578.007	515.029	18.668	497.087	509.105	482.187	8.318

As Figuras 6.1 e 6.2 apresentam os resultados das Consultas 1 a 12 do Experimento 1, que ilustram o desempenho dos *datasets* de mobilidade alta (A), intermediária (I) e baixa (B) no Ambiente 1, com Postgres-XL (PGXL), e no Ambiente 2, com Apache Sedona (Sed). Nos gráficos, é possível visualizar a variabilidade dos resultados e a presença de *outliers* nos tempos de execução, além de facilitar a comparação entre cenários.

Na média dos tempos de execução, o Ambiente 1 (Postgres-XL) se destacou nas consultas sobre observações dos dispositivos, agregações, e consultas envolvendo subconsultas, obtendo melhor desempenho (ou seja, obteve a menor média nos tempos de execução, e o menor registro de tempo) nas Consultas 1, 3, 5, 6, 7, 8, 9, 10 e 11. O Ambiente 2 teve melhor desempenho em algumas operações com maior ênfase nos aspectos espaciais, como as Consultas 2, 4 e 12.

Para a Consulta 1, o Postgres-XL destacou-se pelo menor tempo de execução e pelo tempo médio mais baixo. Isso pode ser devido ao fato de que, embora a busca envolvesse um atributo espacial, o filtro tenha sido aplicado em um atributo do tipo inteiro e outro do tipo *timestamp*. Na média dos tempos de execução, os resultados obtidos para os *datasets* com mobilidade alta (Tabela 6.7) e intermediária (Tabela 6.8) foram melhores do que os do *dataset* de baixa mobilidade (Tabela 6.9).

As Consultas 2 e 10 empregam um filtro espacial (*ST_Within*). Na Consulta 2 o Apache Sedona teve melhor desempenho em todos os contextos analisados, sendo que o Postgres-XL teve os piores tempos médios nos *datasets* de alta mobilidade. Já na consulta 10, envolvendo além do filtro espacial um filtro de valor em um campo específico, o Postgres-XL se saiu melhor.

As Consultas 3 e 11 empregam um filtro espacial (`ST_DWithin`), e o Postgres-XL apresentou melhor desempenho para o menor tempo e tempo médio, apesar de, na variação, ter tido o pior (maior) registro de tempo de uma consulta. A Consulta 4 (kNN) busca os três registros mais próximos de um ponto de referência. O Postgres-XL teve desempenho bastante inferior ao do Apache Sedona, devido a um *loop* aninhado. Já na Consulta 12, apesar de o Apache Sedona ainda ter se saído melhor, os filtros não espaciais empregados ajudaram a reduzir a diferença.

6.5.2 Experimento 2

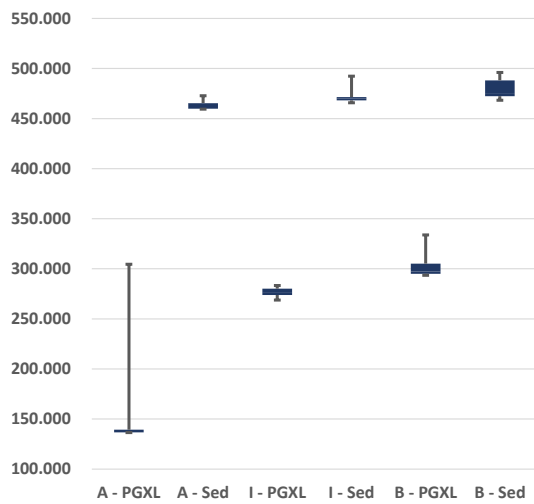
No Experimento 2, o objetivo foi avaliar o efeito da indexação espacial nas consultas realizadas. Para viabilizar sua execução, utilizou-se uma versão reduzida dos *datasets* originais, dado que o hardware disponível não apresentou capacidade suficiente para indexar o volume completo de dados dos *datasets* do Experimento 1.

Além dos ambientes computacionais já experimentados para o Postgres-XL e Apache Sedona, testou-se o uso de índices espaciais. No Postgres-XL foi aplicado o índice GiST [273], índice espacial utilizado pelo PostGIS, e no Apache Sedona foi empregado o índice R-Tree [97], utilizado em vários trabalhos mapeados na literatura sobre dados geoespaciais de IoT. Em ambas as ferramentas, o índice foi aplicado no atributo espacial (*location*).

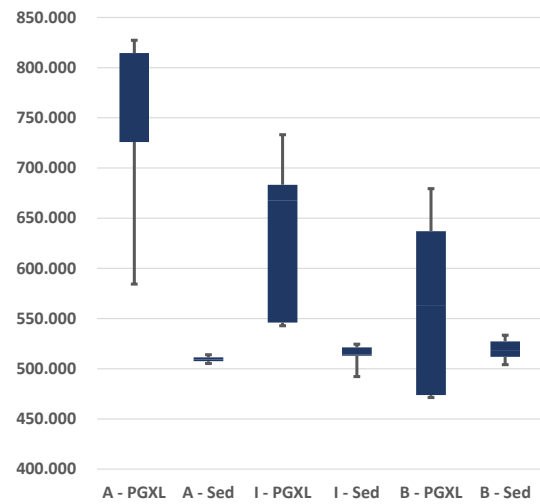
Entretanto, em testes preliminares, verificou-se que a criação do índice GiST no Postgres-XL exigiu tempo elevado e provocou aumento significativo nos tempos de execução das consultas em relação ao cenário sem índice, tornando inviável a sua utilização no experimento. Ainda assim, os resultados do Postgres-XL sem índice foram registrados para servir de referência na comparação com os resultados obtidos pelo Apache Sedona, tanto com quanto sem índices.

Nesse Experimento, também foram avaliados *datasets* com diferentes níveis de mobilidade (alta, intermediária e baixa), porém com volume reduzido em comparação ao Experimento 1. A Tabela 6.10 apresenta as características desses *datasets* simulados. Para cada conjunto de dados, são indicados o tamanho do arquivo original e o espaço ocupado nos dois ambientes avaliados: o Ambiente 1, baseado em Postgres-XL com PostGIS, e o Ambiente 2, utilizando Apache Sedona em conjunto com HDFS.

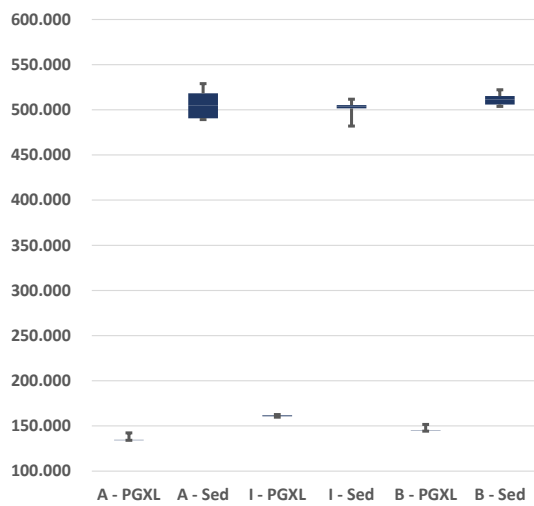
De forma análoga ao Experimento 1, as 12 consultas foram executadas 10 vezes cada, em cada uma das três situações avaliadas – Ambiente 1, com o Postgres-XL sem indexação; Ambiente 2, com o Apache Sedona sem indexação; e Ambiente 2, com indexação. Os tempos das consultas (maior, menor e médio) em milissegundos e o desvio padrão (DP) nos três ambientes para os *datasets* de mobilidade alta, intermediária e baixa estão representados nas Tabelas 6.11 a 6.14. Novamente, as células em negrito indicam, para cada



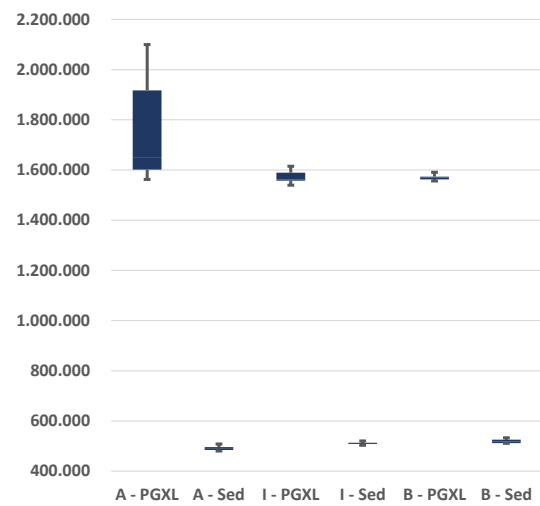
(a) Consulta 1



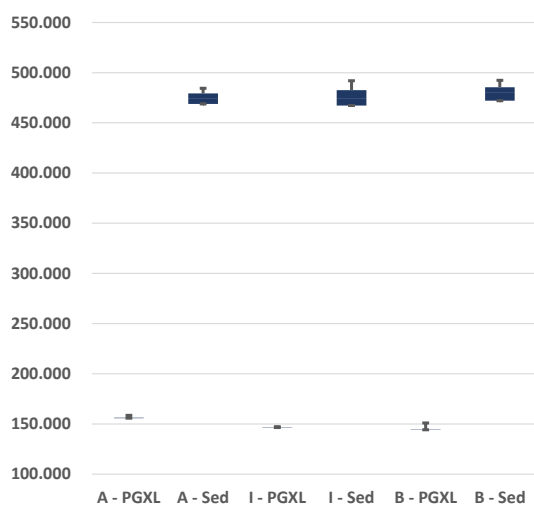
(b) Consulta 2



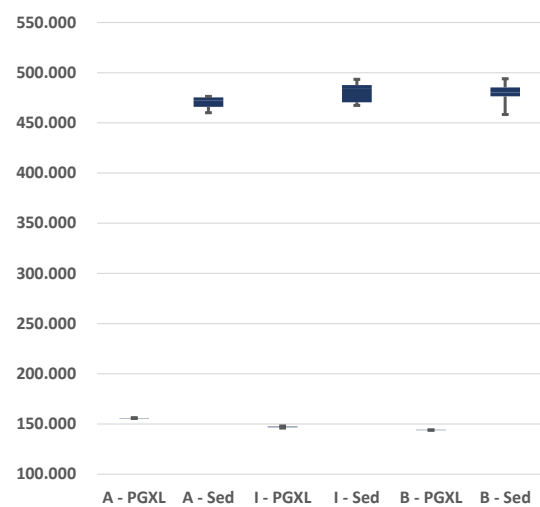
(c) Consulta 3



(d) Consulta 4

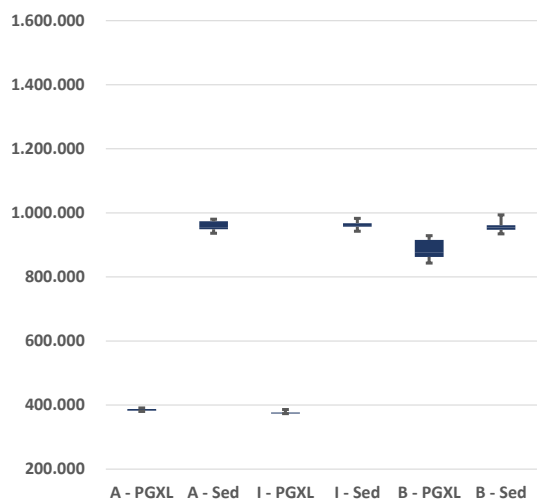


(e) Consulta 5

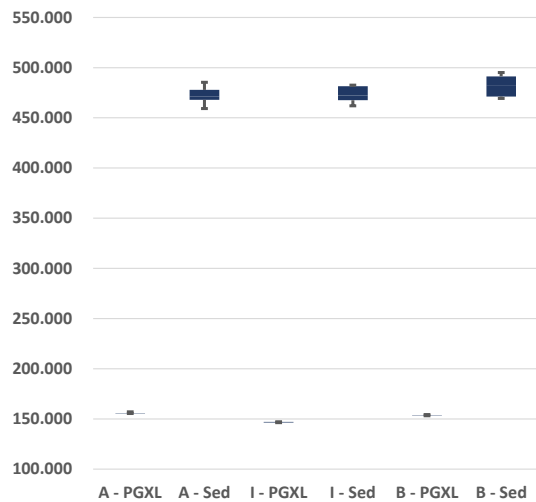


(f) Consulta 6

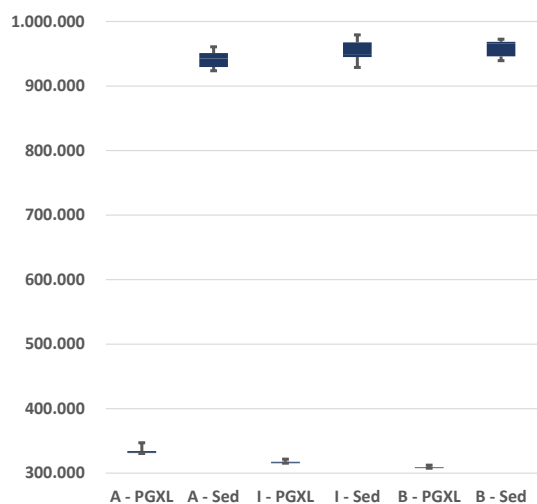
Figura 6.1: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 1.



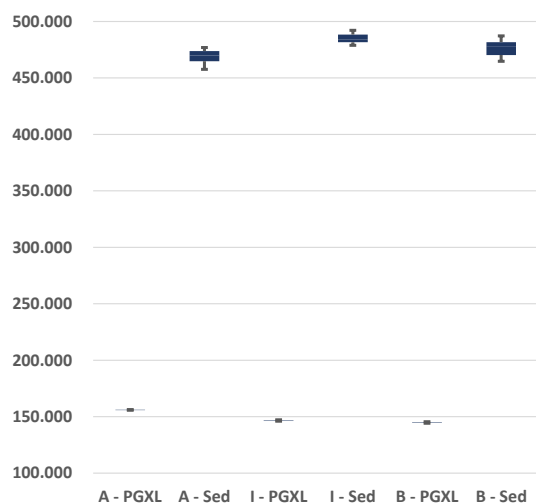
(a) Consulta 7



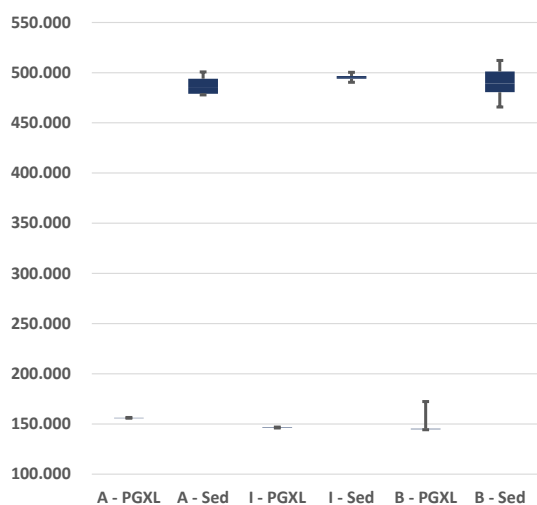
(b) Consulta 8



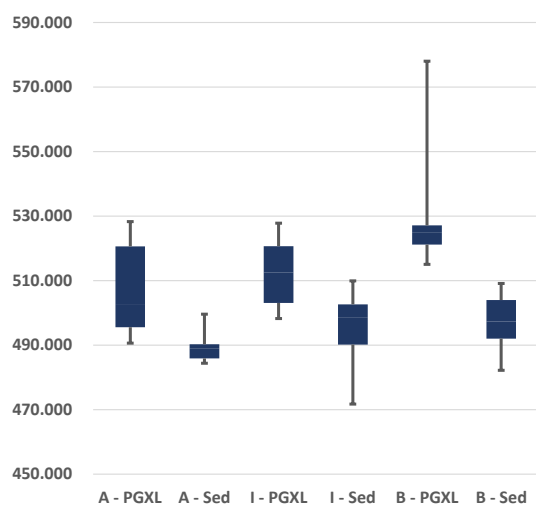
(c) Consulta 9



(d) Consulta 10



(e) Consulta 11



(f) Consulta 12

Figura 6.2: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 1 (continuação).

Tabela 6.10: Características dos *datasets* simulados - Experimento 2.

Dataset	Tamanho		
	Arquivo original	Ambiente 1 (Postgres-XL e PostGIS)	Ambiente 2 (Apache Sedona e HDFS)
Mobilidade alta	≈ 100 GB	≈ 217 GB	≈ 280 GB
Mobilidade intermediária	≈ 100 GB	≈ 219 GB	≈ 280 GB
Mobilidade baixa	≈ 100 GB	≈ 221 GB	≈ 280 GB

consulta, o menor tempo de execução e a menor média de tempo obtidos, considerando todos os ambientes avaliados.

Tabela 6.11: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de alta mobilidade (Postgres-XL).

	Ambiente 1 (Postgres-XL e PostGIS)			
	Média	Maior	Menor	DP
C1	277.848	281.547	273.439	2.234
C2	280.588	283.423	279.382	1.219
C3	71.325	71.445	71.260	61
C4	731.462	746.139	714.377	11.030
C5	71.266	71.626	71.087	174
C6	71.337	71.575	71.200	133
C7	173.877	174.615	173.070	530
C8	71.481	71.781	71.202	228
C9	147.769	148.225	147.418	239
C10	71.340	71.498	71.164	103
C11	71.407	71.801	71.141	240
C12	227.423	233.299	222.357	3.572

As Figuras 6.3 a 6.6 apresentam os resultados das Consultas 1 a 12 para o Experimento 2, com as caixas representando os resultados para os *datasets* de mobilidade alta (A), intermediária (I) e baixa (B) para o Ambiente 1, com o Postgres-XL (PGXL), o Ambiente 2, com o Apache Sedona (Sed), e novamente o Ambiente 2, mas com o Sedona utilizando índices (Sed (i)).

No Experimento 2, o Apache Sedona apresentou melhor desempenho na Consulta 1, a busca por localizações. Assim como no Experimento 1, teve um desempenho superior nas Consultas 2 e 4. A Consulta 4 (kNN) apresentou o maior tempo médio de execução no Postgres-XL, mas ao contrário do Experimento 1, sua versão com um filtro não espacial (Consulta 12) apresentou desempenho melhor que no Apache Sedona, com ou sem utili-

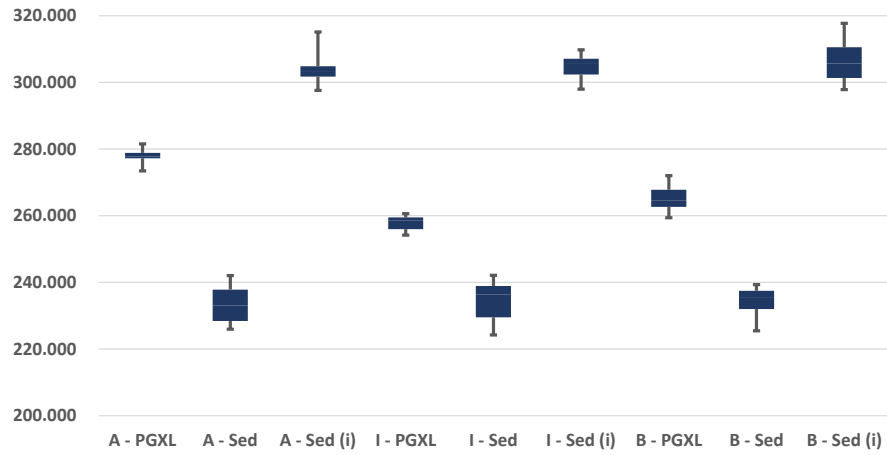
Tabela 6.12: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de alta mobilidade (Apache Sedona).

	Ambiente 2 (Apache Sedona e HDFS)				Ambiente 2 (Apache Sedona (indexado) e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	233.408	242.046	225.955	5.431	304.701	315.145	297.596	5.121
C2	253.200	260.137	236.671	7.122	339.998	348.164	334.987	4.308
C3	244.696	253.890	229.969	8.399	325.878	331.215	318.942	4.021
C4	249.668	264.515	235.216	7.776	327.230	333.680	320.262	4.403
C5	233.171	241.032	224.521	6.311	302.329	310.823	294.733	5.328
C6	233.692	239.411	226.665	4.056	302.511	310.647	296.104	4.517
C7	471.254	486.442	462.465	6.788	603.293	609.417	593.108	4.602
C8	236.681	248.411	227.936	6.860	303.159	310.623	298.865	3.455
C9	466.934	488.867	451.263	10.975	604.749	612.725	596.168	6.097
C10	233.204	247.145	221.343	8.479	311.296	319.156	302.578	5.124
C11	233.449	245.096	223.792	7.515	324.029	328.686	317.853	2.831
C12	238.824	248.380	232.123	4.820	308.778	318.461	297.836	6.331

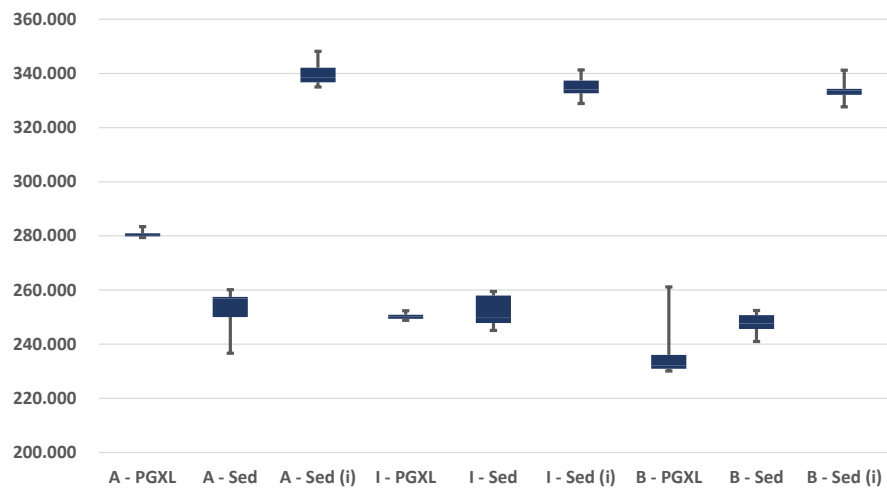
Tabela 6.13: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de mobilidade média (Postgres-XL).

	Ambiente 1 (Postgres-XL e PostGIS)			
	Média	Maior	Menor	DP
C1	257.764	260.632	254.169	2.198
C2	250.269	252.357	248.820	1.096
C3	72.165	72.504	71.938	177
C4	734.242	761.411	720.433	12.628
C5	71.915	72.218	71.682	201
C6	71.864	72.162	71.714	184
C7	175.700	177.028	174.361	845
C8	72.281	72.681	71.934	281
C9	152.100	152.679	151.757	314
C10	73.229	77.825	71.955	1.874
C11	72.310	72.604	72.048	180
C12	231.106	236.022	228.018	3.037

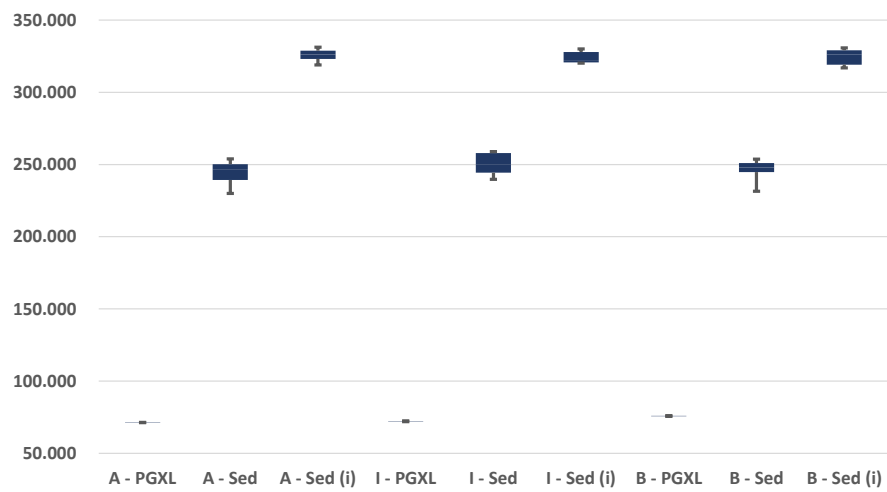
zação do índice. O Postgres-XL também apresentou o melhor desempenho para todas as consultas envolvendo filtros não espaciais.



(a) Consulta 1

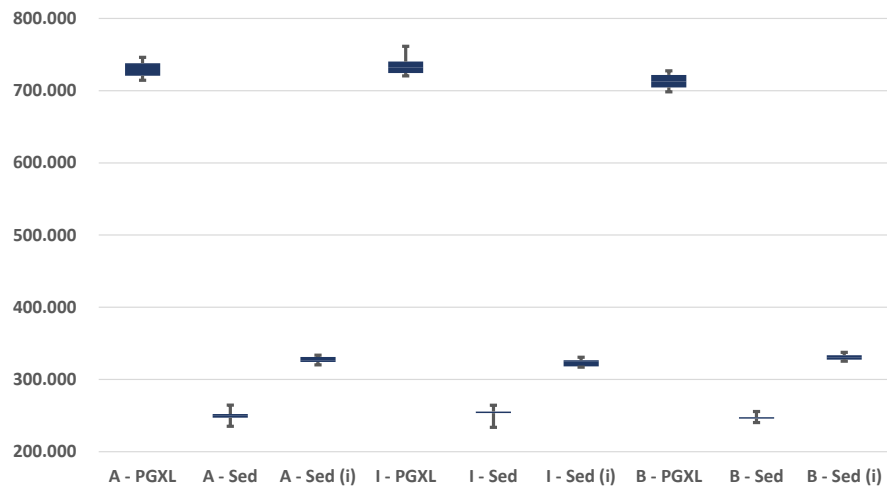


(b) Consulta 2

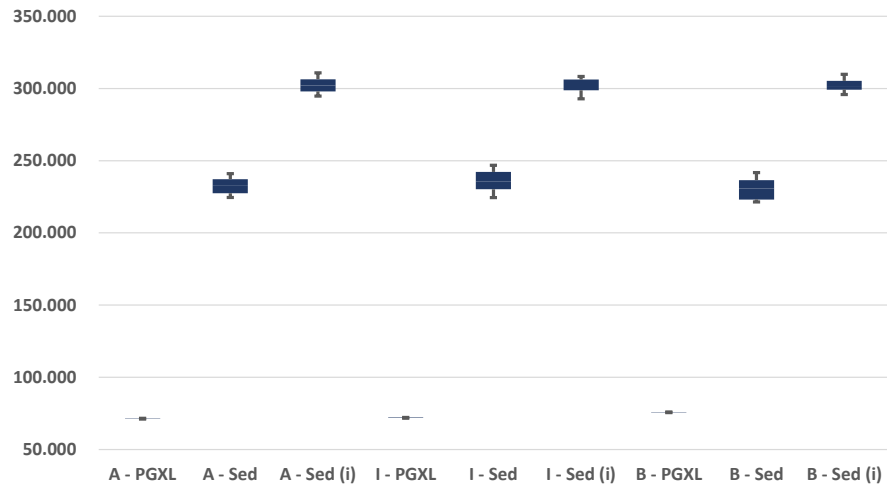


(c) Consulta 3

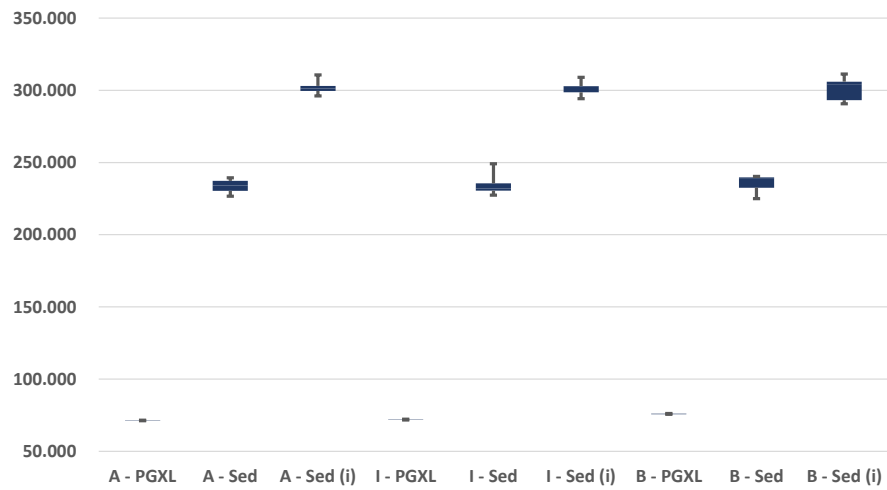
Figura 6.3: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2.



(a) Consulta 4

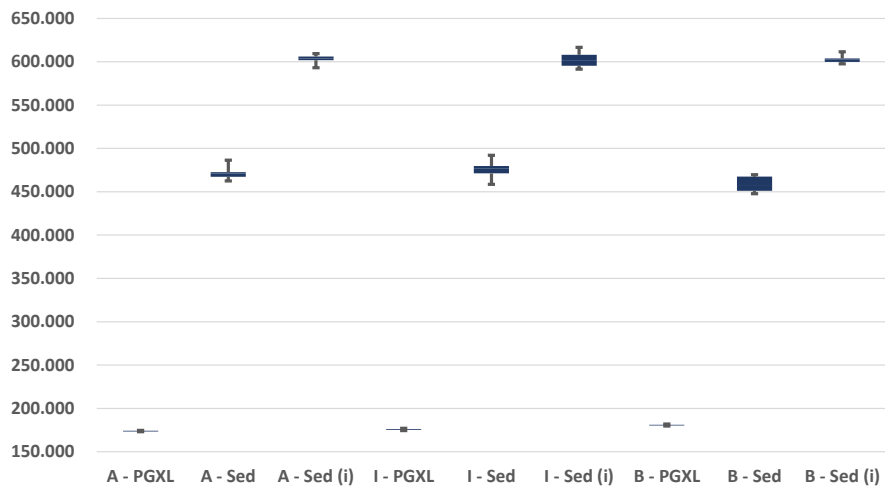


(b) Consulta 5

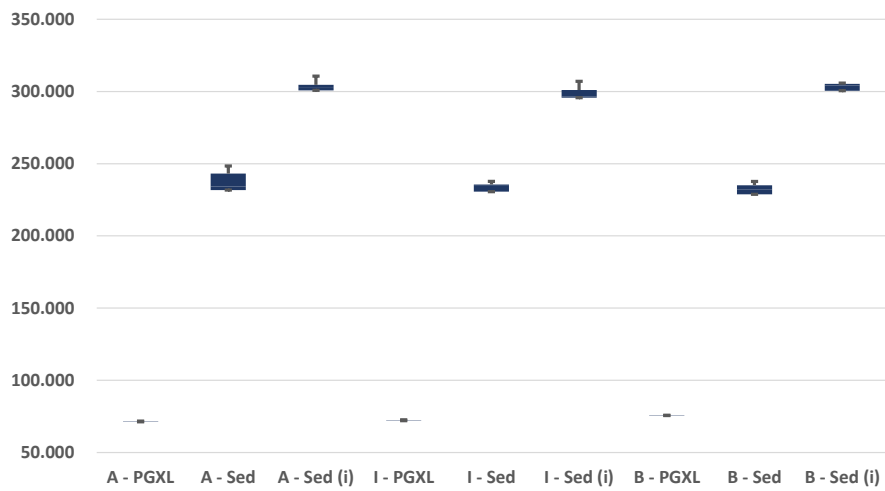


(c) Consulta 6

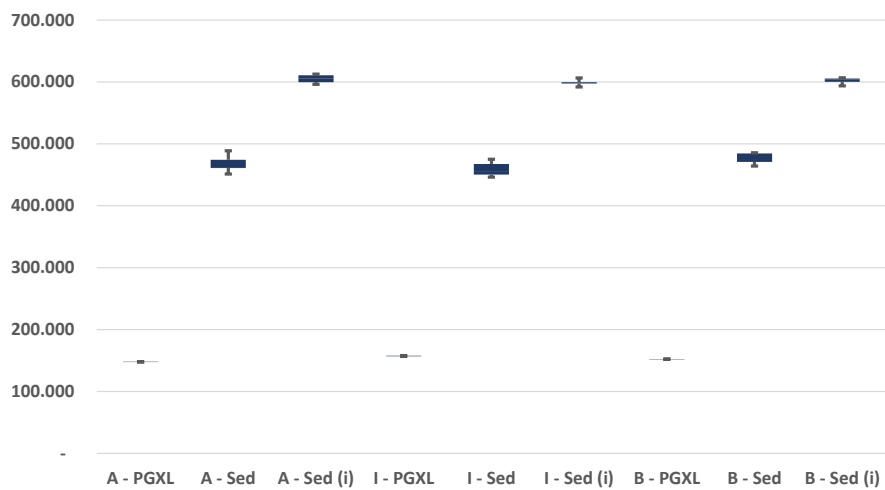
Figura 6.4: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).



(a) Consulta 7

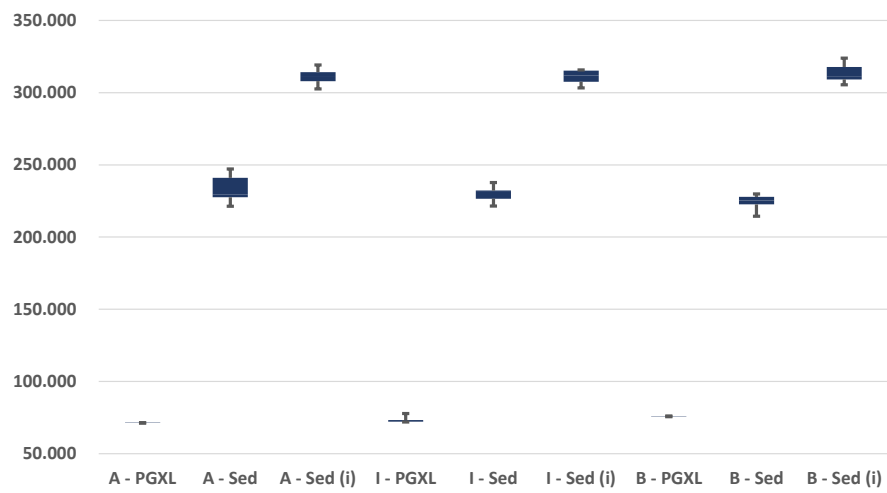


(b) Consulta 8

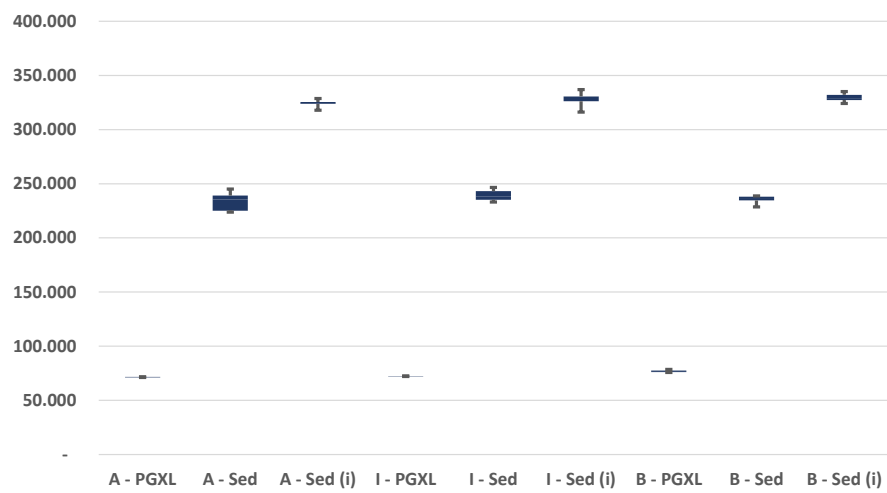


(c) Consulta 9

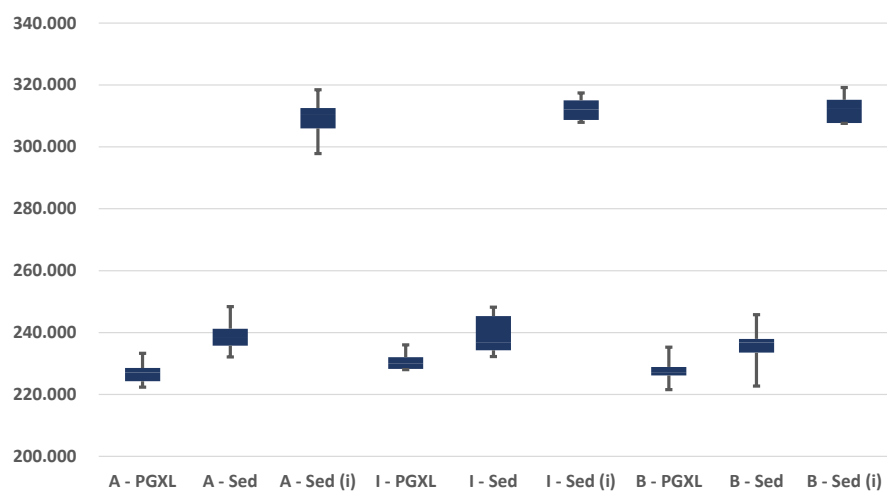
Figura 6.5: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).



(a) Consulta 10



(b) Consulta 11



(c) Consulta 12

Figura 6.6: Gráficos de caixas dos tempos de execução (em milissegundos) das consultas no Experimento 2 (continuação).

Tabela 6.14: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de mobilidade intermediária (Apache Sedona).

	Ambiente 2 (Apache Sedona e HDFS)				Ambiente 2 (Apache Sedona (indexado) e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	236.446	253.236	223.566	8.787	304.897	309.797	297.969	3.624
C2	251.850	259.263	245.076	5.312	334.823	341.310	328.890	3.826
C3	250.341	258.939	239.733	7.291	323.832	330.018	320.175	3.793
C4	253.577	264.356	233.748	8.336	324.046	330.817	317.075	5.327
C5	235.503	246.825	224.433	7.903	301.442	308.321	292.840	4.962
C6	233.907	249.132	227.395	6.379	301.595	308.980	294.190	4.322
C7	475.106	492.086	458.545	9.544	602.119	616.646	591.446	8.043
C8	233.271	237.756	228.233	3.323	298.683	307.037	294.989	4.053
C9	476.271	485.635	464.176	7.707	602.417	606.789	593.685	4.921
C10	230.590	237.755	221.484	5.086	310.942	315.733	303.285	4.354
C11	238.955	246.592	233.113	4.760	328.266	336.923	316.131	5.649
C12	239.033	248.206	232.248	5.926	312.229	317.422	307.930	3.503

Tabela 6.15: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de baixa mobilidade (Postgres-XL).

	Ambiente 1 (Postgres-XL e PostGIS)			
	Média	Maior	Menor	DP
C1	265.291	272.035	259.392	4.242
C2	236.527	261.165	230.094	9.843
C3	75.825	76.112	75.603	170
C4	712.909	727.446	698.352	9.626
C5	75.734	76.007	75.560	155
C6	75.917	76.217	75.653	198
C7	180.721	181.952	179.653	720
C8	75.686	75.807	75.519	104
C9	157.235	158.023	156.581	546
C10	75.821	76.136	75.622	167
C11	77.075	78.596	75.752	998
C12	227.925	235.277	221.564	4.049

6.5.3 Análise dos Resultados

As subseções anteriores apresentaram os resultados gerais dos experimentos realizados. Nesta seção, realiza-se uma análise detalhada desses resultados, abordando o desempenho geral das ferramentas avaliadas, a variabilidade dos resultados, o consumo de espaço de armazenamento, o desempenho em função da mobilidade, e a influência da indexação nos

Tabela 6.16: Experimento 2: Resultados (tempo em milissegundos) para o *dataset* de baixa mobilidade (Apache Sedona).

	Ambiente 2 (Apache Sedona e HDFS)				Ambiente 2 (Apache Sedona (indexado) e HDFS)			
	Média	Maior	Menor	DP	Média	Maior	Menor	DP
C1	234.178	239.336	225.463	4.305	306.548	317.733	297.847	6.815
C2	247.412	252.446	240.993	3.857	333.915	341.196	327.653	4.220
C3	246.681	253.711	231.471	6.843	324.672	330.741	316.894	5.245
C4	247.670	255.673	240.363	4.403	330.818	337.594	325.351	4.192
C5	230.386	241.757	221.428	7.418	302.229	309.802	295.806	4.824
C6	235.399	240.377	224.969	6.177	300.826	311.255	290.663	7.327
C7	458.680	469.679	447.717	8.231	602.849	611.509	597.576	4.113
C8	232.204	237.700	226.228	4.073	302.273	305.797	294.396	3.858
C9	458.949	475.162	446.234	10.680	598.903	606.512	591.938	3.788
C10	224.417	229.866	214.426	4.615	312.966	323.894	305.448	6.013
C11	235.379	238.760	228.606	3.633	329.321	335.103	324.103	3.613
C12	236.128	245.778	222.712	6.370	312.215	319.180	307.589	4.062

tempos de execução e na manipulação dos dados.

Desempenho Geral das Ferramentas

Em ambos os experimentos, nas consultas de observações (Consultas 5, 6, 7 e 8), todas baseadas em filtros sobre atributos não espaciais, o Postgres-XL apresentou desempenho superior (menor registro de tempo de execução, e menor média nos tempos) ao do Apache Sedona. Em todas as consultas envolvendo agregações (Consulta 5 - contagem, Consulta 6 - média, Consulta 7 - mínimo e Consulta 9 - máximo) o Postgres-XL se sobressaiu. Isso pode estar relacionado ao tratamento das agregações parciais, realizadas em cada nó trabalhador (Coordenador no Postgres-XL, e *Workers* no Apache Sedona).

O Postgres-XL apresentou desempenho superior em grande parte das consultas baseadas em filtros não espaciais. Também destacou-se positivamente em operações de agregação (Consultas 5, 6, 7 e 9), nas quais a distribuição dos dados e a execução paralela favoreceram a redução dos tempos de resposta. Por outro lado, seu desempenho foi limitado em consultas do tipo kNN devido à necessidade de execução de *loops* aninhados, o que resultou em maior tempo de processamento.

O Apache Sedona, por sua vez, apresentou melhor desempenho em consultas que envolviam filtros espaciais, beneficiando-se de seus mecanismos otimizados para o processamento geoespacial em larga escala. Esse resultado reforça o papel do Sedona como ferramenta especializada para cenários em que a dimensão espacial constitui fator central de análise.

Na Consulta 12, observou-se que o Ambiente 1 se sobressaiu no Experimento 1, enquanto o Ambiente 2 (Apache Sedona sem índice) obteve o melhor resultado no Experimento 2. A diferença entre os tempos observados nos dois ambientes foi relativamente pequena quando comparada às discrepâncias verificadas nas demais consultas. Apesar de se tratar de uma operação de kNN — tipo de consulta em que o Ambiente 1 apresentou menor eficiência —, a aplicação do filtro não espacial (no atributo *observedProperty*) contribuiu para reduzir a diferença entre os resultados.

Variabilidade dos Resultados

Na análise dos resultados, não foi identificada nenhuma relação consistente entre a variabilidade dos tempos de execução e o aumento ou diminuição do nível de mobilidade dos *datasets* utilizados. Foi possível observar a variabilidade (grau de oscilação em relação à média) dos tempos de execução nos experimentos por meio da observação dos gráficos de caixas (Figuras 6.1 a 6.6), assim como pela análise do desvio padrão nas tabelas de resultados (Tabelas 6.7 a 6.9, e 6.11 a 6.16). Nos três níveis de mobilidade considerados, o Ambiente 1 apresentou um maior número de consultas com baixa variabilidade, na ordem de centenas de milissegundos (Consultas 3, 5, 6, 8, 10 e 11 no Experimento 1, e Consultas 3, 5, 6, 7, 8, 9, 10 e 11 no Experimento 2). Esse comportamento indica que, apesar dos tempos de execução no Ambiente 1 serem geralmente menores do que no Ambiente 2, a variabilidade dos tempos (representada pelo desvio padrão) também se manteve significativamente mais baixa.

No caso do Postgres-XL, observou-se que a variabilidade foi mais acentuada nas consultas com filtros puramente espaciais, especialmente a Consulta 2 (Filtro *ST_Within* delimitado por um retângulo) e a Consulta 4 (kNN). Um possível fator que contribui para esse comportamento é o fato de o PostGIS ter sido originalmente projetado para o PostgreSQL e não especificamente para o Postgres-XL. Essa diferença pode resultar em adaptações menos otimizadas no ambiente distribuído, impactando a consistência dos tempos de execução em consultas que exploram intensivamente operadores espaciais.

A maior variabilidade geral nos tempos de execução observada no Apache Sedona pode estar relacionada a fatores de infraestrutura e gerenciamento de recursos. Primeiramente, o tamanho dos *dataframes* processados pelo Sedona era superior ao dos bancos armazenados no Postgres-XL, o que aumentou o volume de dados manipulados em memória durante as consultas. Além disso, em ambos os ambientes, o tamanho dos bancos superava a memória física disponível, ocasionando uso intensivo de memória virtual, o que pode ter amplificado a instabilidade no tempo de resposta das operações distribuídas. Outro aspecto relevante é que, enquanto o Postgres-XL executava apenas seus serviços próprios (coordenadores e *datanodes*) em cada nó, o ambiente do Sedona compartilhava recursos

com o Spark e o Hadoop, incluindo o HDFS, intensificando a competição por CPU, memória e I/O. Esses fatores combinados contribuem para explicar a maior dispersão nos tempos de execução das consultas realizadas no Sedona em comparação ao Postgres-XL.

Conforme dito na Seção 6.5.1, os valores extremos (maior e menor) foram desconsiderados na análise dos resultados. É importante destacar, entretanto, que, apesar de apresentar menor variabilidade geral em comparação ao Apache Sedona, o Postgres-XL registrou uma maior ocorrência desses *outliers*, fato que pode ser percebido com clareza nos gráficos de caixas dos resultados e evidencia que sua execução, em alguns casos, pode ser mais suscetível a desvios pontuais de desempenho.

Consumo de Espaço de Armazenamento

Em ambos os experimentos, o Ambiente 2 (HDFS) consumiu mais espaço de armazenamento no *cluster*. Este consumo de espaço mostrou-se mais diretamente relacionado ao tamanho do arquivo do *dataset* utilizado, refletindo principalmente o efeito do fator de replicação adotado (três cópias por bloco de dados) aplicado aos blocos de dados dos arquivos do *dataset* de entrada.

No caso do Postgres-XL, não há a utilização de um fator de replicação fixo como no HDFS. O aumento observado no espaço ocupado está associado a sobrecargas internas do sistema de gerenciamento, tais como metadados e mecanismos de alinhamento dos dados, e não à duplicação sistemática dos registros em múltiplos nós.

O consumo de espaço também foi analisado sob a perspectiva da mobilidade. Investigou-se a influência da variabilidade dos registros de localização — ou seja, a quantidade de coordenadas distintas, que tende a crescer com o aumento do nível de mobilidade — sobre a organização interna do banco, considerando que dados mais repetitivos podem favorecer a compressão e, conseqüentemente, reduzir o espaço ocupado. No Experimento 1, que utilizou *datasets* de maior volume em relação ao Experimento 2, a redução da mobilidade (registros de localização mais repetitivos) resultou em menor espaço ocupado, evidenciando um efeito positivo desse padrão sobre o armazenamento. Entretanto, no Experimento 2, com *datasets* menores, a diminuição da mobilidade foi acompanhada do incremento no tamanho do banco, sugerindo que, em volumes de dados reduzidos, esse efeito não se manifesta da mesma forma e outros fatores de organização interna passam a ter maior influência no consumo de espaço.

Desempenho e Mobilidade

Nos dois experimentos, a relação entre o desempenho das ferramentas e o nível de mobilidade nos ambientes computacionais foi examinada tanto a partir das médias dos tempos

de execução quanto por meio da interpretação dos diagramas de caixas referentes aos resultados das consultas.

No Experimento 1, no Ambiente 1, identificaram-se dois casos em que o aumento da mobilidade do *dataset* foi acompanhado por maiores tempos de execução: as Consultas 2 (com o filtro *ST_Within*) e 4 (kNN). Em contrapartida, em algumas consultas observou-se redução nos tempos de execução à medida que a mobilidade do *dataset* aumentou: as Consultas 1 (busca por localizações dos registros, com filtro temporal) e 12 (kNN com filtro por propriedade observada). Cabe ressaltar que, em ambos os casos, além do componente geoespacial, houve influência de outros filtros aplicados aos atributos *observationTime* e *observedProperty*. A Consulta 7 (menor registro de umidade), por sua vez, apresentou tempos mais elevados no *dataset* de baixa mobilidade, em função de outras características do conjunto de dados, uma vez que não envolveu aspectos geoespaciais. Nas demais consultas, assim como no Ambiente 2 (independentemente da utilização de indexação), a variação de desempenho decorrente da variação no nível de mobilidade foi pouco significativa.

No Experimento 2, as variações nos tempos de execução associadas ao aumento do nível de mobilidade mostraram-se mais sutis, em função do menor tamanho das bases de dados. No Ambiente 1, verificou-se nas Consultas 1 e 2 um incremento nos tempos de execução à medida que cresceu o nível de mobilidade do *dataset*. Nas demais consultas, assim como no Ambiente 2 (independentemente da utilização de indexação), as alterações de desempenho decorrentes da variação no nível de mobilidade mantiveram-se limitadas, o que evidencia a estabilidade do comportamento dessas operações frente a diferentes níveis de mobilidade.

Análise da Indexação

No Experimento 2, a utilização de índices espaciais mostrou-se desfavorável em ambos os ambientes avaliados. No Ambiente 1 (Postgres-XL), o emprego do índice GiST no campo geoespacial sequer foi funcional. No Ambiente 2 (Apache Sedona), a adoção de um índice R-Tree resultou em decréscimo no desempenho, sem benefícios evidentes nos cenários analisados. Tal comportamento pode estar associado à sobrecarga decorrente da construção e manutenção do índice em ambiente distribuído, sobretudo em operações que não se beneficiam diretamente dessa estrutura. Cabe destacar que os experimentos foram realizados em um ambiente com capacidade computacional aquém do recomendado para cargas intensivas, o que pode ter impactado o comportamento observado. Dessa forma, nos contextos avaliados, a aplicação de índices espaciais não se mostrou recomendada para a arquitetura proposta.

6.6 Limitações

Apesar dos resultados alcançados, algumas limitações a respeito do trabalho devem ser apontadas. Uma delas advém da dificuldade de obtenção de um *dataset* amplamente validado, que capture de forma realista a dinâmica dos dados de observação e localização em diferentes níveis de mobilidade. Embora a simulação dos dispositivos tenha sido realizada com base em *datasets* reais, buscando refletir de forma realista o comportamento dos dispositivos, as simplificações inerentes à geração de dados sintéticos podem não abranger toda a complexidade de sistemas reais, como variações ambientais, interferências externas e comportamentos imprevisíveis dos dispositivos.

Na elaboração das consultas, buscou-se cobrir um amplo conjunto de operações sobre os dados, incluindo filtros espaciais, consultas por atributos e análises de trajetória. No entanto, essas consultas podem não refletir demandas e cenários reais típicos de aplicações reais. Todas as operações testadas neste estudo envolveram apenas uma única tabela e utilizaram operações espaciais básicas, o que permitiu uma avaliação inicial do desempenho das ferramentas em consultas simples. No entanto, para uma análise mais abrangente, é essencial verificar em trabalhos futuros o comportamento dessas ferramentas em operações mais complexas, como junções espaciais e consultas envolvendo múltiplas tabelas, de forma a obter uma visão mais completa sobre a escalabilidade e a eficiência de cada tecnologia em cenários de maior complexidade computacional.

Uma limitação adicional deste trabalho está no fato de que nenhuma máquina executava somente um serviço, o que pode ter interferido no desempenho das operações. Nesse cenário, eventuais picos de processamento em um serviço podem impactar o desempenho de outro, gerando variações nos tempos de resposta e dificultando a avaliação isolada de cada componente do sistema. Para minimizar esse impacto, foram realizados testes preliminares para ajustar os parâmetros das ferramentas e amenizar a sobrecarga nas máquinas. No entanto, futuros experimentos com maior quantidade de recursos podem explorar a atribuição de máquinas exclusivas a determinados serviços.

Apesar das limitações, a abordagem adotada permitiu uma exploração abrangente, garantindo que os experimentos fossem reproduzíveis e pudessem ser ajustados para diferentes parâmetros. Futuras pesquisas podem se beneficiar da validação com aplicações reais, permitindo uma análise mais aprofundada da aplicabilidade das consultas e dos modelos de mobilidade em cenários do mundo real.

É importante destacar uma limitação deste trabalho relacionada ao levantamento bibliográfico. O mapeamento sistemático descrito no Capítulo 3 teve como base uma busca realizada em outubro de 2024, refletindo, portanto, o estado da arte até aquele momento. Ainda que o processo metodológico completo não tenha sido repetido desde então, a bibliografia vem sendo continuamente atualizada. Para reduzir possíveis lacunas, foi con-

duzida uma busca manual complementar, incluindo um processo de *forward snowballing*, contemplando publicações até o primeiro semestre de 2025.

6.7 Ameaças à Validade

Apesar dos rigor metodológico adotado na condução da pesquisa, é importante considerar possíveis ameaças à validade de seus resultados.

O Postgres-XL foi adotado na implementação da arquitetura por se tratar de uma versão distribuída do PostgreSQL, um sistema relacional amplamente utilizado para dados geoespaciais em conjunto com o *plugin* PostGIS. Entretanto, a ferramenta foi descontinuada no ano de 2024. Apesar de sua utilização ter enriquecido a experimentação, ao permitir observar o comportamento de um banco de dados relacional no contexto de *big data* geoespacial de IoT, a descontinuação compromete a replicação dos experimentos e a aplicabilidade dos resultados em trabalhos futuros.

A elaboração das consultas usadas nos experimentos pode influenciar diretamente os resultados e interpretações, introduzindo vieses na seleção dos dados ao privilegiar subconjuntos específicos de registros ou operadores espaciais. A escala dos experimentos impediu que uma maior quantidade de consultas pudesse ser testada, e limitou a quantidade de iterações das consultas nos experimentos. Para evitar que isso afetasse a representatividade dos resultados, buscou-se abranger uma ampla diversidade de operações na formulação das consultas para garantir que os resultados refletissem, de forma equilibrada, diferentes aspectos dos dados geoespaciais e observacionais considerados.

O uso de imagens Docker permite a replicação dos testes em diferentes ambientes de maneira padronizada e controlada. No entanto, essa abordagem pode representar uma ameaça à validade dos resultados, especialmente no que diz respeito à eficiência da comunicação entre os serviços e à capacidade das ferramentas de explorar a localidade dos dados nos diversos nós da rede. O uso do Docker Swarm para a orquestração dos contêineres pode introduzir sobrecarga e latências adicionais em relação a abordagens que operam instaladas diretamente na máquina, sem abstrações de sistema operacional ou infraestrutura de rede. Para mitigar isso, buscou-se otimizar a configuração da rede no Docker Swarm, fixando serviços e volumes de dados em nós específicos.

A configuração dos parâmetros dos diferentes serviços utilizados nos experimentos pode introduzir vieses nos resultados, uma vez que ajustes específicos podem influenciar o desempenho dos serviços de forma não intencional. Alguns exemplos de parâmetros que podem afetar o desempenho dos serviços são o número de *threads*, o número de núcleos de processamento utilizados, e a alocação de memória. Para mitigar esse risco, foram realizados diversos testes preliminares antes da execução de cada experimento, visando definir

valores para os parâmetros de modo a equilibrar desempenho e consistência, identificando configurações mais adequadas para as cargas de trabalho esperadas.

6.8 Resultados Acadêmicos

No decorrer do doutorado, foram publicados os seguintes trabalhos relacionados à pesquisa sendo realizada:

- Silva, Dennis Sávio e Holanda, Maristela. *Applications of Geospatial Big Data in the Internet of Things*. Transactions in GIS, 26(1):41–71, 2021.
- Dias, Lucas; Silva, Dennis; Sousa Júnior, Rafael e Holanda, Maristela. *C*DynaConf: An Apache Cassandra Auto-tuning Tool for Internet of Things Data*. 6th International Conference on Internet of Things, Big Data and Security, 2021, Online Streaming. Proceedings of the 6th International Conference on Internet of Things, Big Data and Security, 2021.
- Silva, Dennis Sávio e Holanda, Maristela. *Uma Arquitetura para Big Data Geoespacial de Dispositivos IoT com Níveis Heterogêneos de Mobilidade*. Simpósio Brasileiro de Banco de Dados (SBBD), 2025.

O seguinte trabalho, embora não relacionado diretamente ao tema desta tese, foi publicado durante o doutorado:

- Silva, Dennis Sávio M.; da Silva, Waldeyr M. C.; Ruizhe, Guo; Bernardi, Ana Paula; Mariano, Ari Melo; Holanda, Maristela. *Big Data Trends in Bioinformatics*. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019.

6.9 Considerações Finais

Este capítulo apresentou o processo de experimentação conduzido nesta tese para validar a arquitetura proposta no Capítulo 5. Foi feita uma contextualização dos experimentos e da preparação do ambiente computacional onde funcionaram os ambientes propostos para a camada de Armazenamento e Análise de Dados. Em seguida, foram apresentadas as consultas empregadas na experimentação, que abrangeram tanto dados geoespaciais quanto de observações realizadas pelos dispositivos IoT. Então foram apresentados os experimentos realizados, os resultados obtidos e suas implicações. Por fim, discutiram-se as limitações do trabalho, as iniciativas de mitigação adotadas, as ameaças à validade e

as considerações finais do capítulo. Os resultados descritos neste capítulo, assim como o levantamento de ferramentas descrito no Capítulo 4 e a descrição da arquitetura no Capítulo 5, foram publicados em [274].

Capítulo 7

Conclusão

Esta tese investigou a aplicação de *big data* geoespacial em IoT, e teve como objetivo propor e implementar uma arquitetura para o armazenamento e a consulta de grandes volumes de dados geoespaciais oriundos de dispositivos IoT com diferentes níveis de mobilidade. O tema foi explorado sob perspectivas teóricas, por meio de um mapeamento de literatura e um levantamento de ferramentas, e práticas, ao propor, implementar e validar o protótipo de uma arquitetura para *big data* geoespacial de IoT oriundo de dispositivos com diferentes níveis de mobilidade.

O mapeamento sistemático da literatura investigou a aplicação de *big data* geoespacial em IoT. No total, identificaram-se 196 trabalhos, classificados em seis grupos conforme seus objetivos e contribuições. Desses, extraíram-se os principais domínios de aplicação de dados geoespaciais em IoT, *datasets* disponíveis, formatos de arquivos e ferramentas aplicadas, além de serem identificadas tendências de pesquisa. O levantamento das ferramentas elicitou alguns dos *frameworks* e SGBDs para *big data* geoespacial mais citados na literatura, suas principais características e suporte a recursos geoespaciais como índices, geometrias, linguagens de consulta, operações espaciais e formatos de arquivos. Também foi exposta a integração entre essas ferramentas.

Foi proposta uma abordagem para *big data* geoespacial de IoT que considera o nível de mobilidade dos dispositivos na escolha do mecanismo de armazenamento, na forma de uma arquitetura em duas camadas que contemplam armazenamento e análise dos dados. Foi implementado um protótipo da arquitetura, e nele foi realizada uma avaliação do desempenho e utilização de espaço de armazenamento de algumas das principais ferramentas de *big data*. A experimentação e análise dos resultados permitiu avaliar o comportamento de diferentes ambientes computacionais de análise e armazenamento de dados, submetidos a diversos níveis de mobilidade e volumes de dados, considerando tanto dados de observações quanto de localização. A arquitetura proposta pode beneficiar gestores de sistemas que lidam com *big data* geoespacial ao oferecer uma abordagem flexível para o

armazenamento e processamento dessas informações. Ao considerar o nível de mobilidade dos dispositivos IoT gerenciados, permite a escolha de soluções de armazenamento mais adequadas, equilibrando desempenho e eficiência. Ao propor o uso de ferramentas voltadas a *big data* geoespacial, oferece escalabilidade e desempenho no armazenamento e análise dos dados.

7.1 Trabalhos Futuros

Com base no que foi apresentado nesta tese, alguns trabalhos podem ser propostos para a continuidade da pesquisa:

- **Oportunidades de Evolução da Arquitetura** – Na camada de Aplicação, trabalhos futuros podem contemplar o avanço no desenvolvimento da interface, incluindo a criação de uma ferramenta web que permita consultas aos dados da aplicação e a geovisualização dos resultados das operações. Também pode ser considerada a integração da arquitetura a um *middleware* IoT, possibilitando a coleta direta e em tempo real dos dados gerados pelos dispositivos. Além disso, podem ser exploradas modelagens alternativas dos dados na camada de Armazenamento e Análise de Dados e realizados experimentos adicionais envolvendo operações de maior complexidade.
- **Realização de experimentos com novos ambientes na camada de Armazenamento e Análise de Dados** – Novas combinações de ferramentas podem ser avaliadas nessa camada. Como exemplo, destaca-se a possibilidade de utilização da GeoMesa, amplamente empregada na análise de dados geoespaciais, em conjunto com bancos de dados NoSQL, como o Redis ou o Apache Cassandra.
- **Proposição e validação de métricas de nível de mobilidade para dispositivos IoT** – Trabalhos futuros podem contemplar o desenvolvimento de métricas destinadas a apoiar análises comparativas entre diferentes cenários de nível de mobilidade, mensurando esse aspecto tanto no escopo de dispositivos individuais quanto do sistema como um todo.
- **Aplicação da arquitetura em um projeto real** – Implementação da arquitetura em um contexto prático, de modo a verificar em um cenário real seu desempenho, a integração entre as camadas e a utilidade das ferramentas de armazenamento, análise e geovisualização de dados geoespaciais.

Referências

- [1] al Jawarneh, Isam, Paolo Bellavista, Antonio Corradi e Luca Foschini: *Big Spatial Data Management for the Internet of Things: A Survey*. Journal of Network and Systems Management, 28(4):990–1035, outubro 2020. 1, 8, 23, 30, 33
- [2] Jo, Junghee e Kang Woo Lee: *MapReduce-Based D_ELT Framework to Address the Challenges of Geospatial Big Data*. ISPRS International Journal of Geo-Information, 8(11):475, outubro 2019. 1, 7, 9, 22, 23
- [3] Liu, Zhen, Huadong Guo e Changlin Wang: *Considerations on Geospatial Big Data*. IOP Conference Series: Earth and Environmental Science, 46:012058, novembro 2016. 1, 7, 9
- [4] Dritsas, Elias e Maria Trigka: *Database Systems in the Big Data Era: Architectures, Performance, and Open Challenges*. IEEE Access, 13:95068–95084, 2025. 1, 8, 9, 10
- [5] Kamilaris, Andreas e Frank Ostermann: *Geospatial Analysis and the Internet of Things*. ISPRS International Journal of Geo-Information, 7(7):269, julho 2018, ISSN 2220-9964. 1, 13, 19, 26
- [6] Armstrong, Marc P., Shaowen Wang e Zhe Zhang: *The internet of things and fast data streams: prospects for geospatial data science in emerging information ecosystems*. Cartography and Geographic Information Science, 46(1):39–56, 2019. 1, 8, 9, 11, 12
- [7] Zee, Erik van der e Henk Scholten: *Spatial Dimensions of Big Data: Application of Geographical Concepts and Spatial Technology to the Internet of Things*, páginas 137–168. Springer International Publishing, 2014, ISBN 978-3-319-05029-4. 1
- [8] Ge, Mouzhi, Hind Bangui e Barbora Buhnova: *Big data for internet of things: A survey*. Future Generation Computer Systems, 87:601–614, 2018, ISSN 0167-739X. <https://www.sciencedirect.com/science/article/pii/S0167739X17316953>. 1, 9, 13, 19
- [9] Ding, Zhiming, Zhikui Chen e Qi Yang: *IoT-SVKSearch: A Real-Time Multimodal Search Engine Mechanism for the Internet of Things*. Int. J. Commun. Syst., 27(6):871–897, jun 2014. 1, 11, 19, 21, 23, 25, 27
- [10] Isikdag, Umit: *An IoT Architecture for Facilitating Integration of Geoinformation*. International Journal of Engineering and Geosciences, 5:15–25, fevereiro 2020. 1, 11, 12, 13, 19, 20, 21, 22, 23, 26, 28

- [11] Guo, Dongming e Erling Onstein: *State-of-the-art geospatial information processing in nosql databases*. ISPRS International Journal of Geo-Information, 9(5), 2020, ISSN 2220-9964. 1, 7, 8
- [12] Shah, Purnimabahen Rasiklal: *Developing Big Data Analytics Architecture for Spatial Data*. Tese de Doutoramento, Ahmedabad University, 2019. 1, 11, 33, 42
- [13] Li, Xin, Huayan Yu, Ligang Yuan e Xiaolin Qin: *Query optimization for distributed spatio-temporal sensing data processing*. Sensors, 22(5), 2022, ISSN 1424-8220. <https://www.mdpi.com/1424-8220/22/5/1748>. 1
- [14] Jo, Junghee e Kang Woo Lee: *High-performance geospatial big data processing system based on mapreduce*. ISPRS International Journal of Geo-Information, 7(10), 2018, ISSN 2220-9964. 1
- [15] Taha, Kamal: *Big data analytics in iot, social media, nlp, and information security: trends, challenges, and applications*. Journal of Big Data, 12(1):150, Jun 2025, ISSN 2196-1115. <https://doi.org/10.1186/s40537-025-01192-9>. 1, 8
- [16] Li, Pengfei, Hua Lu, Qian Zheng, Long Yang e Gang Pan: *Lisa: A learned index structure for spatial data*. Em Maier, David, Rachel Pottinger, AnHai Doan, Wang-Chiew Tan, Abdussalam Alawini e Hung Q. Ngo (editores): *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*, páginas 2119–2133. ACM, 2020. 2, 27
- [17] Dai, Jing e CHang Tien Lu: *Dime: Disposable index for moving objects*. Em *2011 IEEE 12th International Conference on Mobile Data Management*, volume 1, páginas 68–77, 2011. 2
- [18] Hevner, Alan R., Salvatore T. March, Jinsoo Park e Sudha Ram: *Design science in information systems research*. MIS Quaterly, 28(1):75–105, março 2004, ISSN 0276-7783. 4
- [19] Wieringa, Roel J.: *Design Science Methodology for Information Systems and Software Engineering*. Springer Publishing Company, Incorporated, 2014, ISBN 3662438380. 4
- [20] Wohlin, Claes, Per Runeson, Martin Hst, Magnus C. Ohlsson, Björn Regnell e Anders Wessln: *Experimentation in Software Engineering*. Springer Publishing Company, Incorporated, 2012, ISBN 3642290434. 4
- [21] Almeida, Damião Ribeiro de, Cláudio de Souza Baptista, Fabio Gomes de Andrade e Amílcar Soares: *A survey on big data for trajectory analytics*. ISPRS International Journal of Geo-Information, 9(2), 2020, ISSN 2220-9964. 7, 9
- [22] Clementini, Eliseo, Paolino Di Felice e Peter van Oosterom: *"a small set of formal topological relationships suitable for end-user interaction"*. Em Abel, David e Beng Chin Ooi (editores): *Advances in Spatial Databases*, páginas 277–295, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg, ISBN 978-3-540-47765-5. 8

- [23] Rothermel, Kurt, Otthein Herzog e Siegfried Wu Zhiqiang: *Proximity-based service discovery for distributed digital twin systems*. Discover Internet of Things, 5(1):14, Feb 2025, ISSN 2730-7239. <https://doi.org/10.1007/s43926-025-00103-x>. 8
- [24] Herring, John R.: *Opengis implementation standard for geographic information — simple feature access — part 2: Sql option*. Relatório Técnico OGC 06-104r4, Open Geospatial Consortium, 2010. <http://www.opengeospatial.org/standards/sfs>, Status: Approved Corrigendum. 8
- [25] Li, Zhenlong: *Geospatial big data handling with high performance computing: Current approaches and future directions*. CoRR, abs/1907.12182, 2019. <http://arxiv.org/abs/1907.12182>. 8, 9, 20
- [26] Barik, Rabindra Kumar, Harishchandra Dubey, Kunal Mankodiya, Sapana Sasane e Chinmaya Misra: *Geofog4health: a fog-based SDI framework for geospatial health big data analysis*. Journal of Ambient Intelligence and Humanized Computing, 10(2):551–567, 2019. 8
- [27] Chinnaswamy, Anitha K., Armando Papa, Luca Dezi e Alberto Mattiacci: *Big data visualisation, geographic information systems and decision making in healthcare management*. Management Decision, 57, 2019. 8, 20, 21, 23
- [28] Lü, Guonian, Michael Batty, Josef Strobl, Hui Lin, A Xing Zhu e Min Chen: *Reflections and Speculations on the Progress in Geographic Information Systems (GIS): a Geographic Perspective*. International Journal of Geographical Information Science, 33(2):346–367, 2019. 9
- [29] Barik, Rabindra Kumar, Harishchandra Dubey, Arun B. Samaddar, Rajan D. Gupta e Prakash K. Ray: *Foggis: Fog computing for geospatial big data analytics*. Em *2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON)*, páginas 613–618, 2016. 9
- [30] Wu, Jiayang, Wensheng Gan, Han Chieh Chao e Philip S. Yu: *Geospatial big data: Survey and challenges*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17:17007–17020, 2024. 9
- [31] Detti, Andrea, Nicola Blefari Melazzi, Michele Orru, Riccardo Paolillo e Giulio Rossi: *Opengeobase: Information centric networking meets spatial database applications*. Em *2016 IEEE Globecom Workshops (GC Wkshps)*, páginas 1–7, 2016. 9, 32
- [32] Saha, A. K., A. Kumar, V. Tyagi e S. Das: *Big data and internet of things: A survey*. Em *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, páginas 150–156, 2018. 9, 12
- [33] Shrivastava, Swapnil: *A review of spatial big data platforms, opportunities, and challenges*. IETE Journal of Education, 61(2):80–89, 2020. 9, 37

- [34] Bovkir, Rabia e Arif Cagdas Aydinoglu: *Big urban data visualization approaches within the smart city: Gis-based open-source dashboard example*. Em *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 46, páginas 125–130, 2021. 9
- [35] Shah, Purnima: *Developing Big Data Analytics Architecture for Spatial Data*. Em Bartolini, Ilaria e Feifei Li (editores): *Proceedings of the VLDB 2019 PhD Workshop, co-located with the 45th International Conference on Very Large Databases (VLDB 2019), Los Angeles, California, USA, August 26-30, 2019*, volume 2399 de *CEUR Workshop Proceedings*. CEUR-WS.org, 2019. 9
- [36] C. Costa, Rogério Luís de, José Moreira, Paulo Pintor, Veronica dos Santos e Sérgio Lifschitz: *A survey on data-driven performance tuning for big data analytics platforms*. *Big Data Research*, 25:100206, 2021, ISSN 2214-5796. 9
- [37] Al-Yadumi, Sohaib, Tan Ee Xion, Sharon Goh Wei Wei e Patrice Boursier: *Review on Integrating Geospatial Big Datasets and Open Research Issues*. *IEEE Access*, 9:10604–10620, 2021, ISSN 2169-3536. 10, 30
- [38] Breunig, Martin, Patrick Erik Bradley, Markus Jahn, Paul Kuper, Nima Mazroob, Norbert Rösch, Mulhim Al-Doori, Emmanuel Stefanakis e Mojgan Jadidi: *Geospatial Data Management Research: Progress and Future Directions*. *ISPRS International Journal of Geo-Information*, 9(2), 2020. 10, 30
- [39] Eldawy, Ahmed e Mohamed F. Mokbel: *The era of big spatial data: A survey*. *Foundations and Trends® in Databases*, 6(3-4):163–273, 2016, ISSN 1931-7883. 11, 58
- [40] Finogeev, Alexey, Anton Finogeev, Ludmila Fionova, Artur Lyapin e Kirill A. Ly-chagin: *Intelligent monitoring system for smart road environment*. *Journal of Industrial Information Integration*, 15:15–20, 2019, ISSN 2452-414X. 11, 19, 20, 21, 22, 23, 27, 28
- [41] Kahraman İbrahim, Alper Köse, Mutlu Koca e Emin Anarim: *Age of information in internet of things: A survey*. *IEEE Internet of Things Journal*, 11(6):9896–9914, 2024. 11
- [42] Mattern, Friedemann e Christian Floerkemeier: *From the Internet of Computers to the Internet of Things*, páginas 242–259. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, ISBN 978-3-642-17226-7. 11
- [43] Atmani, Amine, Ibtissame Kandrouch, Nabil Hmina e Habiba Chaoui: *Big data for internet of things: A survey on iot frameworks and platforms*. Em Ezziyyani, Mostafa (editor): *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019)*, páginas 59–67, Cham, 2020. Springer International Publishing, ISBN 978-3-030-33103-0. 12
- [44] Sinda, Matthew e Qi Liao: *Spatial-Temporal Anomaly Detection using Security Visual Analytics via Entropy Graph and Eigen Matrix*. Em *Proceedings - 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing*,

- 2017 IEEE 15th International Conference on Pervasive Intelligence and Computing, 2017 IEEE 3rd International Conference on Big Data Intelligence and Computing and 2017 IEEE Cyber Science and Technology Congress, DASC-PICom-DataCom-CyberSciTec 2017*, volume 2018-January, páginas 511–518. IEEE, 2017. 12
- [45] Azad, Poopak, Nima Jafari Navimipour, Amir Masoud Rahmani e Arash Sharifi: *The Role of Structured and Unstructured Data Managing Mechanisms in the Internet of Things*. Cluster Computing, 23:1185–1198, 2020. 12, 13, 33
 - [46] Emani, Cheikh Kacfeh, Nadine Cullot e Christophe Nicolle: *Understandable big data: A survey*. Computer Science Review, 17:70 – 81, 2015, ISSN 1574-0137. <http://www.sciencedirect.com/science/article/pii/S1574013715000064>. 12
 - [47] Sarwat, Mohamed: *Spatial data systems support for the internet of things: Challenges and opportunities*. SIGSPATIAL Special, 12(2):42–47, outubro 2020. 13
 - [48] Sarkar, Subhadeep, Subarna Chatterjee e Sudip Misra: *Assessment of the Suitability of Fog Computing in the Context of Internet of Things*. IEEE Transactions on Cloud Computing, 6(1):46–59, 2018. 13
 - [49] Wan, Wei, Xingqiang Du, Xinwei Zhao e Zhenkun Yang: *A Cloud-Enabled Collaborative Hub for Analysis of Geospatial Big Data*. Em *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021*, páginas 1–5, 2021. 13
 - [50] Almutairi, Jaber e Mohammad Aldossary: *A novel approach for IoT tasks offloading in edge-cloud environments*. Journal of Cloud Computing, 10(1):28, Apr 2021. 13
 - [51] Mondragón-Ruiz, Giovanny, Alonso Tenorio-Trigoso, Manuel Castillo-Cara, Blanca Caminero e Carmen Carrión: *An experimental study of fog and cloud computing in CEP-based Real-Time IoT applications*. Journal of Cloud Computing, 10(1):32, Jun 2021, ISSN 2192-113X. 13
 - [52] Ding, Zhiming, Jiajie Xu e Qi Yang: *SeaCloudDM: A Database Cluster Framework for Managing and Querying Massive Heterogeneous Sensor Sampling Data*. The Journal of Supercomputing, 66(3):1260–1284, December 2013, ISSN 0920-8542. 13, 21, 22, 23, 27
 - [53] Chen, Xiaoyan, Xiaomin Xu, Sheng Huang, Weiming Ye, Lance Feagan, Lalitha Krishnamoorthy e Mark (M.A.) Ashworth: *STaaS: SpatioTemporal Historian as a Service*. Em *2015 IEEE International Conference on Web Services (ICWS)*, páginas 747–750. IEEE, julho 2015, ISBN 978-1-4673-7272-5. 13, 21, 22, 23
 - [54] Tenedório, José António e Jorge Rocha: *Introductory chapter: Spatial analysis, modelling, and planning*. Em Rocha, Jorge e José António Tenedório (editores): *Spatial Analysis, Modelling and Planning*, capítulo 1. IntechOpen, Rijeka, 2018. 13
 - [55] Silva, Dennis Sávio e Maristela Holanda: *Applications of Geospatial Big Data in the Internet of Things*. Transactions in GIS, 26(1):41–71, 2022. 14, 29

- [56] Kitchenham, Barbara e Stuart Charters: *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Relatório Técnico, Keele University and Durham University, 2007. Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report. 14
- [57] Petersen, Kai, Robert Feldt, Shahid Mujtaba e Michael Mattsson: *Systematic Mapping Studies in Software Engineering*. Em *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, página 68–77, 2008. 14
- [58] Eck, Nees Jan van e Ludo Waltman: *Vosviewer: Visualizing Scientific Landscapes*, 2010. <https://www.vosviewer.com>. 17
- [59] Youzhuo, Zheng, Fu Yu, Zhang Yang, Hao Shuqing e Cai Yongxiang: *Research and Practice of Intelligent Operation and Maintenance System for Essential Load Guarantee Based on 5G and IOT Technology*. Em *2020 4th International Conference on Power and Energy Engineering (ICPEE)*, páginas 53–57, Nov 2020. 19, 23, 28
- [60] Kim, Sung Ah, Dongyoun Shin, Yoon Choe, Thomas Seibert e Steffen P. Walz: *Integrated Energy Monitoring and Visualization System for Smart Green City Development: Designing a Spatial Information Integrated Energy Monitoring Model in the Context of Massive Data Management on a Web Based Platform*. *Automation in Construction*, 22:51 – 59, 2012, ISSN 0926-5805. 19, 21, 22, 23, 25
- [61] Rathore, M. Mazhar, Awais Ahmad, Anand Paul e Seungmin Rho: *Urban planning and building smart cities based on the Internet of Things using Big Data analytics*. *Computer Networks*, 101:63–80, 2016, ISSN 1389-1286. *Industrial Technologies and Applications for the Internet of Things*. 19, 22, 23, 27, 28
- [62] Anejionu, Obinna C.D., Piyushimita (Vonu) Thakuriah, Andrew McHugh, Yeran Sun, David McArthur, Phil Mason e Rod Walpole: *Spatial urban data system: A cloud-enabled big data infrastructure for social and economic urban analytics*. *Future Generation Computer Systems*, 98:456 – 473, 2019. 19, 21, 22, 23, 28
- [63] Wang, Zhoufeng, Jianwei Xu, Xiangqi He e Yujun Wang: *Analysis of Spatiotemporal Influence Patterns of Toxic Gas Monitoring Concentrations in an Urban Drainage Network Based on IoT and GIS*. *Pattern Recognition Letters*, 138:237 – 246, 2020. 19, 23
- [64] Wen, Rui, Hui Wang, Chunlin Wang, Yong Xiang e Chengke Zhou: *A Smart Cable Management System in support of the Smart City*. Em *CIREN Workshop 2016*, volume 2016, páginas 1–4. IET, junho 2016. 19, 22, 23
- [65] Tang, Bo, Zhen Chen, Gerald Heffernan, Tao Wei, Haibo He e Qing Yang: *A Hierarchical Distributed Fog Computing Architecture for Big Data Analysis in Smart Cities*. Em *Proceedings of the ASE BigData & SocialInformatics 2015*, número 6. Association for Computing Machinery, 2015. 19, 23

- [66] Alvarez, Manuel Garcia, Javier Morales e Menno Jan Kraak: *Integration and exploitation of sensor data in smart cities through event-driven applications*. Sensors (Basel, Switzerland), 19, 2019. 19, 20, 23, 27, 28
- [67] Volland, Patrick e Hartmut Asche: *Processing and Visualizing Floating Car Data for Human-Centered Traffic and Environment Applications: A Transdisciplinary Approach*. International Journal of Agricultural and Environmental Information Systems, 8(2):32–49, 2017, ISSN 1947-3192. 19, 21, 23, 28
- [68] Erdelić, Tomislav, Tonči Carić, Martina Erdelić, Leo Tišljarić, Ana Turković e Niko Jelušić: *Estimating congestion zones and travel time indexes based on the floating car data*. Computers, Environment and Urban Systems, 87:101604, 2021, ISSN 0198-9715. 19
- [69] Unal, Perin, Yunuscan Kocak e Yunus Donmez: *A Smart Winter Service Platform and Route Planning Algorithm*. Em *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, páginas 192–196. IEEE, agosto 2018, ISBN 978-1-5386-7503-8. 19, 22, 23, 28
- [70] Zhong, Yunqin, Jinyun Fang e Xiaofang Zhao: *VegaIndexer: A Distributed composite index scheme for big spatio-temporal sensor data on cloud*. Em *2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, páginas 1713–1716. IEEE, julho 2013, ISBN 978-1-4799-1114-1. 19, 23, 25
- [71] Wang, Cheng, Zhenzhen Zhang, Lu Shao e MengChu Zhou: *Estimating travel speed via sparse vehicular crowdsensing data*. Em *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, páginas 643–648, dezembro 2016. 19, 23
- [72] Fang, Shifeng, Lida Xu, Yunqiang Zhu, Yongqiang Liu, Zhihui Liu, Huan Pei, Jianwu Yan e Huifang Zhang: *An Integrated Information System for Snowmelt Flood Early-Warning Based on Internet of Things*. Information Systems Frontiers, 17(2):321–335, apr 2013, ISSN 1387-3326. 20, 22, 23
- [73] Zou, Xiangyu, Jinjin Zhao, Duan Zhao, Bin Sun, Yongxin He e Stelios Fuentes: *Air quality prediction based on a spatiotemporal attention mechanism*. Mobile Information Systems, 2021:6630944, Feb 2021, ISSN 1574-017X. 20, 22, 23
- [74] Xuyao, Yu, Yu Hui, Xu Kexin, Su Yijin e Li Jinhang: *Design and implementation of on-line automatic monitoring system for malodor pollution incidents based on ArcGIS*. Em *2013 IEEE 11th International Conference on Electronic Measurement Instruments*, volume 2, páginas 524–528. IEEE, agosto 2013. 20, 21, 22, 23
- [75] Moussa, Rim: *Scalable analytics of air quality batches with apache spark and apache sedona*. Em *DEBS 2021 - Proceedings of the 15th ACM International Conference on Distributed and Event-Based Systems*, páginas 154–159, 2021. 20, 23
- [76] Howell, Shaun, Yacine Rezgui e Thomas Beach: *Water utility decision support through the semantic web of things*. Environmental Modelling & Software, 102:94–114, 2018, ISSN 1364-8152. 20, 21, 23, 28

- [77] Suárez, José P., Agustín Trujillo, Conrado Domínguez e Pablo Fernández José Miguel Santana: *Managing and 3D visualization of real-time big geo-referenced data from Las Palmas port through a flexible open source computer architecture*. Em *2015 1st International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM)*, páginas 72–82, abril 2015, ISBN 978-989-758-142-7. 20, 21, 22, 23, 25, 27, 28
- [78] Kang, Yong Shin, Il Ha Park, Jongtae Rhee e Yong Han Lee: *MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data*. *IEEE Sensors Journal*, 16(2):485–497, janeiro 2016, ISSN 1558-1748. 20, 21, 23
- [79] Fernández, Pablo, Jose M. Santana, Sebastián Ortega, Agustín Trujillo, Jose P. Suárez, Jaisiel A. Santana, Alejandro Sánchez e Conrado Domínguez: *Web-based gis through a big data open source computer architecture for real time monitoring sensors of a seaport*. Em *The Rise of Big Spatial Data*, páginas 41–53. Springer International Publishing, 2017. 20, 23, 25, 27, 28
- [80] Fernández, Pablo, José Pablo Suárez, Agustín Trujillo, Conrado Domínguez e José Miguel Santana: *3d-monitoring big geo data on a seaport infrastructure based on fiware*. *Journal of Geographical Systems*, 20(2):139–157, April 2018. 20, 21, 22, 23, 25, 26, 27, 28, 37
- [81] Li, Min, Shaoquan Ni, Ling Zhou e Qiang Huang: *IOT Big Data Process Based on Context Temporal-Spatial Characteristics in Petroleum Products Distribution*. Em *Proceedings of the 5th International Conference on Transportation Engineering (ICTE)*, páginas 2065–2074, 2015. 20, 21, 22, 23
- [82] Gupta, Kamal, Sanjay Kumar Sadana e Bhoomi Gupta: *Geospatial Data Preprocessing and Visualization for the Logistics Industry*. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(1):57–64, 2020. 20, 21, 23
- [83] Benedusi, Paolo, Angelo Chianese, Fiammetta Marulli e Francesco Piccialli: *An associative engines based approach supporting collaborative analytics in the internet of cultural things*. Em *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, páginas 533–538, Nov 2015. 20, 21, 23
- [84] Chianese, Angelo, Fiammetta Marulli, Francesco Piccialli, Paolo Benedusi e Jai E. Jung: *An associative engines based approach supporting collaborative analytics in the internet of cultural things*. *Future Generation Computer Systems*, 66:187 – 198, 2017. 20, 21, 23, 28
- [85] Mello, Ronaldo dos Santos, Vania Bogorny, Luis Otavio Alvares, Luiz Henrique Zambom Santana, Carlos Andres Ferrero, Angelo Augusto Frozza, Geomar Andre Schreiner e Chiara Renso: *Master: A multiple aspect view on trajectories*. *Transactions in GIS*, 23(4):805–822, 2019. 20, 23, 24
- [86] Li, Baishou, Mingzhi Liao e Lulu Huang: *Research on Big Data Analysis of Location Service and Intelligent Service Platform of Urban Scenic Spots: A Case Study of Guilin City, Guangxi, China*. Em *International Archives of the Photogrammetry*,

- Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 42, páginas 1135–1138, 2020. 20, 23
- [87] Chunzhong, Wei, Ling Zhong, Jiang Yanbo, Deng Xuan e Wang Yu: *Design of intelligent monitoring system of integrated water supply and drainage based on big data and cloud computing*. Em *2022 2nd International Conference on Networking, Communications and Information Technology (NetCIT)*, páginas 1–7, Dec 2022. 20, 23
 - [88] Zhou, Tong, Wenzhong Shi, Xintao Liu, Fei Tao, Zhen Qian e Ruijia Zhang: *A Novel Approach for Online Car-Hailing Monitoring Using Spatiotemporal Big Data*. IEEE Access, 7:128936–128947, 2019, ISSN 2169-3536. 20, 23, 27
 - [89] Parygin, Danila S, Alexey G Finogeev, Valery A Kamaev, Anton A Finogeev, Elena P Gnedkova e Anton P Tyukov: *A convergent model for distributed processing of big sensor data in urban engineering networks*. Journal of Physics: Conference Series, 803:012112, janeiro 2017. 20, 23, 28
 - [90] Song, Xin, Jing Gao, Jin'an Ma, Shaokai Niu e Huiyuan He: *HTME: A Data Streams Processing Strategy based on Hoeffding tree in MapReduce environment*. Em *Proceedings of the 2016 12th World Congress on Intelligent Control and Automation (WCICA)*, páginas 1042–1045. IEEE, junho 2016, ISBN 978-1-4673-8414-8. 20, 22, 23, 24
 - [91] Hu, Guoqiang, Xin Zhang, Ning Duan e Peng Gao: *Towards Reliable Online Services Analyzing Mobile Sensor Big Data*. Em *2017 IEEE International Conference on Web Services (ICWS)*, páginas 849–852, June 2017, ISBN 978-1-5386-0752-7. 20, 23, 27
 - [92] Antonić, Aleksandar, Martina Marjanović, Krešimir Pripužić e Ivana Podnar Žarko: *A mobile crowd sensing ecosystem enabled by cupus: Cloud-based publish/subscribe middleware for the internet of things*. Future Generation Computer Systems, 56:607 – 622, 2016. 20, 22, 23, 28
 - [93] Yin, Junjun, Yizhao Gao e Shaowen Wang: *Cybergis-Enabled Urban Sensing from Volunteered Citizen Participation using Mobile Devices*. Em *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*, páginas 83–96. Springer International Publishing, 2017. 20, 23, 28
 - [94] Kraft, Robin, Ferdinand Birk, Manfred Reichert, Aniruddha Deshpande, Winfried Schlee, Berthold Langguth, Harald Baumeister, Thomas Probst, Myra Spiliopoulou e Rüdiger Pryss: *Design and Implementation of a Scalable Crowdsensing Platform for Geospatial Data of Tinnitus Patients*. Em *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, páginas 294–299. IEEE, June 2019, ISBN 978-1-7281-2286-1. 20, 21, 22, 23, 28
 - [95] Le-Phuoc, Danh, Hoan Nguyen Mau Quoc, Hung Ngo Quoc, Tuan Tran Nhat e Manfred Hauswirth: *The Graph of Things: A step towards the Live Knowledge Graph of connected things*. Journal of Web Semantics, 37–38:25–35, 2016. 20

- [96] Limkar, Suresh V. e Rakesh Kumar Jha: *A novel method for parallel indexing of real time geospatial big data generated by IoT devices*. Future Generation Computer Systems, 97:433 – 452, 2019, ISSN 0167-739X. 20, 22, 23, 25, 42
- [97] Zhou, Lianjie, Qingquan Li e Wei Tu: *An Efficient Access Model of Massive Spatiotemporal Vehicle Trajectory Data in Smart City*. IEEE Access, 8:52452–52465, 2020, ISSN 2169-3536. 20, 21, 23, 24, 25, 27, 28, 64
- [98] Lwin, Ko Ko, Wataru Takeuchi, Yoshihide Sekimoto e Koji Zettsu: *City Geospatial Dashboard: IoT and Big Data Analytics for Geospatial Solutions Provider in Disaster Management*. Em *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, páginas 1–4. IEEE, dezembro 2019. 20, 23, 25, 26, 28, 37
- [99] Laska, Marius, Stefan Herle, Ralf Klamma e Jörg Blankenbach: *A Scalable Architecture for Real-Time Stream Processing of Spatiotemporal IoT Stream Data-Performance Analysis on the Example of Map Matching*. ISPRS International Journal of Geo-Information, 7(7, 238):238, 2018. 20, 22, 23, 26, 28
- [100] Yang, Jiachen, Yurong Han, Yafang Wang, Bin Jiang, Zhihan Lv e Houbing Song: *Optimization of Real-Time Traffic Network Assignment based on IoT Data using DBN and Clustering Model in Smart City*. Future Generation Computer Systems, 108:976–986, 2020. 20, 23
- [101] You, Linlin, Bige Tunçer, Rui Zhu, Hexu Xing e Chau Yuen: *A Synergetic Orchestration of Objects, Data, and Services to Enable Smart Cities*. IEEE Internet of Things Journal, 6(6):10496–10507, dezembro 2019, ISSN 2327-4662. 20, 23, 28
- [102] Qader, Mohiuddin Abdul e Vagelis Hristidis: *DualDB: An Efficient LSM-based Publish/Subscribe Storage System*. Em *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, 2017, ISBN 9781450352826. 21, 23
- [103] Hulbert, Andrew, Thomas Kunicki, James N. Hughes, Anthony D. Fox e Christopher N. Eichelberger: *An Experimental Study of Big Spatial Data Systems*. Em *2016 IEEE International Conference on Big Data (Big Data)*, páginas 2664–2671, dezembro 2016, ISBN 978-1-4673-9005-7. 21, 23
- [104] Hongming Cai and Shuai Luan and Lihong Jiang and Nazaraf Shah and Ray Farmer and Kuo-Ming Chao and Boyi Xu: *A multi-layer internet of things database schema for online-to-offline systems*. International Journal of Distributed Sensor Networks, 12(8), 2016. 21, 22, 23
- [105] Luan, Shuai, Hongming Cai, Fenglin Bu e Lihong Jiang: *A Four-Layer Flexible Spatial Data Framework towards IoT Application*. Em *2015 IEEE 12th International Conference on e-Business Engineering (ICEBE)*, páginas 339–346. IEEE, 2015, ISBN 978-1-4673-8002-7. 21, 23, 24

- [106] Sun, Ke, Junping Zhang e Yingying Zhang: *Roads and Intersections Extraction from High-Resolution Remote Sensing Imagery Based on Tensor Voting under Big Data Environment*. Wireless Communications and Mobile Computing, 2019, 2019, ISSN 1530-8669. 21, 23
- [107] Wu, Chen, Qing Zhu, Yeting Zhang, Zhiqiang Du, Xinyue ye, Han Qin e Yan Zhou: *A NoSQL-SQL Hybrid Organization and Management Approach for Real-Time Geospatial Data: A Case Study of Public Security Video Surveillance*. ISPRS International Journal of Geo-Information, 6(1):21, janeiro 2017, ISSN 2220-9964. 21, 23, 24
- [108] Open Geospatial Consortium, Inc. (OGC): *Keyhole Markup Language*, 2008. <https://developers.google.com/kml>, Acessado em 21 de janeiro de 2025. 21
- [109] Open Geospatial Consortium, Inc. (OGC): *Geography Markup Language (GML)*, 2000. <https://www.ogc.org/standards/gml>, Acessado em 21 de janeiro de 2025. 21
- [110] Crockford, Douglas: *JavaScript Object Notation (JSON)*, 2000. <https://www.json.org/>, Acessado em 21 de janeiro de 2025. 21
- [111] Leung, Carson K., Peter Braun e Alfredo Cuzzocrea: *AI-Based Sensor Information Fusion for Supporting Deep Supervised Learning*. Sensors, 19(6), março 2019, ISSN 1424-8220. 21, 23, 24, 28
- [112] Butler, Howard, Martin Daly, Allan Doyle, Sean Gillies, Stefan Hagen e Tim Schaub: *The GeoJSON Format*. Relatório Técnico 7946, Internet Engineering Task Force (IETF), agosto 2016. 21
- [113] Neo4j, Inc.: *Neo4j Graph Platform - The Leader in Graph Databases*, 2020. <https://neo4j.com/>. 21
- [114] Quoc, Hoan Nguyen Mau e Danh Le-Phuoc: *An Elastic and Scalable Spatiotemporal Query Processing for Linked Sensor Data*. Em *Proceedings of the 11th International Conference on Semantic Systems*, página 17–24, 2015, ISBN 9781450334624. 21, 23, 27, 28
- [115] Iyer, Anand Padmanabha e Ion Stoica: *A Scalable Distributed Spatial Index for the Internet-of-Things*. Em *Proceedings of the 2017 Symposium on Cloud Computing, SoCC '17*, página 548–560, New York, NY, USA, 2017. Association for Computing Machinery, ISBN 9781450350280. 21, 23, 26
- [116] Xia, Dawen, Binfeng Wang, Huaqing Li, Yantao Li e Zili Zhang: *A Distributed Spatial - Temporal Weighted Model on MapReduce for Short-Term Traffic Flow Forecasting*. Neurocomputing, 179(C):246–263, fevereiro 2016, ISSN 0925-2312. 22, 23, 27
- [117] Sukmaningsih, Dyah Wahyu, Wayan Suparta, Agung Trisetyarso, Bahtiar Saleh Abbas e Chul Ho Kang: *Proposing Smart Disaster Management in Urban Area*, páginas 3–16. Springer International Publishing, Cham, 2020, ISBN 978-3-030-14132-5. 22, 23, 27, 28

- [118] Duan, Xingyin, Yimeng Chen, Jingwen Xu e Ruolan Jiang: *Analysis of Influencing Factors on Urban Traffic Congestion and Prediction of Congestion Time based on Spatiotemporal Big Data*. Em *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, páginas 75–78, 2020. 22, 23
- [119] Coşkun, İ. B., A. Çakır e B. Anbaroğlu: *Performance Matters on Identification of Origin-Destination Matrix on Big Geospatial Data*. Em *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 43, páginas 449–454, 2020. 22, 23, 28
- [120] White, Tom: *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 4th edição, 2015, ISBN 1491901632. 22
- [121] Sambrekar, Kuldeep P. e Vijay S. Rajpurohit: *Fast and Efficient Agro Data Classification Model for Agriculture Management System using Hierarchical Cloud Computing*. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(4S):387–394, fevereiro 2019. 22, 23, 24
- [122] Fang, Shifeng, Yunqiang Zhu, Lida Xu, Jinqu Zhang, Peiji Zhou, Kan Luo e Jie Yang: *An integrated system for land resources supervision based on the IoT and cloud computing*. *Enterprise Information Systems*, 11(1):105–121, 2015. 22, 23, 28
- [123] Hayes, Michael A. e Miriam A.M. Capretz: *Contextual Anomaly Detection in Big Sensor Data*. Em *2014 IEEE International Congress on Big Data*, páginas 64–71, junho 2014, ISBN 978-1-4799-5057-7. 22, 23, 24
- [124] Apache Software Foundation: *HDFS Architecture Guide*, 2006. https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html, Acessado em 21 de janeiro de 2025. 22
- [125] Al Jawarneh, Isam Mashhour, Paolo Bellavista, Luca Foschini e Rebecca Montanari: *Spatial-Aware Approximate Big Data Stream Processing*. Em *2019 IEEE Global Communications Conference (GLOBECOM)*, páginas 1–6. IEEE, dec 2019. 22, 23, 28
- [126] Apache Software Foundation: *Apache Storm*, 2011. <https://storm.apache.org/>. 22
- [127] Apache Software Foundation: *Apache Kafka*, 2011. <https://kafka.apache.org/>. 22
- [128] Chen, Yuting, Qi Jing, Ling Xiao, Yulin Ding, Mingyuan Hu, Weitao Che e Hui Lin: *A multi-level situational awareness method with dynamic multi-modal data visualization for air pollution monitoring*. Em *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, páginas 489–492, 2022. 23
- [129] Muninarayanappa, Vimala e Rajeev Ranjan: *Agriculture Data Analysis using Parallel k-Nearest Neighbour Classification Algorithm*. *International Journal of Reconfigurable and Embedded Systems*, 13(2):332 – 340, 2024, ISSN 20894864. 23

- [130] Friji, Hamdi, Abdullah Khanfor, Hakim Ghazzai e Yehia Massoud: *An End-to-End Smart IoT-Driven Navigation for Social Distancing Enforcement*. IEEE Access, 10:76824–76841, 2022. 23
- [131] Panigrahi, Sunil K., Veena Goswami, Hemant K. Apat, Rabindra K. Barik, Ankit Vidyarthi, Punit Gupta e Meshal Alharbi: *An interconnected IoT-inspired network architecture for data visualization in remote sensing domain*. Alexandria Engineering Journal, 81:17–28, 2023. 23
- [132] Yin, Weibin, Zhongjie Han, Yueliang Feng, Gang Zhou e Yunchao Shen: *Application of Key Technologies of Distributed Storage Based on the Internet of Things in Urban Fire Protection*. Scientific Programming, 2022(1), 2022. 23
- [133] Bonilla, Lander, Juan López-De-Armentia, Gorka Zarate, Ana I. Torre-Bastida, Luis Vigo, Peru Bidaguren, Olga Alonso e Abel Capelastegui: *CCIR: An Architecture for Collecting and Storing Connected Corridor Infrastructure and Mobility Data*. Em *2023 IEEE 26th International Conference on Intelligent Transportation Systems, ITSC 2023*, IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, páginas 5344–5350, United States, 2023. Institute of Electrical and Electronics Engineers Inc. 23
- [134] Sarramia, David, Alexandre Claude, Francis Ogereau, Jérémy Mezhoud e Gilles Mailhot: *CEBA: A Data Lake for Data Sharing and Environmental Monitoring*. Sensors, 22(7), 2022. 23
- [135] Bovkir, Rabia e Arif Cagdas Aydinoglu: *Conceptual modelling of sensor-based geographic data: interoperable approach with real-time air quality index (AQI) dashboard*. Earth Science Informatics, 17(6):5317–5333, Dec 2024, ISSN 1865-0481. <https://doi.org/10.1007/s12145-024-01444-x>. 23
- [136] Zhang, X., X. Gong, Y. N. Du, T. Dong, Q. Wang, B. B. Li e X. R. Yan: *Construction and Application of the Geotechnical and Geological Spatiotemporal Big Data Platform in Xiongan New Area*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-1/W2-2023:323–329, 2023. 23
- [137] Pandya, Sharnil, Hemant Ghayvat, Praveen Kumar Reddy, Thippa Reddy Gadekallu, Muhammad Ahmed Khan e Neeraj Kumar: *COUNTERSAVIOR: AIoMT and IIoT-Enabled Adaptive Virus Outbreak Discovery Framework for Healthcare Informatics*. IEEE Internet of Things Journal, 10(5):4202–4212, 2023. 23
- [138] Sun, Ying: *Development and application of urban intelligent emergency platform system*. Em *ISCTT 2022; 7th International Conference on Information Science, Computer Technology and Transportation*, páginas 1–7, 2022. 23
- [139] Wang, Guang, Yuefei Chen, Shuai Wang, Fan Zhang e Desheng Zhang: *ForE-Taxi: Data-Driven Fleet-Oriented Charging Resource Allocation in Large-Scale Electric Taxi Networks*. ACM Transactions on Sensor Networks, 19(3), março 2023, ISSN 1550-4859. 23, 27

- [140] Das, Jaydeep, Shreya Ghosh, Anwesha Mukherjee, Soumya K. Ghosh e Rajkumar Buyya: *RESCUE: Enabling green healthcare services using integrated IoT-edge-fog-cloud computing environments*. Software: Practice and Experience, 52(7):1615–1642, 2022. 23
- [141] Tang, Ke, Lan Cheng, Xiao Wang, Chenqiang Pan, Liang Lin, Li Dong e Bo Li: *Research on Digital Monitoring Technology for Airport High-Pressure Rotary Jet Piles*. IEEE Access, 11:112731–112752, 2023. 23
- [142] Zhang, Tianyu, Xin Jin, Song Bai, Yuxin Peng, Ye Li e Jun Zhang: *Smart Public Transportation Sensing: Enhancing Perception and Data Management for Efficient and Safety Operations*. Sensors, 23(22), 2023, ISSN 1424-8220. 23
- [143] Ghayvat, Hemant, Sharnil Pandya, Muhammad Awais e Kapal Dev: *STRENUOUS: Edge-Line Computing, AI, and IIoT Enabled GPS Spatiotemporal Data-Based Meta-Transmission Healthcare Ecosystem for Virus Outbreaks Discovery*. IEEE Internet of Things Journal, 10(4):3285–3294, 2023. 23
- [144] Pan, Yu, Jianxin Sun, Hongfeng Yu, Geng Bai, Yufeng Ge, Joe Luck e Tala Awada: *Transforming Agriculture with Intelligent Data Management and Insights*. Em *2023 IEEE International Conference on Big Data (BigData)*, páginas 3489–3498, 2023. 23
- [145] Fang, Shifeng, Li Da Xu, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan e Zhihui Liu: *An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things*. IEEE Transactions on Industrial Informatics, 10(2):1596–1605, May 2014, ISSN 1941-0050. 22, 23
- [146] Min, Huang: *Analysis and Design of Intelligent Environmental Protection IOT System Architecture Based on GIS*. Journal of Physics: Conference Series, 1616:012084, Aug 2020. 23
- [147] Lu, Chunfang, Junfei Liu, Yanhong Liu e Yuming Liu: *Intelligent Construction Technology of Railway Engineering in China*. Frontiers of Engineering Management, 6(4):503–516, dezembro 2019, ISSN 2095-7513. 23
- [148] Terziyski, Atanas, Stoyan Tenev, Vedrin Jeliazkov, Nina Jeliazkova e Nikolay Kochev: *METER.AC: Live open access atmospheric monitoring data for bulgaria with high spatiotemporal resolution*. Data, 5(2):36, 2020. 22, 23, 27
- [149] Wu, Qiulan, Yong Liang, Ying Li e Yusheng Liang: *Research on intelligent acquisition of smart agricultural big data*. Em *2017 25th International Conference on Geoinformatics*, páginas 1–7. IEEE, agosto 2017, ISBN 978-1-5386-2266-7. 23
- [150] Zhao, Zhuofeng, Weilong Ding, Jianwu Wang e Yanbo Han: *A Hybrid Processing System for Large-Scale Traffic Sensor Data*. IEEE Access, 3:2341–2351, 2015. 23
- [151] Lorenz, Felix, Morgan Geldenhuys, Harald Sommersy, Frauke Jakobs, Carsten Lüring, Volker Skwarek, Ilja Behnke e Lauritz Thamsen: *A Scalable and Dependable*

- Data Analytics Platform for Water Infrastructure Monitoring*. Em *2020 IEEE International Conference on Big Data (Big Data)*, páginas 3488–3493, Dec 2020. 23, 28
- [152] Nwankwo, Wilson e Kingsley Eghonghon Ukhurebor: *Big Data Analytics: A Single Window IoT-Enabled Climate Variability System for All-year-round Vegetable Cultivation*. IOP Conference Series: Earth and Environmental Science, 655(1), 2021. 23
- [153] Velásquez, R.M. Arias, Y.L. Romero Ramos e J. Noel: *Citizen Science Approach for Spatiotemporal Modelling of Air Pollution Quality and Traffic in Lima, Peru*. Em *2019 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, páginas 1–4, 2019. 23
- [154] Dao, Minh Son, Ngoc Thanh Nguyen, R. Uday Kiran e Koji Zettsu: *Fusion-3DCNN-max3P: A Dynamic System for Discovering Patterns of Predicted Congestion*. Em *2020 IEEE International Conference on Big Data (Big Data)*, páginas 910–915, Dec 2020. 22, 23, 27
- [155] Chen, Jian, Yan Wang e Xuepeng Jiang: *Research on Key Technologies Of Equipment Support Intelligent System*. Em *2020 IEEE International Conference on Big Data (Big Data)*, páginas 910–915, Dec 2020. 23
- [156] Xu, Cheng, Hengjie Luo, Hong Bao e Pengfei Wang: *STEIM: A Spatiotemporal Event Interaction Model in V2X Systems Based on a Time Period and a Raster Map*. Mobile Information Systems, 2020, Feb 2020. 23
- [157] Lee, K., L. Liu, R. K. Ganti, M. Srivatsa, Q. Zhang, Y. Zhou e Q. Wang: *Lightweight indexing and querying services for big spatial data*. IEEE Transactions on Services Computing, 12(3):343–355, 2019. 23, 27, 33
- [158] Naseer, Salman, William Liu, Nurul I. Sarkar, Muhammad Shafiq e Jin Ghoo Choi: *Smart city taxi trajectory coverage and capacity evaluation model for vehicular sensor networks*. Sustainability (Switzerland), 13(19), 2021. 22, 23, 28
- [159] Chen, Guangsheng, Weitao Zou, Weipeng Jing, Wei Wei e Rafał Scherer: *Improving the efficiency of the ems-based smart city: A novel distributed framework for spatial data*. IEEE Transactions on Industrial Informatics, 19(1):594–604, 2023. 22, 23
- [160] Bujari, Armir, Alessandro Calvio, Luca Foschini, Andrea Sabbioni e Antonio Corradi: *A digital twin decision support system for the urban facility management process*. Sensors, 21(24), 2021. 22, 23
- [161] Shah, Pushpalata: *Igis for managing cities smartly - urban geoinformatics*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 46(4/W5-2021):509–513, 2021. 23
- [162] Hayat, Haroun e Ghomari Abdessamed Réda: *An IoT and spatial Big data based architecture for monitoring Occupational Health Risks exposure*. Em *2022 First International Conference on Big Data, IoT, Web Intelligence and Applications (BIWA)*, páginas 13–18, Dec 2022. 23

- [163] Russell-Pavier, Frederick S, Suresh Kaluvan, David Megson-Smith, Dean T Connor, Samuel J Fearn, Euan L Connolly, Thomas B Scott e Peter G Martin: *A highly scalable and autonomous spectroscopic radiation mapping system with resilient IoT detector units for dosimetry, safety and security*. Journal of Radiological Protection, 43(1), 2023. 23
- [164] Lee, Chunggi, Yeonjun Kim, Seungmin Jin, Dongmin Kim, Ross Maciejewski, David Ebert e Sungahn Ko: *A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion*. IEEE Transactions on Visualization and Computer Graphics, 26(11):3133–3146, 2020. 23
- [165] Felemban, Emad, Faizan Ur Rehman, Asad Ali Biabani, Atif Naseer, Omar Hussain e Ehsan Ullah Warriach: *An Interactive System for Analyzing Movement of Buses in Hajj*. Journal of Theoretical and Applied Information Technology, 98(21):3468–3481, 2020. 23
- [166] Wang, Shaowen, Fangzheng Lyu, Shaohua Wang, Charles E. Catlett, Anand Padmanabhan e Kiumars Soltani: *Integrating CyberGIS and Urban Sensing for Reproducible Streaming Analytics*, páginas 663–681. Springer Singapore, Singapore, 2021, ISBN 978-981-15-8983-6. 23, 28
- [167] Maarala, Altti Ilari, Mika Rautiainen, Miikka Salmi, Susanna Pirttikangas e Jukka Rieki: *Low Latency Analytics for Streaming Traffic Data with Apache Spark*. Em *2015 IEEE International Conference on Big Data (Big Data)*, páginas 2855–2858, 2015. 23, 28
- [168] Bhuvaneswari, Anbalagan e Chinnaiah Valliyammai: *Social IoT-Enabled Emergency Event Detection Framework using Geo-Tagged Microblogs and Crowdsourced Photographs*. Em *Emerging Technologies in Data Mining and Information Security*, páginas 151–162. Springer Singapore, 2019. 23
- [169] Wang, Zuchao, Min Lu, Xiaoru Yuan, Junping Zhang e Huub van de Wetering: *Visual Traffic Jam Analysis Based on Trajectory Data*. IEEE Transactions on Visualization and Computer Graphics, 19(12):2159–2168, 2013. 23
- [170] Cao, Hung e Monica Wachowicz: *The Design of an IoT-GIS Platform for Performing Automated Analytical Tasks*. Computers, Environment and Urban Systems, 74:23–40, 2019, ISSN 0198-9715. 22, 23, 28
- [171] Zadeh, Reza Bairam, Atabak Elmi, Valeh Moghaddam e Somaiyeh MahmoudZadeh: *A conceptual high level multiagent system for wildfire management*. IEEE Transactions on Geoscience and Remote Sensing, 63:1–15, 2025. 22, 23
- [172] Lakshmi, T.K., Naresh Kumar Mangalapu, Mary Fathima Anuja Joseph Swaminathan, S. Chandravadhana, R. Pari e T. Ranjana: *Ai for disaster response: Developing intelligent systems for real-time emergency management*. Em *2025 International Conference on Next Generation Communication Information Processing (INCIP)*, páginas 868–874, Jan 2025. 23

- [173] Ramesh, Azmeera, Zufarova Gulmira, Ammar Hameed Shnain, R. Ramya, P. Nagaveni e Aanandha Saravanan K: *Big data and machine learning for climate change prediction: An integrated approach to environmental monitoring*. Em *2025 International Conference on Automation and Computation (AUTOCOM)*, páginas 1384–1389, March 2025. 22, 23
- [174] Zhou, Lianjie, Wei Tu, Qingquan Li e Dongjie Guan: *A Heterogeneous Streaming Vehicle Data Access Model for Diverse IoT Sensor Monitoring Network Management*. *IEEE Internet of Things Journal*, 11(16):26929–26943, 2024. 23
- [175] Deshpande, Kartik Vishal e Dheeraj Kumar: *A Novel Algorithm for Evaluating Clustering Propensity of IoT-Generated Spatio-Temporal Data Geared for Distributed Systems*. *IEEE Sensors Journal*, 24(14):21949–21963, 2024. 23
- [176] Xu, Qi, Longgang Xiang, Huayi Wu, Chuang Tao, Haocheng Wang, Liebing Yu, Quan Liu e Xumei Wang: *Dynamic heatmap pyramid computation for massive high-parallel spatial streaming in urban environments*. *International Journal of Digital Earth*, 17(1):2368099, 2024. 23
- [177] Rammer, Daniel, Thilina Buddhika, Matthew Malensek, Shrideep Pallickara e Sangmi Lee Pallickara: *Enabling Fast Exploratory Analyses Over Voluminous Spatiotemporal Data Using Analytical Engines*. *IEEE Transactions on Big Data*, 8(1):213–228, 2022. 23
- [178] Mete, Muhammed: *Geospatial Big Data Analytics for Sustainable Smart Cities*. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W7-2023:141–146, junho 2023. 23
- [179] Wang, En, Mijia Zhang, Bo Yang, Yongjian Yang e Jie Wu: *Large-scale spatiotemporal fracture data completion in sparse crowdsensing*. *IEEE Transactions on Mobile Computing*, 23(7):7585–7601, 2024. 23, 27
- [180] Zhang, Jin, Xun Xu e Liu Ding: *Regional Traffic Flow Prediction Model Based on Mobile Signaling Data- A Case Study of Chongqing City*. Em *2023 7th International Conference on Transportation Information and Safety (ICTIS)*, páginas 475–482, 2023. 23
- [181] Zhou, Bao, Junsan Zhao, Guoping Chen e Ying Yin: *Security Authentication Mechanism of Spatio-Temporal Big Data Based on Blockchain*. *Applied Sciences*, 13(11), 2023, ISSN 2076-3417. 23
- [182] Lymperis, Dimitrios e Christos Goumopoulos: *SEDIA: A Platform for Semantically Enriched IoT Data Integration and Development of Smart City Applications*. *Future Internet*, 15(8), 2023, ISSN 1999-5903. 23
- [183] Wang, Jinbao, Zuobin Xiong, Qilong Han, Xixian Han e Donghua Yang: *Top-k Socially Constrained Spatial Keyword Search in Large SIoT Networks*. *IEEE Internet of Things Journal*, 9(12):9280–9289, 2022. 23, 28

- [184] Ma, Wenjie, Yanguo Peng, Ximeng Liu e Jiangtao Cui: *VeriRange: A Verifiable Range Query Model on Encrypted Geographic Data for IoT Environment*. IEEE Internet of Things Journal, 11(2):3068–3081, 2024. 23, 28
- [185] Lv, Zhihan, Xiaoming Li, Haibin Lv e Wenqun Xiu: *BIM Big Data Storage in WebVRGIS*. IEEE Transactions on Industrial Informatics, 16(4):2566–2573, apr 2019. 23, 25, 26, 28
- [186] Sun, Wei, Jun Zhu, Ning Duan, Peng Gao, Guo Qiang Hu, Wei Shan Dong, Zhi Hu Wang, Xin Zhang, Peng Ji, Chun Yang Ma e Jing Chang Huang: *Moving Object Map Analytics: A Framework Enabling Contextual Spatial-Temporal Analytics of Internet of Things Applications*. Em *Proceedings of the 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, páginas 101–106. IEEE, julho 2016, ISBN 978-1-5090-2927-3. 23, 28
- [187] Wu, Siqi, Luca Morandini e Richard O. Sinnott: *Smash: A cloud-based architecture for big data processing and visualization of traffic data*. Em *2015 IEEE International Conference on Data Science and Data Intensive Systems*, páginas 53–60, Dec 2015. 23, 27, 28
- [188] Hu, Linshu, Feng Zhang, Mengjiao Qin, Zhiyi Fu, Zhende Chen, Zhenhong Du e Renyi Liu: *A Dynamic Pyramid Tilling Method for Traffic Data Stream Based on Flink*. IEEE Transactions on Intelligent Transportation Systems, 2021. 22, 23
- [189] Liu, Zhaoman, Lei Wu, Weizhi Meng, Hao Wang e Wei Wang: *Accurate range query with privacy preservation for outsourced location-based service in iot*. IEEE Internet of Things Journal, páginas 1–1, 2021. 23
- [190] Li, Shiqiang, Weize Wang, Jiawei Shan, Heng Qi, Yanming Shen e Baocai Yin: *An Effective Spatio-Temporal Query Framework for Massive Trajectory Data in Urban Computing*. Em *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, volume 2019-December, páginas 586–593, 2019. 23, 28
- [191] Baig, Furqan, Dejun Teng, Jun Kong e Fusheng Wang: *SPEAR: Dynamic Spatio-Temporal Query Processing over High Velocity Data Streams*. Em *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, páginas 2279–2284, April 2021. 23, 27, 28
- [192] Chen, Yan, Tian Shu, Xiaokang Zhou, Xuzhe Zheng, Akira Kawai, Kaoru Fueda, Zheng Yan, Wei Liang e Kevin I Kai Wang: *Graph Attention Network With Spatial-Temporal Clustering for Traffic Flow Forecasting in Intelligent Transportation System*. IEEE Transactions on Intelligent Transportation Systems, páginas 1–11, 2022, ISSN 1558-0016. 22, 23, 28
- [193] Leite, Daniel, Igor Škrjanc, Sašo Blažič, Andrej Zdešar e Fernando Gomide: *Interval incremental learning of interval data streams and application to vehicle tracking*. Information Sciences, 630:1–22, 2023. 22, 23, 28

- [194] Barik, Rabindra Kumar, Sudhansu Shekhar Patra, Priva Kumari, Sachi Nandan Mohanty e Abdulsattar Abdullah Hamad: *A New Energy Aware Task Consolidation Scheme for Geospatial Big Data Application in Mist Computing Environment*. Em *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, páginas 48–52, March 2021. 23
- [195] Cheng, Hongju, Yushi Shi, Leihuo Wu, Yingya Guo e Naixue Xiong: *An intelligent scheme for big data recovery in Internet of Things based on Multi-Attribute assistance and Extremely randomized trees*. *Information Sciences*, 557:66–83, MAY 2021, ISSN 0020-0255. 23
- [196] Zhou, Lianjie, Wei Tu, Chisheng Wang e Qingquan Li: *A heterogeneous access metamodel for efficient iot remote sensing observation management: Taking precision agriculture as an example*. *IEEE Internet of Things Journal*, 9(11):8616–8632, JUN 1 2022, ISSN 2327-4662. 23
- [197] Chen, Yuting, Qi Jing, Ling Xiao, Yulin Ding, Mingyuan Hu, Weitao Che e Hui Lin: *A multi-level situational awareness method with dynamic multi-modal data visualization for air pollution monitoring*. Em *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, páginas 489–492, July 2022. 23, 28
- [198] Zhu, Shunzhi, Dahan Wang, Lijuan Liu, Yan Wang e Danhuai Guo: *Inferring Region Significance by using Multi-Source Spatial Data*. *Neural Computing and Applications*, 32(11):6523–6531, 2019. 23
- [199] Shen, Jun Hong, Mu Yen Chen, Ching Ta Lu e Rou Hua Wang: *Monitoring Spatial Keyword Queries Based on Resident Domains of Mobile Objects in IoT Environments*. *Mobile Networks and Applications*, 2020. 23
- [200] Hagan, Ronald D., Charles A. Phillips, Michael A. Langston e Bradley J. Rhodes: *Multiscale Graph Theoretical Tools Reveal Subtle Patterns in Big Geospatial Data*. Em *2017 IEEE International Conference on Big Data (Big Data)*, volume 2018-January de *IEEE International Conference on Big Data*, páginas 3422–3425, 2017. 23
- [201] Wu, Xuesong, Mengyun Xu, Jie Fang e Xiongwei Wu: *A Multi-Attention Tensor Completion Network for Spatiotemporal Traffic Data Imputation*. *IEEE Internet of Things Journal*, 9(20):20203–20213, 2022. 23
- [202] Chen, Xiaobo, Shurong Liang, Zhihao Zhang e Feng Zhao: *A novel spatiotemporal data low-rank imputation approach for traffic sensor network*. *IEEE Internet of Things Journal*, 9(20):20122–20135, 2022. 23
- [203] Du, Xiaoxin, Hui Zhu, Yandong Zheng, Rongxing Lu, Fengwei Wang e Hui Li: *A Semantic-Preserving Scheme to Trajectory Synthesis Using Differential Privacy*. *IEEE Internet of Things Journal*, 10(15):13784–13797, 2023. 23
- [204] Cai, Sujin, Xin Lyu, Xin Li, Duohan Ban e Tao Zeng: *A Trajectory Released Scheme for the Internet of Vehicles Based on Differential Privacy*. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):16534–16547, 2022. 23

- [205] Mashhadi, Afra, Ali Tabaraei, Yuting Zhan e Reza M. Parizi: *An Auditing Framework for Analyzing Fairness of Spatial-Temporal Federated Learning Applications*. Em *2022 IEEE World AI IoT Congress (AIIoT)*, páginas 699–707, 2022. 23
- [206] Ma, Haoyuan, Mintao Zhou, Xiaodong Ouyang, Du Yin, Renhe Jiang e Xuan Song: *Forecasting Regional Multimodal Transportation Demand with Graph Neural Networks: An Open Dataset*. Em *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, páginas 3263–3268, 2022. 23
- [207] Chen, Yan, Tian Shu, Xiaokang Zhou, Xuzhe Zheng, Akira Kawai, Kaoru Fueda, Zheng Yan, Wei Liang e Kevin I Kai Wang: *Graph Attention Network With Spatial-Temporal Clustering for Traffic Flow Forecasting in Intelligent Transportation System*. *IEEE Transactions on Intelligent Transportation Systems*, 24(8):8727–8737, 2023. 23
- [208] Srinivas, Kunal G e Nagendra Nath Giri: *K-Means Clustering driven Deep Spatiotemporal Learning Model for PM2.5 Prediction*. Em *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICD-CECE)*, páginas 1–8, 2022. 23
- [209] Abidi, Jassuer e Fethi Filali: *Mobility Analytics of Fans During the 2021 FIFA Arab Cup Football Tournament in Qatar*. *IEEE Open Journal of Intelligent Transportation Systems*, 4:653–662, 2023. 23
- [210] Ding, Weilong, Tianpu Zhang, Honghao Gao, Qi Yu, Jianwu Wang e Zhuofeng Zhao: *Multi-Graph Spatio-Temporal Convolution for Traffic Flow Prediction Focusing on Edge Derived Imbalanced Data From Highway Electronics*. *IEEE Transactions on Consumer Electronics*, páginas 1–1, 2024. 23
- [211] Santos Sá, Joiner dos, Edinho do Nascimento da Silva, Leonardo Nunes Gonçalves, Caio Mateus Machado Cardoso, Andréia Antloga do Nascimento, Gervásio Protásio dos Santos Cavalcante, Maria Emília de Lima Tostes, Jasmine Priscyla Leite de Araújo, Fabricio José Brito Barros e Fabricio de Souza Farias: *Multimodal urban mobility solutions for a smart campus using artificial neural networks for route determination and an algorithm for arrival time prediction*. *Engineering Applications of Artificial Intelligence*, 137:109074, 2024, ISSN 0952-1976. 23
- [212] Musa, Usman Shuaibu, Yifan Guo, Cheng QianT e Wei Yu: *Optimal sampling for Moving Object Trajectory Tracking in Smart Transportation Systems: A Transformer-based Approach*. Em *2023 IEEE International Conference on Big Data (BigData)*, páginas 1230–1235, Los Alamitos, CA, USA, dezembro 2023. IEEE Computer Society. 23
- [213] Dewage, Prabuddha M. H., Lakitha O. H. Wijeratne, Xiaohe Yu, Mazhar Iqbal, Gokul Balagopal, John Waczak, Ashen Fernando, Matthew D. Lary, Shisir Ruwali e David J. Lary: *Providing Fine Temporal and Spatial Resolution Analyses of Airborne Particulate Matter Utilizing Complimentary In Situ IoT Sensor Network and Remote Sensing Approaches*. *Remote Sensing*, 16(13):2454, julho 2024. 23

- [214] Mo, Wen, Zeyuan Li, Zhiwen Zeng, Neal N. Xiong, Shaobo Zhang e Anfeng Liu: *SCTD: A spatiotemporal correlation truth discovery scheme for security management of data platform*. *Future Generation Computer Systems*, 139(C):109–125, fevereiro 2023, ISSN 0167-739X. 23
- [215] Wang, En, Mijia Zhang, Yuanbo Xu, Haoyi Xiong e Yongjian Yang: *Spatiotemporal Fracture Data Inference in Sparse Urban CrowdSensing*. Em *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, páginas 1499–1508, 2022. 23
- [216] Chatterjee, Kalyan, Muntha Raju, Machakanti Navya Thara, Mandadi Sriya Reddy, M. Priyadharshini, N. Selvamuthukumar, Saurav Mallik, Haya Mesfer Alshahrani, Mohamed Abbas e Ben Othman Soufiene: *Toward Cleaner Industries: Smart Cities' Impact on Predictive Air Quality Management*. *IEEE Access*, 12:78895–78910, 2024. 23
- [217] Nanda, Sandeep, Veena Goswami, Aditya Narayan Brahma, Sudhansu Shekhar Patra e Rabindra K. Barik: *Towards Efficient and Dynamic Allocations of Mist Nodes for IoST Devices*. Em *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, páginas 1–5, 2022. 23
- [218] Wang, Jianwei, Mengyan Cui, Huimin Wang, Hao Yang, Xiaowei Guo, Xuxu Liu e Xin Fu: *Trip Purpose Inference and Spatio-Temporal Characterization Based on Anonymized Trip Data: Empirical Study from Dockless Shared Bicycle Dataset in Xi'an, China*. *Transportation Research Record*, 2678(9):251–266, 2024. 23
- [219] Qiu, T., J. Liu, W. Si, M. Han, H. Ning e M. Atiquzzaman: *A data-driven robustness algorithm for the internet of things in smart cities*. *IEEE Communications Magazine*, 55(12):18–23, 2017. 23, 24
- [220] Song, X., C. Wang e J. Gao: *An integrated framework for analysis and mining of the massive sensor data using feature preserving strategy on cloud computing*. Em *2014 Seventh International Symposium on Computational Intelligence and Design*, volume 2, páginas 337–340, Dec 2014. 23
- [221] Chen, Zhaoya, Wei Huang, Lijun Ma, Hao Xu e Yaheng Chen: *Application and Development of Big Data in Sustainable Utilization of Soil and Land Resources*. *IEEE Access*, 8:152751–152759, 2020. 23
- [222] Zhang, Zhang e Liang Tian: *Data Visualization Association Algorithm Based on Spatio-Temporal Association Big Data*. Em *The 2020 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy*, páginas 314–319. Springer International Publishing, 2021. 23
- [223] Song, Jun, Ke Han e Marc Stettler: *Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing*. *IEEE Internet of Things Journal*, páginas 1–1, 2020. 23
- [224] Zhong, Tao, Kshitij Doshi, Gang Deng, Xiaoming Yang e Hegao Zhang: *High volume geospatial mapping for internet-of-vehicle solutions with in-memory map-reduce processing*. Em *2014 IEEE International Conference on Big Data (Big Data)*, páginas 802–807. IEEE, outubro 2014. 23

- [225] Chiwewe, Tapiwa M. e Jeofrey Ditsela: *Machine learning based estimation of ozone using spatio-temporal data from air quality monitoring stations*. Em *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, páginas 58–63, July 2016. 23, 24
- [226] Yang, Jiachen, Yurong Han, Yafang Wang, Bin Jiang, Zhihan Lv e Houbing Song: *Optimization of real-time traffic network assignment based on iot data using dbn and clustering model in smart city*. *Future Generation Computer Systems*, 108:976 – 986, 2017. 23
- [227] Machado, Kássio, Azzedine Boukerche, Eduardo Cerqueira e Antonio Loureiro: *A Data-Centric Approach for Social and Spatiotemporal Sensing in Smart Cities*. *IEEE Internet Computing*, 23(1):9–18, Jan 2019. 23
- [228] Bermudez-Edo, Maria: *Analysing Real World Data Streams with Spatio-Temporal Correlations: Entropy vs. Pearson Correlation*. *Automation in Construction*, 88:87–100, 2018. 23
- [229] Kumar, Neeraj, Sudip Misra, Joel J.P.C. Rodrigues e Mohammad S. Obaidat: *Coalition Games for Spatio-Temporal Big Data in Internet of Vehicles Environment: A Comparative Analysis*. *IEEE Internet of Things Journal*, 2(4):310–320, Aug 2015. 23
- [230] Jiang, Hao, Lixia Li, Haoran Xian, Yulin Hu, Hehe Huang e Juzhen Wang: *Crowd Flow Prediction for Social Internet-of-Things Systems Based on the Mobile Network Big Data*. *IEEE Transactions on Computational Social Systems*, páginas 1–12, 2021. 23, 24
- [231] Moon, Aekyeung, Jaeyoung Kim, Jialing Zhang e Seung Woo Son: *Evaluating Fidelity of Lossy Compression on Spatiotemporal Data from an IoT Enabled Smart Farm*. *Computers and Electronics in Agriculture*, 154:304–313, 2018. 23, 27
- [232] Moon, Aekyeung, Jaeyoung Kim, Jialing Zhang e Seung Woo Son: *Lossy Compression on IoT Big Data by Exploiting Spatiotemporal Correlation*. Em *2017 IEEE High Performance Extreme Computing Conference (HPEC)*, páginas 1–7. IEEE, Sep 2017. 23, 27
- [233] Shuja, Junaid, Mohammad Ali Humayun, Waleed Alasmay, Hassan Sinky, Eisa Alanazi e Muhammad Khurram Khan: *Resource Efficient Geo-Textual Hierarchical Clustering Framework for Social IoT Applications*. *IEEE Sensors Journal*, páginas 1–1, 2021. 23, 24
- [234] Moon, Aekyeung, Jaeyoung Kim, Jialing Zhang, Hang Liu e Seung Woo Son: *Understanding the Impact of Lossy Compressions on IoT Smart Farm Analytics*. Em *2017 IEEE International Conference on Big Data (Big Data)*, IEEE International Conference on Big Data, páginas 4602–4611, 2017. 23
- [235] Loo, Becky P Y, Feiyang Zhang, Janet Hsiao, Antoni B Chan e Hui Lan: *Applying the Hidden Markov Model to Analyze Urban Mobility Patterns: An Interdisciplinary Approach*. *Chinese Geographical Science*, 31(1), 2021. 23

- [236] Chen, Mingdong: *Grassland Data Analysis and Calculation Based on the Internet of Things and Cloud Computing*. Em *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, páginas 1207–1210, Feb 2022. 23
- [237] Zhao, Wei, Meini Wang e V. T. Pham: *Unmanned aerial vehicle and geospatial analysis in smart irrigation and crop monitoring on iot platform*. *Mobile Information Systems*, 2023, 2023. 23
- [238] Jeong, Myeong Hun, Clair J. Sullivan e Shaowen Wang: *Complex Radiation Sensor Network Analysis with Big Data Analytics*. Em *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, páginas 1–4, 2015. 23
- [239] Kan, Zihan, Luliang Tang, Mei Po Kwan, Chang Ren, Dong Liu e Qingquan Li: *Traffic congestion analysis at the turn level using Taxis' GPS trajectory data*. *Computers, Environment and Urban Systems*, 74:229–243, 2019, ISSN 0198-9715. 23
- [240] Jia, Jia, Linghui Li, Ximing Li, Ning Wang, Binsi Cai, Xu Zhang e Pengfei Qiu: *Auto-gan: Gan-based self-supervised collaborative learning for robust spatio-temporal trajectory classification in iot*. *IEEE Internet of Things Journal*, 12(15):29642–29655, Aug 2025, ISSN 2327-4662. 23
- [241] Hu, Yang, Shaobo Li, Dawen Xia, Wenyong Zhang, Panliang Yuan, Fengbin Wu e Huaqing Li: *A multiview spatial-temporal adaptive transformer-gru framework for traffic flow prediction*. *IEEE Internet of Things Journal*, 12(6):7114–7132, March 2025, ISSN 2327-4662. 23
- [242] Yadav, Sonal, Yacine Belal, Brent Lagesse e Afra Mashhadi: *Benchmarking clustered federated learning algorithms for next-point prediction*. Em *2025 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, páginas 18–26, June 2025. 23
- [243] Zhang, Ke, Hongjin Ren, Jinbiao Kang, Cai Guo, Weiming Chen, Ming Tao, Hong Ning Dai, Shaohua Wan e Haiyong Bao: *TST-Trans: A Transformer Network for Urban Traffic Flow Prediction*. *IEEE Internet of Things Journal*, 12(7):8276–8287, April 2025, ISSN 2327-4662. 23
- [244] Yang, Shiyu, Qunying Wu, Mengmeng Li e Yu Sun: *Temporal identity interaction dynamic graph convolutional network for traffic forecasting*. *IEEE Internet of Things Journal*, 12(11):15057–15072, June 2025, ISSN 2327-4662. 23
- [245] Al Jawarneh, Isam Mashhour, Paolo Bellavista, Antonio Corradi, Luca Foschini e Rebecca Montanari: *Locality-preserving spatial partitioning for geo big data analytics in main memory frameworks*. Em *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, páginas 1–6, 2020. 23, 24, 28
- [246] Wang, Hailun, Mingming Zhang, Renyu Yang, Xuelian Lin, Tianyu Wo, Rajiv Ranjan e Jie Xu: *SMTP: An Optimized Storage Method for Vehicle Trajectory Data Exploiting Trajectory Patterns*. Em *2016 IEEE 18th International Conference on*

- High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, páginas 773–780, 2016. 23
- [247] Zhou, Lianjie, Wei Tu e Qingquan Li: *SMART-IMPALA: Efficient Querying of hyper Massive Spatiotemporal Trajectory Data*. Em *2021 28th International Conference on Geoinformatics*, páginas 1–5, Nov 2021. 22, 23
- [248] Godhbani, Salah, Sabeur, Wonho Suh e Seongkwan M. Lee: *ADL Based Framework For Multimodal Data Fusion in Traffic Jam Prediction*. Em *2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, páginas 126–132, Dec 2022. 22, 23
- [249] Wang, Shaowen: *Cybergis and spatial data science*. *GeoJournal*, 81:965–968, 2016. 23
- [250] Barik, Rabindra K., Sudhansu Shekhar Patra, Rasmita Patro, Sachi Nandan Mohanty e Abdulsattar Abdullah Hamad: *Geobd²: Geospatial big data deduplication scheme in fog assisted cloud computing environment*. Em *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, páginas 35–41, March 2021. 23
- [251] Fawzy, Dina, Sherin Moussa e Nagwa Badr: *The Spatiotemporal Data Fusion (STDF) Approach: IoT-Based Data Fusion using Big Data Analytics*. *Sensors*, 21(21), 2021. 23
- [252] Zhou, Yuchao, Suparna De, Wei Wang, Klaus Moessner e Marimuthu S Palaniswami: *Spatial Indexing for Data Searching in Mobile Sensing Environments*. *Sensors (Switzerland)*, 17(6), junho 2017. 23, 25
- [253] Huang, Chih Yuan e Yu Jui Chang: *An Adaptively Multi-Attribute Index Framework for Big IoT Data*. *Computers and Geosciences*, 155, 2021. 23
- [254] Jamjoum, Ghaith, Abdalrahman Eltahir, Mohannad Janahi, Sebti Foufou e Isam Mashhour Al Jawarneh: *A novel method for efficient approximate join queries on big hyperlocal geo data*. Em *2025 International Conference on Multimedia Computing, Networking and Applications (MCNA)*, páginas 25–30, 2025. 23, 25
- [255] Hossam, Aya, Ahmed Magdy, Ahmed Fawzy e Shriene M. Abd El-Kader: *An Integrated IoT System to Control the Spread of COVID-19 in Egypt*. Em *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, páginas 336–346. Springer International Publishing, 2020. 23, 25, 26, 28
- [256] Yu, Renxin, Qinghuang Yao, Zhong Tianrong, Wei Li e Ying Ma: *Visualized Panoramic Display Platform for Transmission Cable Based on Space-Time Big Data*. *Communications in Computer and Information Science*, 1123 CCIS:314–323, 2019. 23, 25, 26

- [257] Wang, Shaohua, Ershun Zhong, Wenwen Cai, Qiang Zhou, Hao Lu, Yongquan Gu, Weiying Yun, Zhongnan Hu e Liang Long: *A visual analytics framework for big spatiotemporal data*. Em *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Analytics for Local Events and News*, LENS'18, New York, NY, USA, 2018. Association for Computing Machinery, ISBN 9781450360357. 23, 28
- [258] Yu, Jia, Zongsi Zhang e Mohamed Sarwat: *Geosparkviz: A scalable geospatial data visualization framework in the apache spark ecosystem*. Em *Proceedings of the 30th International Conference on Scientific and Statistical Database Management, SSDBM '18*, New York, NY, USA, 2018. Association for Computing Machinery, ISBN 9781450365055. 23
- [259] Krommyda, Maria e Verena Kantere: *Spatial data management in iot systems: Solutions and evaluation*. *International Journal of Semantic Computing*, 15(1):117–139, 2021. 26
- [260] Junior, Braulio L. D. C., Edward David Moreno, Douglas D. J. Macedo, Diego Kreutz e Mario A. R. Dantas: *Towards a Hybrid Storage Architecture for IoT*. Em *2018 IEEE Symposium on Computers and Communications (ISCC)*, páginas 00470–00473, 2018. 26
- [261] Shin, Hansub, Kisung Lee e Hyuk Yoon Kwon: *A comparative experimental study of distributed storage engines for big spatial data processing using geospark*. *The Journal of Supercomputing*, 78(2):2556–2579, 2022, ISSN 1573-0484. 32, 43
- [262] Pandey, Varun, Andreas Kipf, Thomas Neumann e Alfons Kemper: *How good are modern spatial analytics systems?* *Proc. VLDB Endow.*, 11(11):1661–1673, julho 2018, ISSN 2150-8097. 35
- [263] Jordahl, Kelsey, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Adrian Garcia Badaracco, Carson Farmer, Geir Arne Hjelle, Alan D. Snow, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Nick Eubank, maxalbert, Aleksey Bilogur, Sergio Rey, Christopher Ren, Dani Arribas-Bel, Leah Wasser, Levi John Wolf, Martin Journois, Joshua Wilson, Adam Greenhall, Chris Holdgraf, Filipe e François Leblanc: *"geopandas/geopandas: v0.8.1"*, julho 2020. <https://doi.org/10.5281/zenodo.3946761>. 37
- [264] Portal, Chicago Data: *Postgres-xl*. <https://git.postgresql.org/gitweb/?p=postgres-xl.git>. Acessado em 10 de março de 2025. 41
- [265] The PostgreSQL Global Development Group: *PostgreSQL: Documentation: Multi-Version Concurrency Control*, 2001. <https://www.postgresql.org/docs/7.1/mvcc.html>, Acessado em 14 de setembro de 2025. 41
- [266] OSGeo, PostGIS PSC : *Postgis*. <https://postgis.net/>. Acessado em 10 de março de 2025. 42
- [267] Apache Sedona e Yitao Li: *apache.sedona: R Interface for Apache Sedona*, 2023. <https://sedona.apache.org/>. 43

- [268] Consortium, Open Geospatial: *Ogc sensorthings api part 1: Sensing version 1.1*, 2021. <https://docs.ogc.org/is/18-088/18-088.html>. 46
- [269] Portal, Chicago Data: *Chicago - bike routes*. <https://data.cityofchicago.org/Transportation/Bike-Routes/3w5d-sru8>. Acessado em 21 de janeiro de 2025. 55
- [270] Ahmed, Mahmuda, Sophia Karagiorgou, Dieter Pfoser e Carola Wenk: *A comparison and evaluation of map construction algorithms using vehicle tracking data*. *GeoInformatica*, 19(3):601–632, Jul 2015, ISSN 1573-7624. 56
- [271] Chicago, University of: *Array of things*. <https://arrayofthings.github.io/>. Acessado em 21 de janeiro de 2025. 56
- [272] Burrough, P.A., R.A. McDonnell e C.D. Lloyd: *Principles of Geographical Information Systems*. OUP Oxford, 2015, ISBN 9780198742845. 56
- [273] *Postgresql documentation: Gist indexes*. <https://www.postgresql.org/docs/current/gist.html>, 2025. Acesso em: 30 ago. 2025. 64
- [274] Silva, Dennis Sávio e Maristela Holanda: *Uma arquitetura para big data geoespacial de dispositivos iot com níveis heterogêneos de mobilidade*. Em *Anais do XL Simpósio Brasileiro de Bancos de Dados*, páginas 42–55, Porto Alegre, RS, Brasil, 2025. SBC. <https://sol.sbc.org.br/index.php/sbbd/article/view/37227>. 81