



NOVA IMS
INFORMATION MANAGEMENT SCHOOL

Modelagem de dados de Produtividade Agrícola em Goiás

Trabalho para obtenção de créditos da cadeira de Bases de Dados
Geoespaciais sob orientação do Prof. Dr. André Oliveira

Antonio Henrique Xavier da SILVA
20240915

30 de novembro de 2024

Sumário

1	Introdução	1
2	Área de Estudo	2
3	Modelo Relacional	3
3.1	Perspectiva administrativa	4
3.2	Perspectiva da infraestrutura	4
3.3	Perspectiva produtiva	5
3.4	Perspectiva natural	5
4	Implementação	6
5	Resultados	7
6	Conclusão	9
A	Diagrama de Entidades e Relacionamentos	11
B	Configuração e Inicialização da Base de Dados	12

Capítulo 1

Introdução

O Cerrado é um bioma que se estende pela maior parte do Brasil Central, caracterizando-se como a savana com maior biodiversidade do mundo. Apesar de suas condições naturais inicialmente desfavoráveis à agricultura, os esforços de modernização e adaptação tecnológica transformaram essa região em uma importante produtora de commodities agrícolas, especialmente soja e cana-de-açúcar [1].

A produção de alimentos é uma atividade econômica essencial para o desenvolvimento regional, sendo fundamental assegurar que essa produtividade ocorra de maneira sustentável e inclusiva, envolvendo produtores rurais de diferentes portes. O aumento da oferta agrícola, por meio de práticas eficientes e diversificadas, contribui para a estabilização de preços e o acesso a alimentos para a população local. A promoção da diversidade na produção também é necessária para evitar a formação de regiões de monocultura, que podem reduzir a oferta de alimentos básicos e impactar a segurança alimentar.

Além disso, é indispensável considerar os impactos ambientais e sociais decorrentes da exploração agropecuária. A expansão desordenada pode acarretar graves consequências, como perda de biodiversidade, alterações nos regimes de chuvas e contaminação de corpos d'água. Assim, o monitoramento constante e a análise dos impactos das atividades agrícolas são fundamentais para a formulação de políticas públicas que equilibrem o aumento da produtividade e a preservação ambiental.

Este trabalho propõe uma estrutura de dados voltada para a representação das unidades de produção agrícola e da infraestrutura associada, como redes viárias, silos de armazenamento de grãos e áreas irrigadas. Além disso, contempla a identificação e monitoramento de diferentes commodities agropecuárias, permitindo a obtenção de dados de produtividade média por localidade e a representação espacial por meio de cartas temáticas.

Como parte essencial da modelagem espacial, o processo de Extração, Transformação e Carga (ETL) é abordado neste estudo, evidenciando como a aplicação de um Sistema Gerenciador de Bases de Dados Espaciais (SGBDE) pode otimizar a indexação e organização de dados com base em critérios espaciais e semânticos. Essa abordagem acelera o tempo de consulta e permite análises geoespaciais detalhadas e eficazes, apoiando a tomada de decisões informadas sobre a produtividade e a sustentabilidade no setor agrícola.

Capítulo 2

Área de Estudo

Este estudo foca no estado de Goiás, localizado no Brasil Central e amplamente coberto pelo bioma Cerrado, uma região de grande relevância para a produção de alimentos no país. Goiás é um dos principais fornecedores agrícolas do mercado nacional e internacional, destacando-se pelo segundo maior rebanho bovino e pela quarta maior produção de grãos entre os estados brasileiros. O setor agropecuário é responsável por 11,4% do Produto Interno Bruto (PIB) do estado, refletindo sua significativa contribuição para a economia nacional [2].

A escolha de Goiás como área de estudo justifica-se por sua importância histórica e econômica no cenário agropecuário brasileiro. Como parte de uma fronteira agrícola que vem se expandindo rumo aos biomas Pantanal e Amazônico, Goiás representa uma região com alta concentração fundiária e forte dependência de monoculturas de baixo valor agregado. Essa dependência aumenta os desafios ambientais, uma vez que práticas agrícolas intensivas nessa região podem causar degradação do solo, desmatamento e desequilíbrios ecológicos.

Compreender a dinâmica da produção agrícola e suas infraestruturas auxiliares em Goiás é essencial para avaliar o impacto do setor sobre o Cerrado, um bioma com alta biodiversidade e grande fragilidade ambiental. Por meio da análise das características fundiárias, da infraestrutura de apoio e da produtividade agrícola local, este estudo busca oferecer subsídios para práticas agrícolas mais sustentáveis e eficientes, auxiliando na criação de políticas públicas que promovam o uso responsável dos recursos naturais e o desenvolvimento socioeconômico equilibrado.

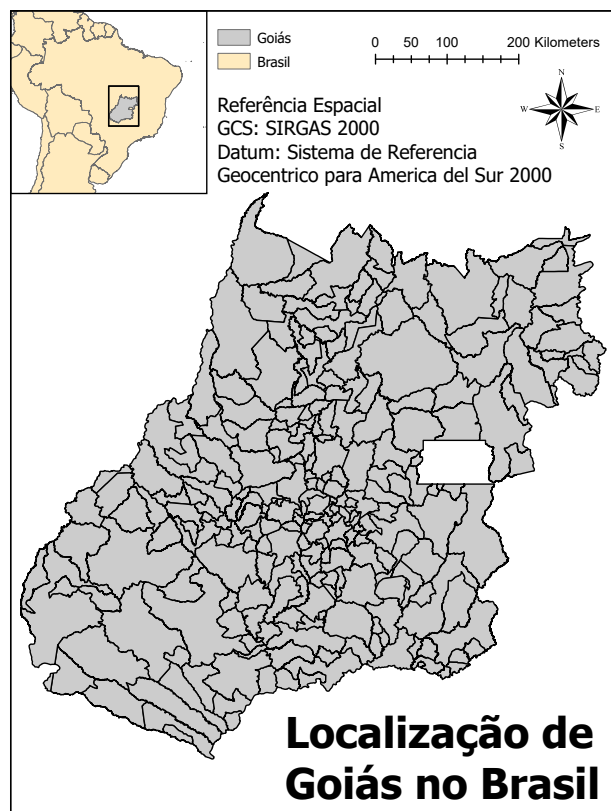


Figura 2.1. Localização de Goiás no Brasil

Capítulo 3

Modelo Relacional

Para o desenvolvimento do modelo de dados deste estudo, foram considerados dados abertos sobre a produção agrícola do estado de Goiás, disponibilizados por diversas instituições públicas, como o Instituto Brasileiro de Geografia e Estatística (IBGE), o Instituto Nacional de Colonização e Reforma Agrária (INCRA) e o Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). A modelagem que será apresentada foi inspirada no trabalho de Řezník et al. [3], apesar de não ser tão abrangente quanto.

A Figura 3.1 apresenta o modelo conceitual adotado, que foi estruturado em quatro perspectivas principais: administrativa, infraestrutura, produtiva e natural. Cada uma dessas perspectivas é composta por diferentes entidades, descritas detalhadamente nas seções subsequentes. As listagens 3.1, 3.2, 3.3 e 3.4 apresentam as entidades associadas a cada perspectiva, bem como seus respectivos atributos, o que facilita a compreensão das relações e interdependências entre os diferentes elementos do modelo.

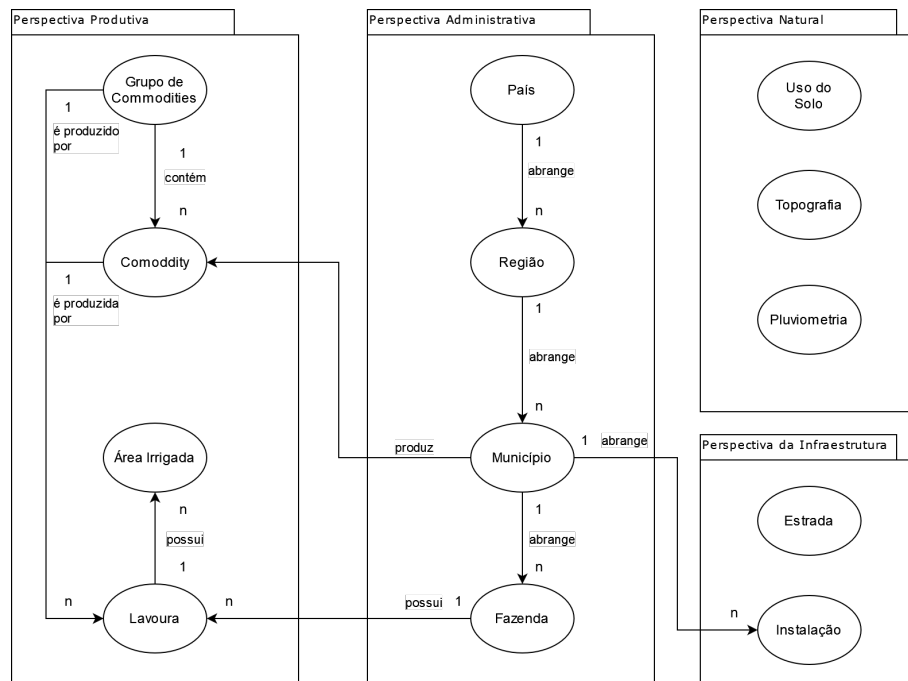


Figura 3.1. Modelo conceitual dos dados.

3.1 Perspectiva administrativa

```
Country(id, country_name)
Region(id, region_name, country_id)
Municipality(id, municipality_name, region_id, geom)
Farm(id, farm_name, farm_type, municipality_id, geom)
```

Listing 3.1. Entidades da perspectiva administrativa.

A perspectiva administrativa organiza as unidades agrícolas de forma hierárquica em quatro níveis: País, Região, Município e Fazenda. Para este trabalho, foram considerados apenas um país (Brasil) e uma região (Goiás), abrangendo um total de 246 municípios e mais de 90 mil fazendas cadastradas. Considera-se como "fazenda" qualquer propriedade rural registrada no sistema [SIGEF](#), mantido pelo INCRA para a gestão fundiária no território brasileiro. Os limites municipais foram obtidos a partir da [Malha Municipal](#) de 2022, disponibilizada pelo IBGE.

3.2 Perspectiva da infraestrutura

```
Facility(id, facility_name, facility_type, facility_capacity,
         municipality_id, geom)
Road(id, road_name, road_type, geom)
```

Listing 3.2. Entidades da perspectiva da infraestrutura.

Como discutido por Castillo [4], aspectos logísticos desempenham papel fundamental no desenvolvimento agrícola de Goiás, uma vez que as dinâmicas produtivas são fortemente impactadas pelo interesse de grandes empresas dos mais variados setores - como armazéns de estocagem, transportadoras e agroindústrias - que dependem da infraestrutura fornecida pelo estado para operar.

Neste sentido foram incluídas duas categorias de infraestrutura no modelo: Instalações de Estoque e Rodovias. A escolha dessas categorias fundamenta-se nas dificuldades logísticas da produção agropecuária em Goiás, onde é esperado que fazendas localizadas próximas a instalações de estocagem (como silos para grãos e frigoríficos para carnes) e a grandes centros consumidores (como Goiânia e Brasília) apresentem maior produtividade.

A perspectiva de infraestrutura, portanto, permite a análise de como a localização e o acesso a esses recursos influenciam a eficiência produtiva.

3.3 Perspectiva produtiva

```
Commodity_Group(ibge_code, group_name)
Commodity(id, commodity_name, group_id)
Commodity_Productivity(commodity_id, municipality_id, year,
    cultivated_area, production_value)
Irrigation(id, irrigation_description, farm_id, geom)
Plot(id, farm_id, geom)
```

Listing 3.3. Entidades da perspectiva produtiva.

Os dados produtivos foram obtidos a partir da [Pesquisa de Produção Agrícola Municipal \(PAM\)](#), que abrange informações sobre área plantada, área colhida e valor da produção de 71 produtos agrícolas para os 246 municípios goianos. Apesar de não serem georreferenciados, esses dados identificam cada município, permitindo análises de produtividade por área.

Também foram usados dados sobre pivôs centrais de irrigação, disponíveis pelo sistema [SIGA](#) que é mantido pela Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável de Goiás, considerando-se que cada área irrigada pertence à fazenda onde está localizada.

3.4 Perspectiva natural

```
Map_LandUsage(Tile_Id, filename, rast)
Map_Precipitation(Tile_Id, filename, rast)
Map_Temperature(Tile_Id, filename, rast)
Map_Elevation(Tile_Id, filename, rast)
```

Listing 3.4. Entidades da perspectiva natural.

Para representar características ambientais que influenciam a produção agrícola, foram utilizados dados em formato raster, que contêm informações sobre clima e aspectos geográficos da área de estudo. Esses dados auxiliam tanto na visualização exploratória quanto na modelagem espacial por meio de métodos como a Regressão Geograficamente Ponderada ou técnicas de interpolação, como o Kriging. Camadas de dados sobre uso do solo, elevação, precipitação e temperatura foram obtidas a partir da plataforma [MapBiomass](#) e o trabalho de Amaral et al. [5]. Essas camadas possibilitam a identificação de diferentes categorias de cobertura do solo e a análise das variáveis ambientais que podem influenciar a produtividade agrícola.

Capítulo 4

Implementação

O Diagrama de Entidades e Relacionamentos resultante do modelo lógico encontra-se disponível no Apêndice [A](#), onde são descritas as principais associações e dependências entre as diferentes entidades, bem como a estrutura hierárquica estabelecida para otimizar a organização e o acesso aos dados.

A implementação do modelo lógico foi realizada utilizando a base de dados [PostgreSQL](#) com a extensão [PostGIS](#). A escolha por essas tecnologias de código aberto reflete sua ampla adoção na comunidade científica e a facilidade de replicação dos resultados deste estudo, permitindo que diferentes atores se beneficiem das conclusões obtidas. Detalhes sobre a configuração e a inicialização do servidor encontram-se no Apêndice [B](#).

Para garantir uma representação precisa do modelo conceitual, foram seguidas as melhores práticas de modelagem relacional, considerando também as limitações das ferramentas utilizadas. Nas entidades com características espaciais, foram adicionados atributos específicos com tipos de dados complexos oferecidos pelo PostGIS, permitindo expressar a extensão territorial ou a localização de cada entidade.

A representação das entidades da perspectiva natural foi feita de forma a maximizar a eficiência e clareza das operações. Embora fosse possível consolidar todas as entidades ambientais em uma única tabela, optou-se por organizá-las em tabelas distintas, facilitando a ingestão e o particionamento dos dados. Essa abordagem modular permite uma administração mais eficiente dos dados matriciais e garante maior flexibilidade na realização de consultas e na execução de análises geoespaciais complexas.

Capítulo 5

Resultados

Após a inserção dos dados na base de dados, foram realizados experimentos para demonstrar a capacidade do sistema em analisar e visualizar informações geospaciais. Esses experimentos foram conduzidos utilizando um script Python e diversas bibliotecas de geoprocessamento e conexão com banco de dados, como [Geopandas](#), [Rasterio](#) e [Psycpg2](#).

Os experimentos incluíram a extração de dados do sistema e a criação de cartas temáticas a partir dessas informações. Por exemplo, um estudo explorou a relação entre as características das propriedades rurais e a produtividade da soja nos municípios próximos a Ceres, utilizando a razão entre o valor da produção (em mil reais) e a área plantada (em hectares) como métrica de produtividade. A análise visual das cartas temáticas geradas sugere uma correlação negativa entre a área média das propriedades e a produtividade de soja. Municípios como Ceres, com menor área média de propriedades, apresentaram maior produtividade, enquanto localidades como Nova Glória, com áreas maiores, registraram produtividade inferior.

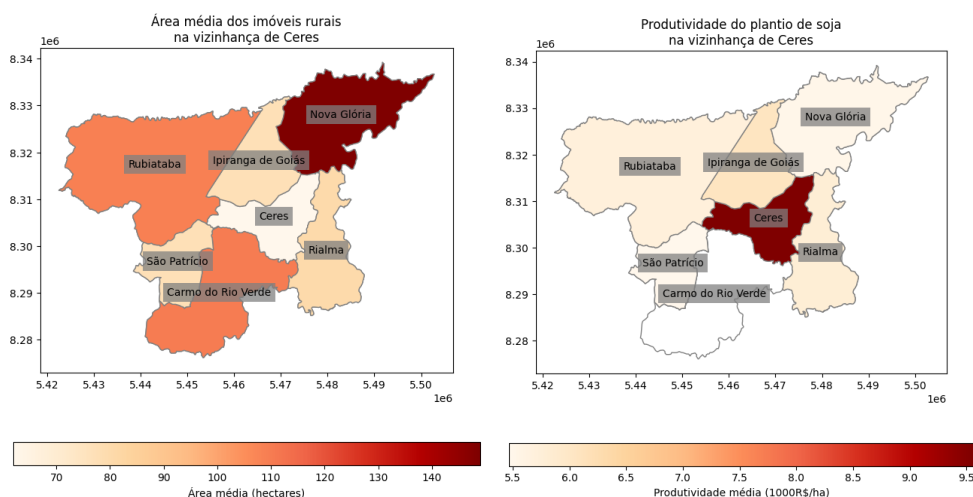


Figura 5.1. Área dos imóveis rurais e Produtividade da Soja na vizinhança de Ceres.

Além disso, foi realizada uma análise de dados matriciais utilizando as ferramentas da extensão [PostGIS Raster](#). Como exemplo, calculou-se a altitude média dos municípios na região de Ceres, permitindo uma compreensão espacial das variáveis ambientais e sua possível influência sobre a produtividade agrícola. A Figura 5.2 ilustra a distribuição da altitude média dos municípios próximos a Ceres, destacando variações topográficas que podem ser relevantes para planejamento agrícola e logístico.

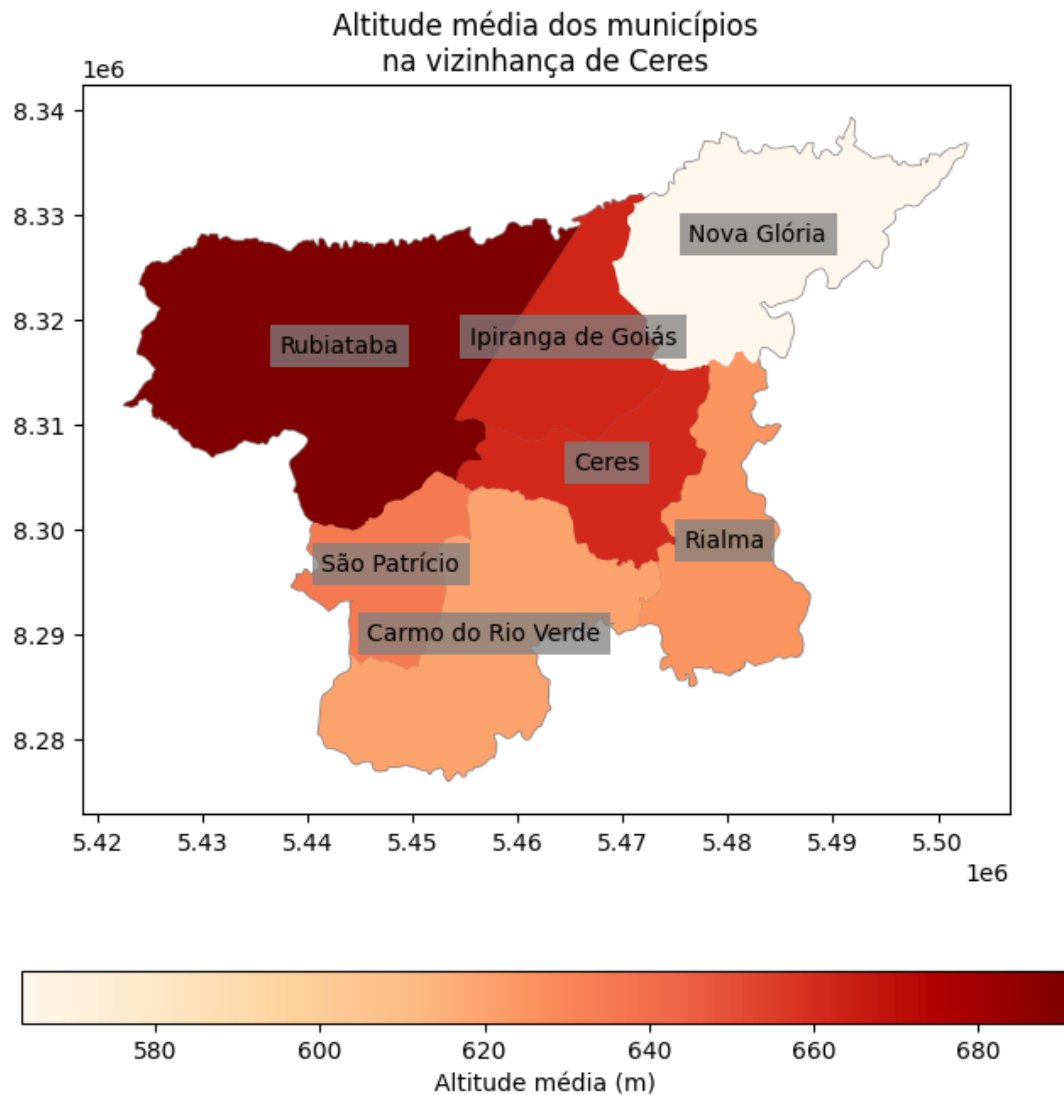


Figura 5.2. Altitude média dos municípios na vizinhança de Ceres

Esses resultados demonstram a eficácia do modelo proposto e sua aplicabilidade em análises complexas no setor agropecuário. A flexibilidade do sistema permite futuras extensões e personalizações, apoiando a tomada de decisões e a formulação de políticas para um desenvolvimento agrícola mais sustentável e eficiente.

Capítulo 6

Conclusão

Este estudo apresentou uma abordagem detalhada para o uso de Bancos de Dados Geoespaciais na gestão de informações relacionadas à produção agrícola, infraestrutura e características ambientais do estado de Goiás. A criação de uma estrutura de dados robusta e a utilização de um Sistema Gerenciador de Banco de Dados Espaciais (SGBDE) permitiram a organização e análise eficiente de dados complexos, essenciais para o monitoramento e o planejamento agrícola.

Os resultados demonstraram que a integração de diferentes perspectivas, incluindo dados administrativos, de infraestrutura, produtivos e naturais, proporciona uma visão holística da produção agrícola no estado. A análise revelou correlações importantes, como a relação inversa entre a área média das propriedades e a produtividade agrícola, destacando o potencial dos bancos de dados geoespaciais para a formulação de políticas públicas e estratégias de desenvolvimento sustentável.

Além disso, o uso de tecnologias como o PostgreSQL e a extensão PostGIS mostrou-se adequado para a manipulação de dados espaciais de alta complexidade, garantindo precisão nas consultas e flexibilidade para futuras expansões do modelo. Este estudo, portanto, contribui para a valorização e adoção de sistemas de banco de dados geoespaciais na agricultura, propondo um modelo replicável para outras regiões e culturas.

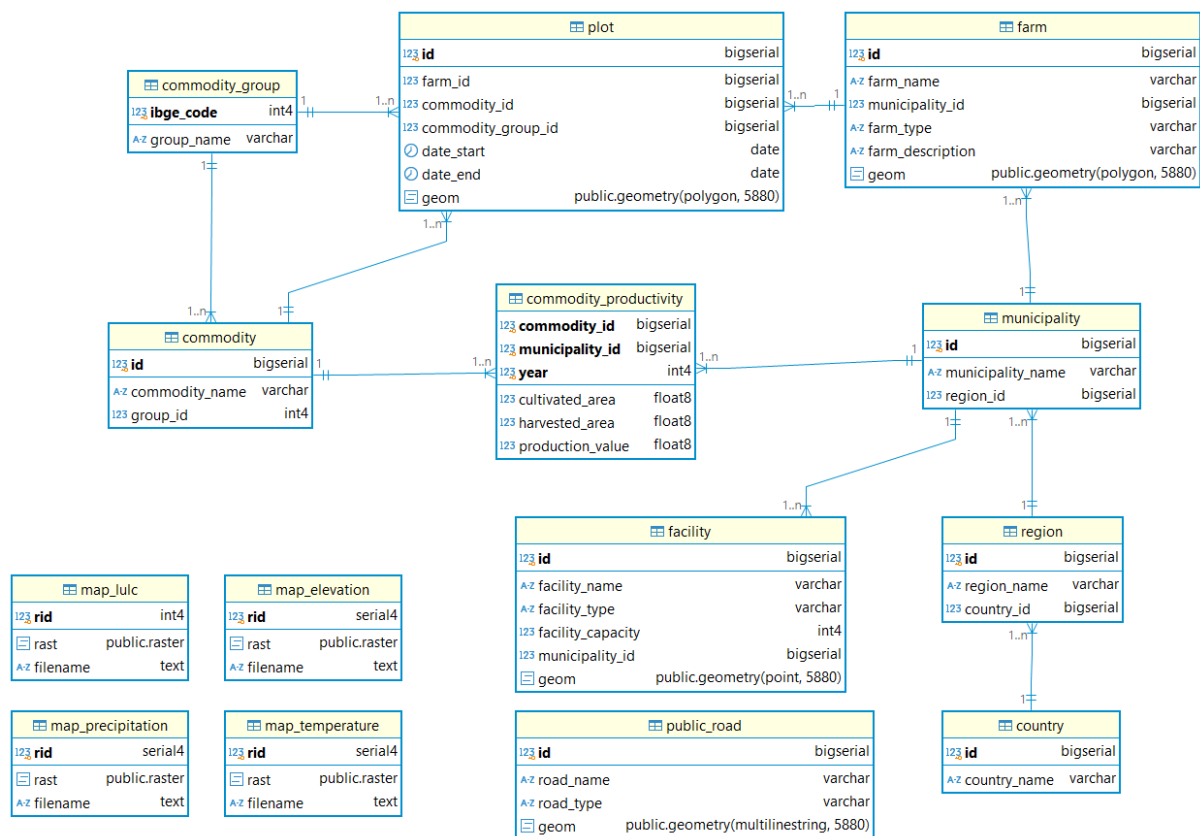
A adoção de bases de dados geoespaciais oferece um diferencial estratégico ao permitir a indexação dos dados conforme critérios espaciais específicos, o que facilita a recuperação de informações em análises complexas. Esse tipo de estrutura não apenas aprimora a eficiência das consultas, mas também viabiliza o uso de paralelismo de dados, permitindo que diversas análises sejam processadas simultaneamente. Com isso, é possível realizar operações em larga escala e obter resultados em menor tempo, fator crucial para decisões ágeis e embasadas em contextos como a gestão agrícola, onde o tempo de resposta pode impactar diretamente a produtividade e a sustentabilidade das operações.

Bibliografia

- [1] Diego Arias e Barbara Farinelli. “Agriculture Productivity Growth in Brazil”. en. Em: (set. de 2017) (ver p. 1).
- [2] “Sobre Goiás”. pt-BR. Section: Goiás - Visão Geral. Jul. de 2018. URL: <https://goias.gov.br/imb/sobre-goias/> (acesso em 10/11/2024) (ver p. 2).
- [3] Tomáš Řezník et al. “Open Data Model for (Precision) Agriculture Applications and Agricultural Pollution Monitoring”. Em: jun. de 2015. DOI: [10.2991/ict4s-env-15.2015.12](https://doi.org/10.2991/ict4s-env-15.2015.12) (ver p. 3).
- [4] Ricardo Castillo. “AGRONEGÓCIO E LOGÍSTICA EM ÁREAS DE CERRADO: EXPRESSÃO DA AGRICULTURA CIENTÍFICA GLOBALIZADA”. en. Em: *Revista da ANPEGE* 3.03 (jul. de 2017). Publisher: Universidade Federal da Grande Dourados, pp. 21–27. ISSN: 1679-768X. DOI: [10.5418/RA2007.0303.0003](https://doi.org/10.5418/RA2007.0303.0003). URL: <https://ojs.ufgd.edu.br/index.php/anpege/article/view/6604> (acesso em 10/11/2024) (ver p. 4).
- [5] Silvana Amaral et al. “AMBDATA: Variáveis ambientais para Modelos de Distribuição de Espécies (SDMs)”. pt. Em: Instituto Nacional de Pesquisas Espaciais (INPE), 2013, pp. 6930–6937. ISBN: 978-85-17-00066-9 978-85-17-00065-2. URL: <http://marte2.sid.inpe.br/col/dpi.inpe.br/marte2/2013/05.29.00.11.46/doc/thisInformationItemHomePage.html> (acesso em 10/11/2024) (ver p. 5).

Apêndice A

Diagrama de Entidades e Relacionamentos



Apêndice B

Configuração e Inicialização da Base de Dados

A base de dados apresentada neste trabalho foi construídas utilizando a imagem Docker [postgis/postgis](#), que permite virtualizar uma máquina virtual no qual um serviço PostgreSQL estará sendo executado, já com a extensão PostGIS habilitada.

Um arquivo de configuração Docker Compose foi utilizado para definir um serviço replicável e consistente. Foi criado também um programa Shell script que utiliza as aplicações CLI [gdalwarp](#) e [raster2pgsql](#) para projetar os dados no sistema de referência [SIRGAS 2000](#) e realizar a ingestão dos dados matriciais na base dados.

Todo o código necessário para replicar a base de dados e os experimentos citados no Capítulo 5 estão disponíveis publicamente [neste repositório do Github](#).