

Primeiro Trabalho Prático

As Cinzas do Brasil

Análise dos dados brasileiros de queimadas

Clara Ernesto de Carvalho	14559479
Felipe Carneiro Machado	14569373
Lívia Lelis	12543822

Sumário

1	Descrição do Problema	4
1.1	Fontes dos Dados	4
1.1.1	Queimadas - INPE	4
1.1.2	Clima - SISAM/INPE	4
1.1.3	Unidades Federativas - IBGE	5
1.1.4	Municípios - IBGE	5
1.2	Assuntos de Interesse	6
1.2.1	FRP - Fire Radiative Power	6
1.2.2	Dias sem Chuva	6
1.2.3	Umidade Relativa	7
1.2.4	Risco de Fogo	7
1.3	Dimensões e Fatos	7
1.3.1	Dimensões Conformadas	7
1.3.2	Tabelas de Fato	7
1.3.2.1	Fato Queimadas:	7
1.3.2.2	Fato Clima:	8
1.3.3	Processo de Integração	8
1.3.4	Ausência de Conflitos Semânticos	8
1.4	Mapeamento de Atributos: Fonte para Data Warehouse	8
1.4.1	Dimensão: dim_data	9
1.4.2	Dimensão: dim_local	9
1.4.3	Dimensão: dim_ponto	9
1.4.4	Dimensão: dim_horario	9
1.4.5	Tabela Ponte: bridge_horario	10
1.4.6	Fato: Queimadas	10
1.4.7	Fato: Clima	11
2	Esquemas	11
2.1	Dimensões Conformadas	11
2.1.1	dim_data	11
2.1.1.1	Atributos Analíticos Adicionais	12
2.1.2	dim_local	12
2.1.3	dim_ponto	13
2.2	Slowly Changing Dimensions (SCDs)	13
2.2.1	Justificativa para Não Implementação de SCDs	13
2.2.2	Abordagem Adotada	14
2.2.3	dim_horario	14
2.3	Tabela Ponte (Bridge Table) para Tempo	14
2.3.1	bridge_horario	14
2.3.2	Mecânica da Tabela Ponte	14
2.3.3	Alternativas Arquiteturais Consideradas	15
2.3.3.1	Abordagem 1: Fato Único com Granularidade Mínima	15

2.3.3.2	Abordagem 2: Fato Único com Granularidade Máxima	15
2.3.3.3	Abordagem 3: Dimensões Temporais Separadas	15
2.3.3.4	Justificativa da Solução Adotada	15
2.4	Fato Queimadas	15
2.4.1	Granularidade	15
2.4.2	Fatos	15
2.4.3	Dimensões	16
2.5	Fato Clima	17
2.5.1	Granularidade	17
2.5.2	Fatos	18
2.5.3	Dimensões	19
2.6	Constelação de Fatos	20
3	Consultas	21
3.1	Drill-Down com agregação de contagem	22
3.1.1	Motivação	22
3.2	Roll-Up com agregação de média	22
3.2.1	Motivação	22
3.3	Pivot	22
3.3.1	Motivação	22
3.4	Slice com agregação de média espacial	22
3.4.1	Motivação	22
3.5	Drill-Down com análise hierárquica por bioma	23
3.5.1	Motivação	23
3.6	Slice multi-dimensional com qualidade do ar	23
3.6.1	Motivação	23
3.7	Drill-Across com agregação de média	23
3.7.1	Motivação	23
3.8	Análise de Pico de Queimadas por Hora usando Bridge Table	24
3.8.1	Motivação	24
	REFERÊNCIAS	25

1 Descrição do Problema

O Brasil sofre com o volumoso problema que são as queimadas, segundo dados do Monitor do Fogo do projeto MapBiomas, mais de 30,8 milhões de hectares foram queimados no Brasil entre janeiro e dezembro de 2024, uma área maior que todo o território da Itália. (1)

Nesse contexto, o nosso foco é analisar os dados de queimada e clima informados pelo INPE, com o objetivo de encontrar correlações entre os dados, entender como eles se comportam em relação à localização geográfica, época do ano, e como se alteraram ao longo dos anos; e criar um data warehouse para facilitar essas consultas uma vez que possuímos um grande volume de dados.

Nós usaremos dois datasets, o dataset do INPE de queimadas (2) e o de clima (3). Ambos possuem dados com localização espacial e temporal, o de clima possui dados desde 2000, e o de queimadas desde 2003. Devido à longa coleta de dados, ambos datasets são muito volumosos, sendo que o dataset de queimadas possui 17.542.900 tuplas, e o dataset de clima possui 158.705.816, totalizando em mais de 15GB. Além disso, precisaremos dos diretórios brasileiros de unidades federativas e municípios.

1.1 Fontes dos Dados

1.1.1 Queimadas - INPE

A Tabela 1 apresenta o esquema do dataset de queimadas fornecido pelo INPE, contendo informações sobre focos de incêndio detectados por satélite.

Tabela 1 – Esquema do Dataset de Queimadas - INPE

Atributo	Tipo	Modo	Descrição
ano	INTEGER	NULLABLE	Ano de referência da passagem do satélite segundo o fuso horário de Greenwich (GMT).
mes	INTEGER	NULLABLE	Mês de referência da passagem do satélite segundo o fuso horário de Greenwich (GMT).
data_hora	DATETIME	NULLABLE	Data e hora da passagem do satélite no fuso horário de Greenwich (GMT).
bioma	STRING	NULLABLE	Nome do Bioma.
sigla_uf	STRING	NULLABLE	Sigla da Unidade Federativa.
id_municipio	STRING	NULLABLE	Identificação do Município.
latitude	FLOAT	NULLABLE	Latitude do centro do píxel de fogo ativo em graus decimais.
longitude	FLOAT	NULLABLE	Longitude do centro do píxel de fogo ativo em graus decimais.
satelite	STRING	NULLABLE	Nome do algoritmo utilizado e referência ao satélite provedor da imagem.
dias_sem_chuva	FLOAT	NULLABLE	Número de dias sem chuva até a detecção do foco.
precipitacao	FLOAT	NULLABLE	Precipitação acumulada no dia até o momento da detecção do fogo.
risco_fogo	FLOAT	NULLABLE	Valor do Risco de Fogo previsto para o dia da detecção do fogo.
potencia_radiativa_fogo	FLOAT	NULLABLE	Fire Radiative Power, MW (megawatts).

1.1.2 Clima - SISAM/INPE

A Tabela 2 apresenta o esquema do dataset de clima fornecido pelo Sistema de Informações Ambientais Integrado à Saúde Ambiental (SISAM) do INPE, contendo dados

meteorológicos e de qualidade do ar.

Tabela 2 – Esquema do Dataset de Clima - SISAM/INPE

Atributo	Tipo	Modo	Descrição
ano	INTEGER	NULLABLE	Ano do registro.
sigla_uf	STRING	NULLABLE	Sigla da Unidade da Federação.
id_municipio	STRING	NULLABLE	ID do Município segundo classificação IBGE (7 dígitos).
data_hora	DATETIME	NULLABLE	Data e hora do registro meteorológico.
co_ppb	FLOAT	NULLABLE	Concentração de monóxido de carbono (CO) em partes por bilhão (ppb).
no2_ppb	FLOAT	NULLABLE	Concentração de dióxido de nitrogênio (NO ₂) em partes por bilhão (ppb).
o3_ppb	FLOAT	NULLABLE	Concentração de ozônio (O ₃) em partes por bilhão (ppb).
pm25_ugm3	FLOAT	NULLABLE	Material particulado fino (PM2.5) em microgramas por metro cúbico ($\mu\text{g}/\text{m}^3$).
so2_ugm3	FLOAT	NULLABLE	Dióxido de enxofre (SO ₂) em microgramas por metro cúbico ($\mu\text{g}/\text{m}^3$).
precipitacao_dia	FLOAT	NULLABLE	Total de precipitação acumulada no momento do registro.
temperatura	FLOAT	NULLABLE	Temperatura do ar (°C).
umidade_relativa	FLOAT	NULLABLE	Umidade relativa do ar (%).
vento_direcao	INTEGER	NULLABLE	Direção do vento (graus azimutais).
vento_velocidade	FLOAT	NULLABLE	Velocidade do vento (m/s).

1.1.3 Unidades Federativas - IBGE

A Tabela 3 apresenta o esquema do dataset de Unidades Federativas fornecido pelo IBGE, contendo informações sobre os estados brasileiros e suas regiões.

Tabela 3 – Esquema do Dataset de Unidades Federativas - IBGE

Atributo	Tipo	Modo	Descrição
id_uf	INTEGER	NOT NULL	Código identificador da Unidade da Federação (2 dígitos).
sigla	STRING	NOT NULL	Sigla da Unidade da Federação.
nome	STRING	NOT NULL	Nome completo da Unidade da Federação.
regiao	STRING	NOT NULL	Nome da região geográfica (Norte, Nordeste, Centro-Oeste, Sudeste, Sul).

1.1.4 Municípios - IBGE

A Tabela 4 apresenta o esquema do dataset de Municípios fornecido pelo IBGE, contendo informações detalhadas sobre todos os municípios brasileiros, incluindo divisões administrativas e geográficas.

Tabela 4 – Esquema do Dataset de Municípios - IBGE

Atributo	Tipo	Modo	Descrição
id_municipio	INTEGER	NOT NULL	Código identificador do município IBGE (7 dígitos).
id_municipio_6	INTEGER	NULLABLE	Código do município com 6 dígitos.
id_municipio_tse	INTEGER	NULLABLE	Código do município no Tribunal Superior Eleitoral.
id_municipio_rf	INTEGER	NULLABLE	Código do município na Receita Federal.
id_municipio_bcb	INTEGER	NULLABLE	Código do município no Banco Central do Brasil.
nome	STRING	NOT NULL	Nome do município.
capital_uf	INTEGER	NULLABLE	Indicador se o município é capital (1) ou não (0).
id_comarca	INTEGER	NULLABLE	Código identificador da comarca.
id_regiao_saude	INTEGER	NULLABLE	Código da região de saúde.
nome_regiao_saude	STRING	NULLABLE	Nome da região de saúde.
id_regiao_imediata	INTEGER	NULLABLE	Código da região geográfica imediata.
nome_regiao_imediata	STRING	NULLABLE	Nome da região geográfica imediata.
id_regiao_intermediaria	INTEGER	NULLABLE	Código da região geográfica intermediária.
nome_regiao_intermediaria	STRING	NULLABLE	Nome da região geográfica intermediária.
id_microrregiao	INTEGER	NULLABLE	Código da microrregião.
nome_microrregiao	STRING	NULLABLE	Nome da microrregião.
id_mesorregiao	INTEGER	NULLABLE	Código da mesorregião.
nome_mesorregiao	STRING	NULLABLE	Nome da mesorregião.
id_regiao_metropolitana	STRING	NULLABLE	Lista de IDs de regiões metropolitanas (separados por vírgula).
nome_regiao_metropolitana	STRING	NULLABLE	Lista de nomes de regiões metropolitanas (separados por vírgula).
ddd	INTEGER	NULLABLE	Código DDD de telefonia.
id_uf	INTEGER	NOT NULL	Código da Unidade da Federação.
sigla_uf	STRING	NOT NULL	Sigla da Unidade da Federação.
nome_uf	STRING	NOT NULL	Nome da Unidade da Federação.
nome_regiao	STRING	NOT NULL	Nome da região geográfica.
amazonia_legal	INTEGER	NULLABLE	Indicador se pertence à Amazônia Legal (1) ou não (0).
centroide	STRING	NULLABLE	Coordenadas geográficas do centroide do município.

1.2 Assuntos de Interesse

1.2.1 FRP - Fire Radiative Power

Impacto da Queimada - Indica (em MW) a potência radiativa da queimada, relacionada com o quanto de material foi consumido, e portanto, o quanto que a queimada impactou. Ela pode ser analisada em relação aos locais em que ocorre, e ao tempo. A métrica (fato) observada é o FRP e a granularidade desta métrica está por minuto e local. As dimensões relacionadas são data, horário e local.

1.2.2 Dias sem Chuva

Analisar o impacto dos períodos de seca nas queimadas, e em quais locais ocorrem mais queimadas mesmo fora destes períodos. Além disso, correlacionar a duração destes períodos com o impacto da queimada (FRP). A métrica (fato) observada é a dias sem chuva e a granularidade desta métrica está por minuto e local. As dimensões relacionadas são data e local.

1.2.3 Umidade Relativa

Analisar umidade em relação ao local, data e horário para encontrar locais e momentos mais afetados por baixa umidade e correlacionar com o impacto da queimada (FRP). A métrica (fato) observada é a `umidade_relativa` e a granularidade desta métrica está por hora e local. As dimensões relacionadas são data, horário e local.

1.2.4 Risco de Fogo

Analisar a previsão do risco de queimada em relação a local e data, e analisar se há correlação com o impacto das queimadas. (Se a previsão da ocorrência de queimadas se correlaciona com o impacto das queimadas existentes no mesmo período).

A métrica (fato) observada é a `risco_fogo` e a granularidade desta métrica está por minuto e local. As dimensões relacionadas são data e local.

1.3 Dimensões e Fatos

Para a modelagem do Data Warehouse, os dados serão organizados em uma constelação de fatos, composta por duas tabelas de fatos principais (Queimadas e Clima) e um conjunto de dimensões conformadas. As dimensões conformadas são compartilhadas entre as tabelas de fato, garantindo consistência e permitindo análises integradas (drill-across).

1.3.1 Dimensões Conformadas

As dimensões que darão contexto aos fatos são:

- **Dimensão de Tempo (Data e Horário):** Para analisar os eventos ao longo do tempo, com granularidades que vão de minuto a ano. Uma vez que os fatos de Queimadas e Clima possuem granularidades temporais distintas (minuto e hora, respectivamente), será utilizada uma tabela ponte para compatibilizá-las.
- **Dimensão de Localização:** Para análises geográficas, com hierarquias que vão desde coordenadas/municípios até regiões e biomas.

1.3.2 Tabelas de Fato

As tabelas de fato conterão as métricas quantitativas dos eventos. A correta classificação da aditividade de cada fato é crucial para a análise.

1.3.2.1 Fato Queimadas:

- **Potência Radiativa do Fogo (FRP):** É um fato aditivo. A energia liberada pode ser somada através de qualquer dimensão (e.g., total de FRP para um estado em um mês).
- **Dias sem Chuva:** É um fato semi-aditivo. Pode ser agregado por média entre locais, mas não é somável ao longo do tempo. Para análises temporais, utilizam-se funções como 'MAX', 'MIN' ou a última medição.

- Risco de Fogo: É um fato não aditivo. Não pode ser somado de forma útil. A agregação correta é a média para entender o risco em uma determinada região ou período.

1.3.2.2 Fato Clima:

- Temperatura, Umidade Relativa, Velocidade do Vento: São fatos não aditivos. A agregação deve ser feita por média.
- Precipitação: É um fato semi-aditivo. A precipitação acumulada pode ser somada entre diferentes locais para um mesmo período, mas não ao longo do tempo.

1.3.3 Processo de Integração

A estratégia de integração adotada consiste em:

1. Integração Direta por Chaves Naturais: Os campos `id_municipio` e `sigla_uf` servem como chaves naturais para unificar os dados geográficos entre as fontes.
2. Enriquecimento com Dados IBGE: Os diretórios de municípios e unidades federativas do IBGE são utilizados para enriquecer as dimensões conformadas com informações hierárquicas (região, mesorregião, microrregião).
3. Dimensões Conformadas: A criação de dimensões conformadas (`dim_data`, `dim_local`, `dim_horario`) garante consistência semântica entre os fatos de queimadas e clima.
4. Resolução de Granularidade Temporal: Uma tabela ponte (`bridge_horario`) resolve a diferença de granularidade entre minutos (queimadas) e horas (clima), permitindo análises drill-across.

1.3.4 Ausência de Conflitos Semânticos

Como todos os datasets originam-se de fontes oficiais brasileiras (INPE e IBGE) e seguem padrões governamentais de codificação, não há necessidade de reconciliação semântica ou resolução de conflitos. A integração é essencialmente uma operação de junção por chaves naturais, seguida de desnormalização controlada para o modelo dimensional.

1.4 Mapeamento de Atributos: Fonte para Data Warehouse

Esta seção documenta o mapeamento completo entre os atributos das fontes de dados e os elementos do Data Warehouse (dimensões e fatos).

1.4.1 Dimensão: dim_data

Tabela 5 – Mapeamento de Atributos - dim_data

Atributo DW	Fonte	Atributo Fonte	Transformação
id	-	-	Chave surrogate gerada
ano	Queimadas/Clima	ano	Extração direta
semestre	Queimadas/Clima	data_hora	Calculado: 1 (jan-jun) ou 2 (jul-dez)
trimestre	Queimadas/Clima	data_hora	Calculado: ceil(mes/3)
mês	Queimadas/Clima	mes	Extração direta
dia	Queimadas/Clima	data_hora	Extração do dia do campo datetime
dia_da_semana	Queimadas/Clima	data_hora	Calculado (1=domingo, 7=sábado)
dia_do_ano	Queimadas/Clima	data_hora	Calculado (1-366)
numero_semana	Queimadas/Clima	data_hora	Calculado (ISO week number)
fim_de_semana	Queimadas/Clima	data_hora	Calculado: dia_da_semana IN (1,7)
estacao	Queimadas/Clima	data_hora	Mapeamento por mês (hemisfério sul)

1.4.2 Dimensão: dim_local

Tabela 6 – Mapeamento de Atributos - dim_local

Atributo DW	Fonte	Atributo Fonte	Transformação
id	-	-	Chave surrogate gerada
municipio	Municípios IBGE	nome	Extração direta
estado	Municípios IBGE	nome_uf	Extração direta
região	Municípios IBGE	nome_regiao	Extração direta
Chave Natural de Integração: id_municipio (7 dígitos IBGE)			

1.4.3 Dimensão: dim_ponto

Tabela 7 – Mapeamento de Atributos - dim_ponto

Atributo DW	Fonte	Atributo Fonte	Transformação
id	-	-	Chave surrogate gerada
local_fk	dim_local	id	Lookup via id_municipio
latitude	Queimadas	latitude	Extração direta (graus decimais)
longitude	Queimadas	longitude	Extração direta (graus decimais)
bioma	Queimadas	bioma	Extração direta

1.4.4 Dimensão: dim_horario

Tabela 8 – Mapeamento de Atributos - dim_horario

Atributo DW	Fonte	Atributo Fonte	Transformação
id	-	-	Chave surrogate gerada
hora	Queimadas/Clima	data_hora	Extração da hora do campo datetime
minuto	Queimadas/Clima	data_hora	Extração do minuto do campo datetime

1.4.5 Tabela Ponte: bridge_horario

Tabela 9 – Mapeamento de Atributos - bridge_horario

Atributo DW	Fonte	Atributo Fonte	Transformação
horario_minuto_fk	dim_horario	id	Lookup para granularidade de minuto
horario_hora_fk	dim_horario	id	Lookup para granularidade de hora (minuto=0)
Relação: cada hora (HH:00) mapeia para 60 minutos (HH:00 a HH:59)			

1.4.6 Fato: Queimadas

Tabela 10 – Mapeamento de Atributos - Fato Queimadas

Atributo DW	Tipo	Fonte	Atributo Fonte	Transformação
Chaves Estrangeiras (Dimensões)				
data_fk	FK	Queimadas	data_hora, ano, mes	Lookup em dim_data
ponto_fk	FK	Queimadas	lat, long, bioma, id_municipio	Lookup em dim_ponto
horario_fk	FK	Queimadas	data_hora	Lookup em dim_horario (minuto)
Fatos (Métricas)				
risco_fogo	Fato	Queimadas	risco_fogo	Extração direta (não aditivo)
potencia_radiativa_fogo	Fato	Queimadas	potencia_radiativa_fogo	Extração direta (aditivo, MW)
dias_sem_chuva	Fato	Queimadas	dias_sem_chuva	Extração direta (semi-aditivo)
Granularidade: um foco de incêndio por ponto geográfico, data e minuto				

1.4.7 Fato: Clima

Tabela 11 – Mapeamento de Atributos - Fato Clima

Atributo DW	Tipo	Fonte	Atributo Fonte	Transformação
Chaves Estrangeiras (Dimensões)				
data_fk	FK	Clima	data_hora, ano	Lookup em dim_data
local_fk	FK	Clima	id_municipio, sigla_uf	Lookup em dim_local
horario_fk	FK	Clima	data_hora	Lookup em dim_horario (hora)
Fatos (Métricas)				
temperatura	Fato	Clima	temperatura	Extração direta (°C, não aditivo)
precipitacao	Fato	Clima	precipitacao_dia	Extração direta (mm, semi-aditivo)
umidade_relativa	Fato	Clima	umidade_relativa	Extração direta (% , não aditivo)
vento_velocidade	Fato	Clima	vento_velocidade	Extração direta (m/s, não aditivo)
vento_direcao	Fato	Clima	vento_direcao	Extração direta (graus, não aditivo)
co_ppb	Fato	Clima	co_ppb	Extração direta (ppb, não aditivo)
no2_ppb	Fato	Clima	no2_ppb	Extração direta (ppb, não aditivo)
o3_ppb	Fato	Clima	o3_ppb	Extração direta (ppb, não aditivo)
pm25_ugm3	Fato	Clima	pm25_ugm3	Extração direta ($\mu\text{g}/\text{m}^3$, não aditivo)
so2_ugm3	Fato	Clima	so2_ugm3	Extração direta ($\mu\text{g}/\text{m}^3$, não aditivo)
Granularidade: medições climáticas agregadas por município, data e hora				

2 Esquemas

Os fatos e dimensões foram organizados em uma constelação de fatos composta por 2 esquemas estrela: Queimadas, responsável por armazenar dados relativos à ocorrência de queimadas, e Clima, que armazena dados climáticos históricos. A integração entre os esquemas é realizada através de dimensões conformadas (conformed dimensions), que são compartilhadas entre os fatos, permitindo análises integradas e consistentes.

2.1 Dimensões Conformadas

Para integrar os esquemas Queimadas e Clima de forma consistente, foram criadas dimensões conformadas que são compartilhadas entre ambos os fatos. Esta abordagem permite análises cruzadas e drill-across entre os dois domínios de dados.

2.1.1 dim_data

Dimensão temporal compartilhada por ambos os fatos:

- id: long, chave primária (PK).
- ano: int.

- semestre: int.
- trimestre: int.
- mês: int.
- dia: int.
- dia_da_semana: int (1-7, onde 1=domingo).
- dia_do_ano: int (1-366).
- numero_semana: int (1-53).
- fim_de_semana: boolean.
- estacao: string (Verão, Outono, Inverno, Primavera).
- Hierarquia: dia \rightarrow mês \rightarrow trimestre \rightarrow semestre \rightarrow ano

2.1.1.1 Atributos Analíticos Adicionais

Os atributos `dia_da_semana`, `fim_de_semana`, `numero_semana`, `dia_do_ano` e `estacao` foram incluídos para facilitar análises específicas:

- Análises Sazonais: O atributo `estacao` permite estudos sobre padrões de queimadas e condições climáticas por estações do ano, fundamental para entender ciclos naturais de seca e chuvas.
- Análises Semanais: `dia_da_semana` e `fim_de_semana` facilitam a identificação de padrões relacionados a atividades humanas, como queimadas controladas durante dias úteis.
- Análises de Períodos Específicos: `numero_semana` e `dia_do_ano` permitem comparações diretas entre períodos equivalentes de anos diferentes, essenciais para análises de tendências temporais.

2.1.2 dim_local

Dimensão geográfica conformada, no grão de município.

- id: long, chave primária (PK).
- municipio: string.
- estado: string.
- região: string.
- Hierarquia: município \rightarrow estado \rightarrow região.

2.1.3 dim_ponto

Dimensão essencial para análise de queimadas por biomas e regiões geográficas específicas. Esta dimensão mantém o grão fino necessário para análises ecológicas e de conservação, permitindo estudos sobre padrões de queimadas em diferentes ecossistemas brasileiros.

- id: long, chave primária (PK).
- local_fk: long, FK para dim_local.
- latitude: float.
- longitude: float.
- bioma: string. Fundamental para análises por Amazônia, Cerrado, Mata Atlântica, etc.

Esta estrutura hierárquica (ponto \rightarrow município \rightarrow estado \rightarrow região) é fundamental para o domínio de queimadas, pois permite análises em múltiplas granularidades geográficas e por biomas. Enquanto o fato de clima opera no grão de município, o fato de queimadas requer granularidade mais fina para capturar a diversidade ecológica e espacial dos incêndios. Análises drill-across são realizadas através do relacionamento hierárquico entre dim_ponto e dim_local.

2.2 Slowly Changing Dimensions (SCDs)

Embora as dimensões geográficas (dim_local e dim_ponto) possam, em teoria, sofrer alterações ao longo do tempo (como mudanças nos nomes de municípios, criação de novos municípios, ou alterações em limites geográficos), optamos por não implementar SCDs para essas dimensões neste projeto pelas seguintes razões:

2.2.1 Justificativa para Não Implementação de SCDs

- Escopo do Projeto: O foco principal está na análise de padrões de queimadas e correlações climáticas, onde mudanças geográficas administrativas não impactam significativamente os resultados analíticos.
- Natureza dos Dados: Os dados de queimadas são eventos pontuais no tempo e espaço. O que importa é a localização exata onde o evento ocorreu no momento específico, não mudanças administrativas posteriores.
- Complexidade vs. Benefício: A implementação de SCDs adicionaria complexidade significativa ao modelo sem fornecer valor analítico proporcional para os objetivos deste projeto.
- Integridade Temporal dos Fatos: Os fatos de queimadas sempre referenciam as características geográficas apropriadas ao momento em que o evento ocorreu, garantindo consistência temporal e precisão analítica.

2.2.2 Abordagem Adotada

As dimensões geográficas mantêm seus valores conforme registrados no momento da ocorrência dos eventos. Esta abordagem:

- Preserva a integridade histórica dos dados.
- Mantém a simplicidade do modelo para consultas e análises.
- Assegura que cada fato de queimada seja sempre analisado no contexto geográfico correto de quando ocorreu.
- Facilita análises longitudinais sem a complexidade de versionamento dimensional.

2.2.3 dim_horario

Dimensão de horário compartilhada, com granularidade mínima de minuto:

- id: long, chave primária (PK).
- hora: int.
- minuto: int.
- Hierarquia: minuto \rightarrow hora

2.3 Tabela Ponte (Bridge Table) para Tempo

Devido à diferença de granularidade temporal entre os fatos Queimadas (minuto) e Clima (hora), foi criada uma tabela ponte (bridge_horario) para facilitar consultas que relacionem os dois fatos no domínio temporal. A granularidade de minuto é essencial para queimadas, pois múltiplos focos de incêndio podem ocorrer no mesmo município dentro da mesma hora, e agregá-los resultaria em perda significativa de informações analíticas sobre a intensidade e distribuição temporal dos eventos.

2.3.1 bridge_horario

- horario_minuto_fk: long, FK para dim_horario (granularidade minuto).
- horario_hora_fk: long, FK para dim_horario (granularidade hora).

2.3.2 Mecânica da Tabela Ponte

Esta tabela implementa um relacionamento muitos-para-um (60:1), onde cada registro de hora em dim_horario se relaciona com exatamente 60 registros de minutos correspondentes. Por exemplo:

- Hora 14:00 \rightarrow Minutos 14:00, 14:01, 14:02, ..., 14:59
- Hora 15:00 \rightarrow Minutos 15:00, 15:01, 15:02, ..., 15:59

Esta estrutura permite três tipos de consultas temporais:

1. Análise granular de queimadas: Consultas diretas em granularidade de minuto

2. Análise climática: Consultas em granularidade de hora
3. Análise integrada (drill-across): Agregação de queimadas por hora para comparação com dados climáticos

2.3.3 Alternativas Arquiteturais Consideradas

2.3.3.1 Abordagem 1: Fato Único com Granularidade Mínima

Criar um único fato com granularidade de minuto forçaria a interpolação artificial de dados climáticos horários, introduzindo imprecisão e aumentando desnecessariamente o volume de dados em 60 vezes.

2.3.3.2 Abordagem 2: Fato Único com Granularidade Máxima

Utilizar apenas granularidade de hora resultaria na perda de informações críticas sobre a distribuição temporal de queimadas, essencial para análises de intensidade e padrões de propagação.

2.3.3.3 Abordagem 3: Dimensões Temporais Separadas

Manter dimensões temporais completamente separadas impediria consultas drill-across, eliminando a capacidade de correlacionar eventos climáticos e de queimadas.

2.3.3.4 Justificativa da Solução Adotada

A tabela ponte oferece flexibilidade máxima mantendo a integridade dos dados originais. Permite agregações controladas quando necessário, preserva a granularidade natural de cada domínio, e facilita tanto análises especializadas quanto integradas.

2.4 Fato Queimadas

2.4.1 Granularidade

A tabela de fatos Queimadas opera no grão de um foco de incêndio detectado por satélite em um ponto geográfico específico (latitude, longitude), em um município, em uma data, em um minuto específico. Cada registro representa uma detecção individual de foco de incêndio, com suas características de risco, potência radiativa e contexto de seca. A granularidade de minuto é essencial para preservar a distribuição temporal fina dos eventos de queimadas, permitindo análises de intensidade e padrões de propagação.

2.4.2 Fatos

- risco_fogo: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA

- Justificativa: Representa uma medida de risco que deve ser calculada como média para manter representatividade estatística.
- potencia_radiativa_fogo: float, aditivo.
 - Agregação temporal: SOMA
 - Agregação espacial: SOMA
 - Justificativa: Representa energia liberada em megawatts, que pode ser somada para obter energia total por período ou região.
- dias_sem_chuva: int, semi-aditivo.
 - Agregação temporal: MAX (para obter o período mais longo sem chuva)
 - Agregação espacial: MÉDIA (para obter condição média da região)
 - Justificativa: Temporalmente, o valor máximo indica o período crítico de seca; espacialmente, a média fornece condição representativa da região.

2.4.3 Dimensões

A tabela de fatos se conecta às seguintes dimensões conformadas:

- dim_data
- dim_local
- dim_horario (com granularidade de minuto)

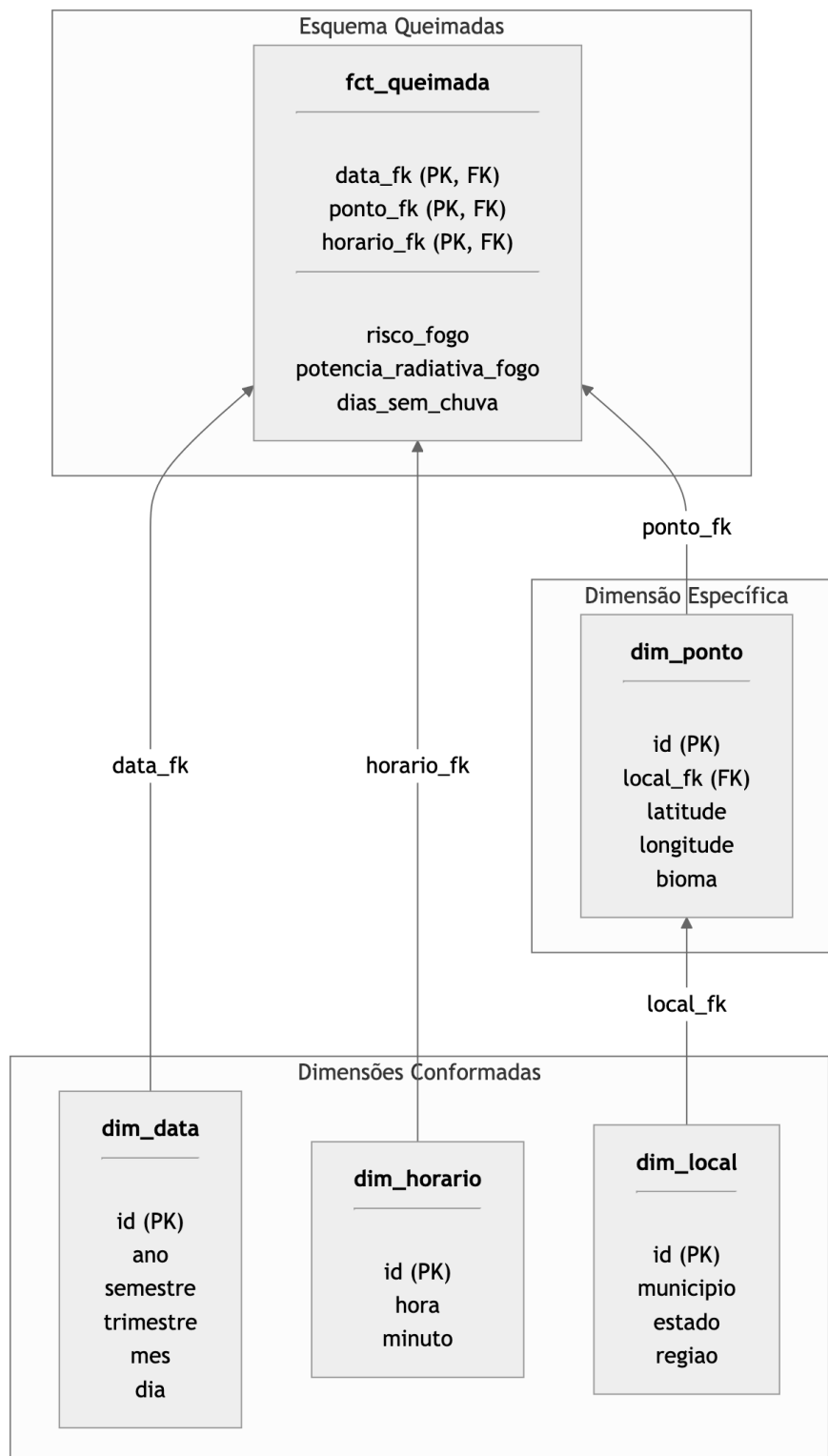


Figura 1 – Esquema Estrela para Queimadas

2.5 Fato Clima

2.5.1 Granularidade

A tabela de fatos Clima opera no grão de medições climáticas agregadas por município, por data, por hora. Cada registro representa as condições climáticas médias observadas em um município durante uma hora específica, incluindo temperatura, precipitação, umidade,

vento e qualidade do ar. A granularidade horária é apropriada para dados climáticos, pois representa a resolução temporal típica das estações meteorológicas e dos modelos climáticos, equilibrando precisão analítica com volume de dados.

2.5.2 Fatos

- temperatura: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Temperatura é uma medida intensiva que deve ser calculada como média ponderada.
- precipitacao: float, semi-aditivo.
 - Agregação temporal: SOMA (para precipitação acumulada no período)
 - Agregação espacial: MÉDIA (para precipitação média da região)
 - Justificativa: Temporalmente representa acumulação; espacialmente representa condição média regional.
- umidade_relativa: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Medida intensiva expressa em percentual, agregada por média.
- vento_velocidade: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Velocidade deve ser agregada por média para manter representatividade física.
- vento_direcao: int, não aditivo.
 - Agregação temporal: MODA ou MÉDIA_CIRCULAR
 - Agregação espacial: MODA ou MÉDIA_CIRCULAR
 - Justificativa: Direção em graus azimutais requer agregação circular ou moda para manter significado físico.
- co_ppb: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Concentração de monóxido de carbono é medida intensiva, agregada por média para representar condições médias de qualidade do ar.
- no2_ppb: float, não aditivo.

- Agregação temporal: MÉDIA
- Agregação espacial: MÉDIA
- Justificativa: Concentração de dióxido de nitrogênio é medida intensiva, agregada por média para análises de qualidade do ar.
- o3_ppb: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Concentração de ozônio é medida intensiva, crucial para análises de qualidade do ar e correlações com queimadas.
- pm25_ugm3: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Material particulado fino (PM2.5) é indicador crítico de qualidade do ar, especialmente relevante em áreas com queimadas.
- so2_ugm3: float, não aditivo.
 - Agregação temporal: MÉDIA
 - Agregação espacial: MÉDIA
 - Justificativa: Concentração de dióxido de enxofre é medida intensiva importante para análises de qualidade do ar e impactos ambientais.

2.5.3 Dimensões

A tabela de fatos se conecta às seguintes dimensões conformadas:

- dim_data
- dim_local
- dim_horario (com granularidade de hora, via bridge_horario)

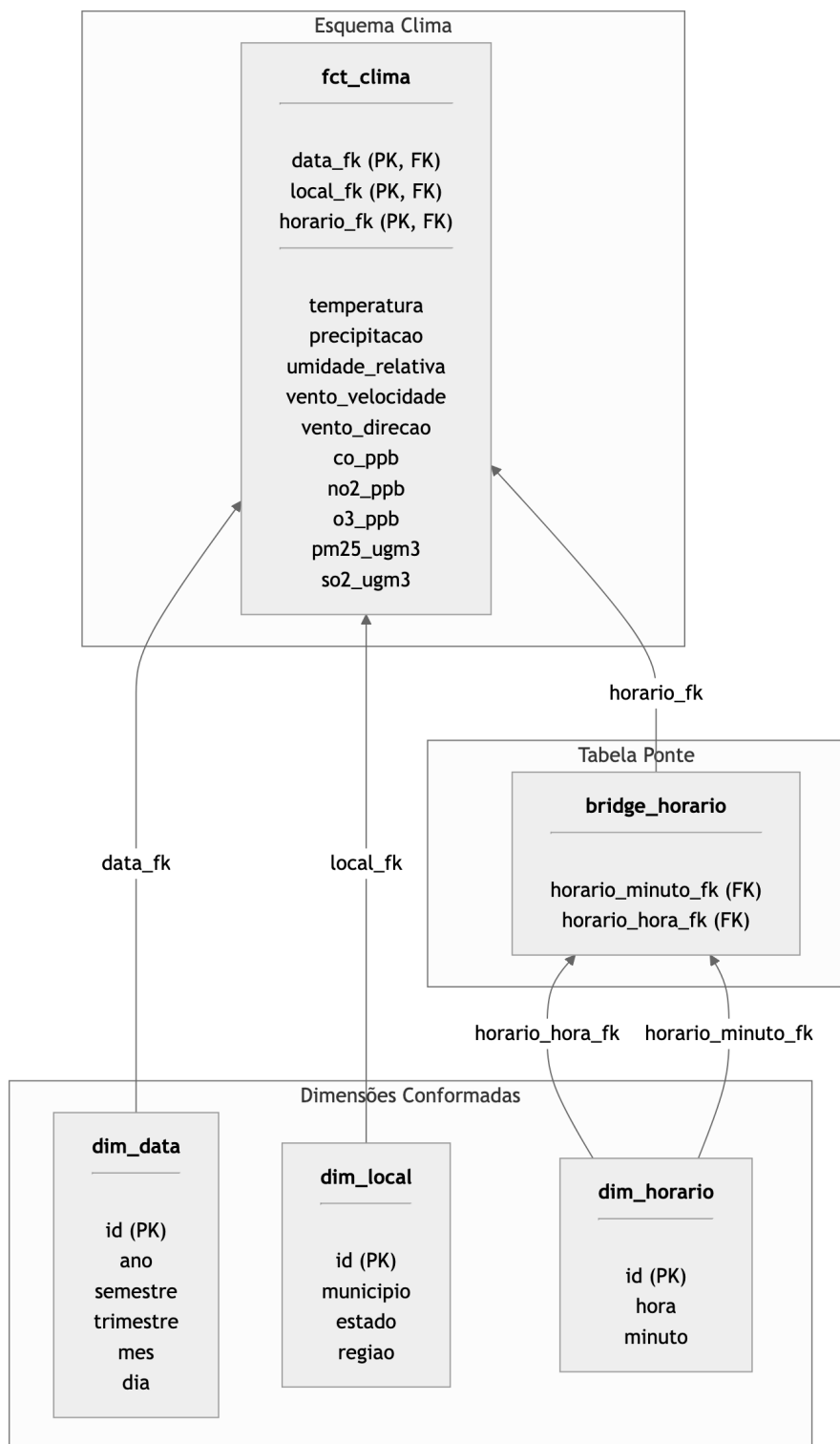


Figura 2 – Esquema Estrela para Clima

2.6 Constelação de Fatos

A constelação final integra os dois fatos através das dimensões conformadas *dim_data* e *dim_local*, e utiliza a tabela ponte *bridge_horario* para resolver a diferença de granularidade temporal. Esta arquitetura permite análises como:

- Correlação entre condições climáticas e ocorrência de queimadas por município e período.

- Análise temporal em múltiplas granularidades (minuto, hora, dia, mês, etc.).
- Drill-across entre fatos para análises integradas de clima e queimadas.

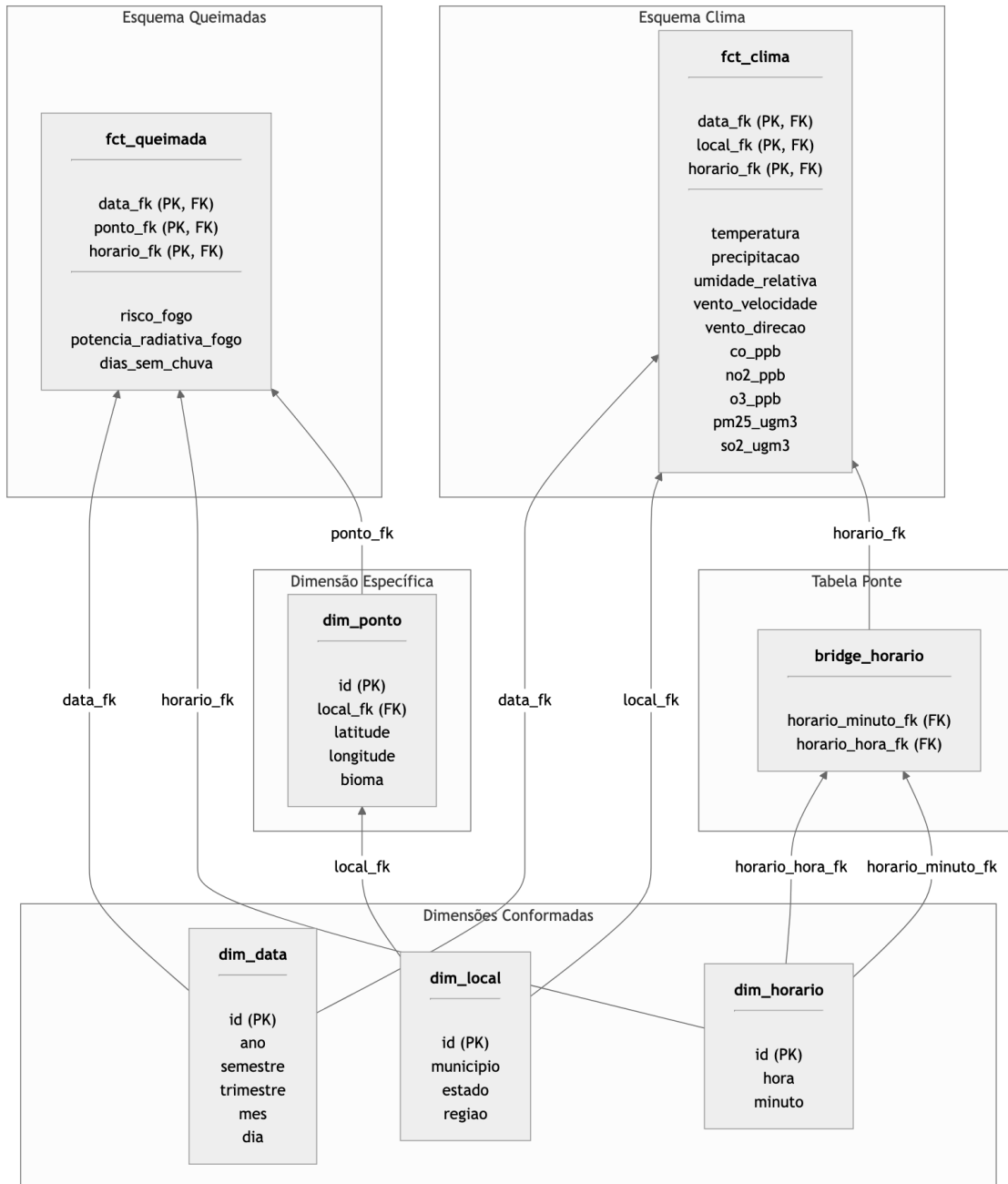


Figura 3 – Constelação de Fatos com Dimensões Conformadas e Tabela Ponte

3 Consultas

A seguir, são apresentados exemplos de operações OLAP que podem ser executadas sobre a constelação de fatos para extrair insights analíticos. As consultas apresentadas são de alto nível, focando nos conceitos e hierarquias, sendo independentes de implementação específica.

3.1 Drill-Down com agregação de contagem

Visão: Número de Focos de Incêndio por Estado por Ano \rightarrow Visão: Número de Focos de Incêndio por Município por Ano.

3.1.1 Motivação

Após identificar Estados com maior incidência de focos de incêndio (contagem de registros na tabela de fatos Queimadas), o analista pode "descer" na hierarquia geográfica de `dim_local` (estado \rightarrow município) para avaliar quais municípios são os maiores contribuintes, permitindo o direcionamento de ações de fiscalização e prevenção.

3.2 Roll-Up com agregação de média

Visão: Média de Precipitação por Estado por Mês \rightarrow Visão: Média de Precipitação por Região por Ano.

3.2.1 Motivação

Agregar dados mensais para uma visão anual (mês \rightarrow ano na hierarquia de `dim_data`) e de estado para região (estado \rightarrow região na hierarquia de `dim_local`) permite a identificação de tendências macro, suavizando variações de curto prazo e revelando padrões climáticos de larga escala. A agregação de precipitação utiliza média espacial (conforme definido no esquema Clima).

3.3 Pivot

Visão: Total de Potência Radiativa por Estado por Mês \rightarrow Visão: Total de Potência Radiativa por Mês por Estado.

3.3.1 Motivação

A primeira visão facilita a análise da série temporal do impacto das queimadas para cada estado, utilizando a medida aditiva `potencia_radiativa_fogo` do fato Queimadas. Ao "pivotar" a tabela (reorientando as dimensões `dim_local` e `dim_data`), a nova visão facilita a comparação do impacto entre os estados para um mesmo mês, identificando áreas críticas em cada período.

3.4 Slice com agregação de média espacial

Visão: Precipitação por Estado por Mês \rightarrow Visão: Precipitação Média por Estado para o mês de Setembro.

3.4.1 Motivação

Permite isolar um "pedaço" do cubo de dados para análise focada. Neste caso, seleciona-se o mês de Setembro através de `dim_data` (tipicamente um mês crítico de seca) para analisar a precipitação média espacial (conforme agregação definida no fato Clima) em

cada estado, auxiliando no planejamento de ações preventivas. Esta operação utiliza as dimensões conformadas para filtrar o período de interesse.

3.5 Drill-Down com análise hierárquica por bioma

Visão: Média de Risco de Fogo por Estado por Ano → Visão: Média de Risco de Fogo por Município por Bioma por Ano.

3.5.1 Motivação

Demonstra a utilização da hierarquia geográfica especializada do fato Queimadas: através de `dim_ponto`, que mantém informações de bioma e se relaciona hierarquicamente com `dim_local`. Esta consulta permite descer do nível de estado para município, adicionando a dimensão de bioma, revelando que municípios específicos podem apresentar riscos distintos em diferentes ecossistemas (Amazônia, Cerrado, etc.). A medida `risco_fogo` utiliza agregação por média, conforme definido no esquema. Esta análise é essencial para políticas de conservação direcionadas por tipo de bioma.

3.6 Slice multi-dimensional com qualidade do ar

Visão: Concentrações Médias de PM2.5, CO e O3 por Município por Dia → Visão: Concentrações Médias de PM2.5, CO e O3 para Municípios da Região Norte durante a Estação de Seca (Julho-Setembro).

3.6.1 Motivação

Demonstra uma operação de slice em múltiplas dimensões do fato Clima, isolando região geográfica específica através de `dim_local` e período sazonal através de `dim_data` (utilizando os atributos analíticos de estação e mês). As medidas de qualidade do ar (`pm25_ugm3`, `co_ppb`, `o3_ppb`) são não-aditivas e agregadas por média. Esta análise é crítica para avaliar impactos ambientais e de saúde pública em períodos de maior incidência de queimadas, permitindo comparações entre diferentes poluentes atmosféricos simultaneamente.

3.7 Drill-Across com agregação de média

Visão: Média de Umidade Relativa por Região por Trimestre + Visão: Média de Potência Radiativa por Região por Trimestre → Visão: Média de Umidade Relativa e Média de Potência Radiativa por Região por Trimestre.

3.7.1 Motivação

Permite a análise de correlação entre métricas de diferentes tabelas de fato (Clima e Queimadas) através das dimensões conformadas `dim_local` e `dim_data`. As medidas `umidade_relativa` (fato Clima) e `potencia_radiativa_fogo` (fato Queimadas) podem ser agregadas simultaneamente utilizando a hierarquia região de `dim_local` e trimestre de `dim_data`. A `bridge_horario` permite integrar dados de diferentes granularidades temporais (hora para Clima, minuto para Queimadas) ao agregar para níveis superiores. Esta

consulta é fundamental para validar hipóteses, como a de que trimestres com menor umidade relativa apresentam incêndios com maior potência radiativa.

3.8 Análise de Pico de Queimadas por Hora usando Bridge Table

Visão: Contagem de Focos de Incêndio por Município por Hora com correlação Temperatura e PM2.5 Médias por Município por Hora.

3.8.1 Motivação

Esta consulta demonstra explicitamente o uso da tabela ponte (`bridge_horario`) para integrar eventos de queimadas (granularidade de minuto) com dados climáticos (granularidade de hora). A tabela ponte realiza o mapeamento 60:1 entre minutos e horas, permitindo agregar múltiplos focos de incêndio detectados em diferentes minutos da mesma hora. Por exemplo, se um município teve focos detectados às 14:23, 14:37 e 14:54, a `bridge_horario` permite agregar esses três eventos (via `horario_minuto_fk`) para a hora 14:00 (via `horario_hora_fk`), possibilitando a contagem total de focos por hora e sua correlação com as condições climáticas médias dessa hora (temperatura e concentração de PM2.5). Esta análise é essencial para identificar padrões temporais de intensificação de queimadas e seus impactos imediatos na qualidade do ar, revelando, por exemplo, que determinadas horas do dia apresentam maior concentração de focos e consequentemente piores níveis de poluição atmosférica.

Referências

- 1 MAPBIOMAS. Área queimada no Brasil cresce 79% em 2024 e supera os 30 milhões de hectares. 2025. <<https://brasil.mapbiomas.org/2025/01/22/area-queimada-no-brasil-cresce-79-em-2024-e-supera-os-30-milhoes-de-hectares/>>. Acessado em 01 de outubro de 2025.
- 2 Instituto Nacional de Pesquisas Espaciais. Monitoramento de Queimadas do INPE. 2024. <<https://basedosdados.org/dataset/f06f3cdc-b539-409b-b311-1ff8878fb8d9?table=a3696dc2-4dd1-4f7e-9769-6aa16a1556b8>> ou <https://console.cloud.google.com/bigquery?ws=!1m4!1m3!3m2!1sbasedosdados!2sbr_inpe_sisam>. Dados sobre focos de queimada no Brasil.
- 3 Instituto Nacional de Pesquisas Espaciais. Sistema de Informações Ambientais do INPE (Sisam). 2024. <https://console.cloud.google.com/bigquery?ws=!1m5!1m4!4m3!1sbasedosdados!2sbr_inpe_queimadas!3smicrodados>. Dados climáticos brasileiros.