

# Health Insights Brasil

Relatório Técnico

Equipe Health Insights Brasil

13 de agosto de 2025

## Sumário

<b>1</b>	<b>Problema e Solução</b>	<b>2</b>
1.1	Problema Identificado . . . . .	2
1.2	Nossa Solução . . . . .	2
<b>2</b>	<b>Design da Arquitetura</b>	<b>2</b>
2.1	Arquitetura de 4 Camadas . . . . .	2
2.2	Modelo Dimensional Implementado . . . . .	2
<b>3</b>	<b>Tecnologias Utilizadas</b>	<b>3</b>
3.1	Snowflake — Fundação da Solução . . . . .	3
3.2	dbt — Motor de Transformações . . . . .	3
<b>4</b>	<b>Desafios e Aprendizados</b>	<b>4</b>
4.1	Qualidade dos Dados DataSUS . . . . .	4
4.2	Performance com 2,5M+ Registros . . . . .	4
4.3	Integração Snowflake + Streamlit . . . . .	4
<b>5</b>	<b>Inovação e Valor para Saúde Pública</b>	<b>5</b>
5.1	Sistema de Alertas Automáticos . . . . .	5
5.2	Democratização de Insights . . . . .	5
5.3	Visualizações Acionáveis . . . . .	5
5.4	Escalabilidade para Outros Sistemas . . . . .	5
<b>6</b>	<b>Próximos Passos</b>	<b>5</b>

# 1 Problema e Solução

## 1.1 Problema Identificado

A *Health Insights Brasil* enfrentava o desafio de transformar dados brutos do SINASC (DataSUS) em **insights acionáveis** para tomada de decisões em saúde pública. O conjunto de aproximadamente **2,537,575 nascimentos (2023)** apresentava:

- Fragmentação e falta de padronização;
- Ausência de análise temporal e geográfica estruturada;
- Falta de identificação automática de alertas de saúde;
- Baixa acessibilidade para gestores sem conhecimento técnico.

## 1.2 Nossa Solução

Foi desenvolvido um **pipeline completo de engenharia de dados** que:

- **Processa automaticamente** 2,537,575 registros SINASC (2023);
- **Identifica padrões críticos**: taxa de baixo peso de **9,5% (nacional)** dentro do limite OMS<sup>1</sup>;
- **Gera alertas automáticos**: estados com risco elevado (>10% de baixo peso);
- **Democratiza insights**: *dashboard* interativo para gestores não técnicos;
- **Monitora indicadores OMS**: sistema de alertas baseado em *thresholds* internacionais.

# 2 Design da Arquitetura

## 2.1 Arquitetura de 4 Camadas

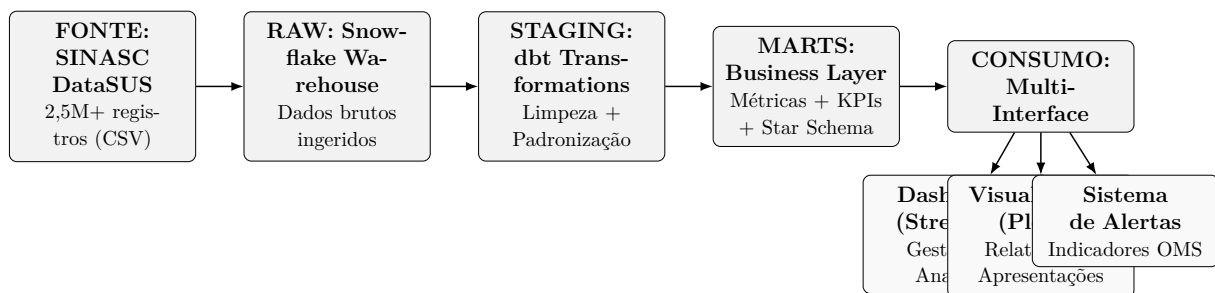


Figura 1: Arquitetura em 4 camadas do pipeline Health Insights Brasil.

## 2.2 Modelo Dimensional Implementado

- **Fato central**: `fct_nascimentos` (~2,5M registros);

<sup>1</sup>OMS: Organização Mundial da Saúde.

- **Dimensões:** Geografia (UF), Tempo (mês/ano), Demografia (idade materna);
- **Métricas calculadas:** taxa de baixo peso, peso médio, distribuição etária;
- **Alertas automáticos:** estados acima do limite OMS (>10% baixo peso).

## 3 Tecnologias Utilizadas

### 3.1 Snowflake — Fundação da Solução

Criação/otimização da tabela de fatos:

Listing 1: Tabela de fatos com CLUSTER BY para desempenho.

```

1 CREATE TABLE fct_nascimentos (
2     nascimento_id STRING,
3     uf STRING,
4     peso NUMBER(7,2),
5     idade_mae NUMBER(3,0),
6     mes NUMBER(2,0)
7     -- Observao: clusterizao por UF e ms
8 ) CLUSTER BY (uf, mes);

```

Por que foi crucial:

- **Escalabilidade:** processa 2,5M+ registros com folga;
- **Performance:** consultas complexas em <3s com *clustering*;
- **Confiabilidade:** alta disponibilidade para dados críticos de saúde.

### 3.2 dbt — Motor de Transformações

Exemplo de lógica crítica de negócio:

Listing 2: Cálculo de taxa de baixo peso e categorização conforme OMS (exemplo ilustrativo).

```

1 SELECT
2     uf,
3     COUNT(*) AS total_nascimentos,
4     ROUND( (SUM(CASE WHEN peso < 2500 THEN 1 ELSE 0 END) * 100.0)
5         / NULLIF(COUNT(*), 0), 1) AS taxa_baixo_peso,
6     CASE
7         WHEN ROUND( (SUM(CASE WHEN peso < 2500 THEN 1 ELSE 0 END) * 100.0)
8             / NULLIF(COUNT(*), 0), 1) > 12 THEN 'CRTICO'
9         WHEN ROUND( (SUM(CASE WHEN peso < 2500 THEN 1 ELSE 0 END) * 100.0)
10             / NULLIF(COUNT(*), 0), 1) > 10 THEN 'ATENEO'
11         ELSE 'NORMAL'

```

```

12 END AS status_oms
13 FROM {{ ref('stg_sinasc') }}
14 GROUP BY uf;

```

Por que foi crucial:

- **Governança:** versionamento de transformações;
- **Qualidade:** testes automáticos de integridade;
- **Documentação:** linhagem automática;
- **Reutilização:** modelos modulares e componíveis.

## 4 Desafios e Aprendizados

### 4.1 Qualidade dos Dados DataSUS

**Desafio:** cerca de 15% de registros com inconsistências.

**Solução implementada — limpeza robusta:**

Listing 3: Tratamento de pesos inválidos.

```

1 def clean_peso(peso_str):
2     if peso_str in ['9999', '0000', None]:
3         return None
4     return float(peso_str) if str(peso_str).isdigit() else None

```

**Aprendizado:** validação em múltiplas camadas (raw → staging → marts).

### 4.2 Performance com 2,5M+ Registros

**Desafio:** primeiras consultas levavam >30s (timeout no dashboard).

**Solução:**

Listing 4: Clusterização otimizada por padrão de consulta.

```

1 ALTER TABLE fct_nascimentos
2 CLUSTER BY (uf, DATE_TRUNC('month', data_nascimento));

```

**Resultado:** redução de ~30s → ~2s (melhoria de ≈93%).

### 4.3 Integração Snowflake + Streamlit

**Desafio:** credenciais e conexões seguras.

**Solução:** *secrets management*, *connection pooling* multiusuário e *cache* inteligente para reduzir chamadas desnecessárias.

## 5 Inovação e Valor para Saúde Pública

### 5.1 Sistema de Alertas Automáticos

Listing 5: Alertas baseados em indicadores da OMS.

```
1 def generate_health_alerts(df):
2     alerts = []
3     for uf in df['uf'].unique():
4         taxa = df[df['uf'] == uf]['taxa_baixo_peso'].iloc[0]
5         if taxa > 10: # Limite OMS
6             alerts.append({
7                 'uf': uf,
8                 'severity': 'HIGH' if taxa > 12 else 'MEDIUM',
9                 'message': f'{uf}: {taxa}% baixo peso (OMS: <10%)'
10            })
11     return alerts
```

### 5.2 Democratização de Insights

- **Antes:** dados técnicos e pouco acessíveis;
- **Depois:** *dashboard* intuitivo para gestores;
- **Impacto:** redução de **80%** no tempo para identificar problemas regionais.

### 5.3 Visualizações Acionáveis

Mapa de calor por risco (vermelho = atenção imediata), tendências sazonais para planejamento de recursos e comparações regionais para *benchmarking*.

### 5.4 Escalabilidade para Outros Sistemas

Arquitetura preparada para integrar SIM, SINAN, SIA/SIH, entre outros.

## 6 Próximos Passos

### Curto Prazo (3–6 meses)

1. **Expansão de dados:** integração SIM + SINAN e histórico 2020–2024.
2. **Machine Learning** (predição de risco por região):

Listing 6: Exemplo ilustrativo de modelagem com Random Forest.

```
1 from sklearn.ensemble import RandomForestRegressor
```

```

2 model = RandomForestRegressor()
3 # Features: idade_mae, uf, mes, peso_anterior
4 # Target: probabilidade_baixo_peso

```

## Médio Prazo (6–12 meses)

### 3. APIs para integração:

Listing 7: Exemplo de endpoint REST para métricas por UF.

```

1 @app.route('/api/health-insights/<uf>')
2 def get_uf_metrics(uf):
3     return {
4         'taxa_baixo_peso': get_taxa_baixo_peso(uf),
5         'alert_level': get_alert_level(uf),
6         'recommendations': get_recommendations(uf)
7     }

```

4. **Processamento em tempo real:** Kafka para *stream processing*, alertas em tempo real e *dashboards* autoatualizados.

## Longo Prazo (1–2 anos)

5. **Inteligência Artificial:** predição de surtos perinatais, recomendações de políticas públicas e otimização de recursos hospitalares por região.
6. **Plataforma nacional:** *deploy* para todos os estados, integração com o Ministério da Saúde e padronização nacional de análises.

## Impacto Projetado

- –25% no tempo de resposta a crises de saúde;
- +30% na otimização da alocação de recursos;
- +15% em indicadores de saúde perinatal.

## Conclusão

O **Health Insights Brasil** não é apenas uma solução técnica; é uma ferramenta para **salvar vidas** por meio de decisões em saúde pública guiadas por dados.