



TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

ARTHUR GUIMARAES MIRANDA PRATES

DANIEL DA SILVA OLIVEIRA

DAVI MONTEIRO CARVALHO

JÂMERSON DE SOUZA SILVA

**ESTUDO CORRETIVO DESEMPENHADO SOBRE BASE DE DADOS
DA FAB UTILIZANDO COMANDOS BÁSICOS MySQL**

GUANAMBI

2025

ARTHUR GUIMARAES MIRANDA PRATES

DANIEL DA SILVA OLIVEIRA

DAVI MONTEIRO CARVALHO

JÂMERSON DE SOUZA SILVA

**ESTUDO CORRETIVO DESEMPENHADO SOBRE BASE DE DADOS
DA FAB UTILIZANDO COMANDOS BÁSICOS MySQL**

Trabalho apresentado ao Instituto Federal de Educação Ciência e Tecnologia Baiano, como requisito de avaliação parcial da disciplina Fundamentos de Bancos de Dados e, por conseguinte, a obtenção dos seus créditos.

Profª. Sarah Moniky Silva Ribeiro

GUANAMBI

2025

SUMÁRIO

1	CONSIDERAÇÕES GERAIS.....	4
2	PROBLEMAS IDENTIFICADOS.....	5
2.1	REDUNDÂNCIA E INCONSISTÊNCIA EM CHAVES DE RELACIONAMENTO	5
2.2	AUSENCIA DE REGISTROS	5
2.3	INCOMPLETUDE DE DADOS E CAMPOS COM VALORES NULOS	5
2.4	INADEQUAÇÃO DO FORMATO DE ARMAZENAMENTO DE DATAS	6
3	PASSO 3: DEFINIÇÃO DE CHAVES PRIMÁRIAS, ESTRANGEIRAS E CORREÇÃO DE TIPOS	7
3.1	IDENTIFICAÇÃO E DEFINIÇÃO DE CHAVES PRIMÁRIAS.....	7
3.2	IMPLEMENTAÇÃO DE CONSTRAINTS UNIQUE E NOT NULL	7
3.3	DEFINIÇÃO DE CHAVES ESTRANGEIRAS	8
3.4	CORREÇÃO DE TIPOS DE DADOS: CONVERSÃO DE DATAS	9
3.5	RESULTADOS E VALIDAÇÕES	10

1 CONSIDERAÇÕES GERAIS

O presente trabalho tem como objetivo a exploração e manipulação de uma base de dados real de ocorrências aeronáuticas no Brasil, disponibilizada pela Força Aérea Brasileira (FAB) através do portal de dados abertos do governo federal. Esta base consolida informações sobre acidentes, incidentes e incidentes graves envolvendo a aviação civil brasileira, constituindo-se em um repositório essencial para análises de segurança operacional e investigação aeronáutica.

Dados reais apresentam desafios significativamente distintos daqueles encontrados em bases didáticas. Inconsistências estruturais, redundâncias e ausência de padronização são características comuns em ambientes de produção, especialmente em sistemas legados ou bases consolidadas a partir de múltiplas fontes. Diante desse cenário, optou-se deliberadamente por manter a estrutura original dos dados, simulando uma situação profissional na qual uma equipe técnica presta serviços de otimização e análise para uma instituição já estabelecida, sem autorização para alterações substanciais no modelo de dados existente.

Esta abordagem metodológica justifica-se pela necessidade de desenvolver competências práticas em ambientes não idealizados. Em contextos profissionais reais, raramente é possível redesenhar completamente uma base de dados em operação. O trabalho técnico frequentemente consiste em compreender limitações estruturais, implementar melhorias incrementais de desempenho e extrair informações relevantes mesmo diante de imperfeições no modelo relacional. Assim, as intervenções realizadas concentraram-se na definição adequada de chaves primárias e estrangeiras, na implementação de restrições de integridade referencial e na otimização de consultas, preservando a totalidade dos dados originais.

Durante o desenvolvimento do projeto, foram identificadas diversas particularidades na estrutura da base, incluindo a utilização de múltiplos códigos de ocorrência para estabelecer relacionamentos entre tabelas e a presença de campos com valores nulos ou inconsistentes. Tais características, embora representem desafios analíticos, não inviabilizaram a execução das tarefas propostas e serão discutidas ao longo deste documento. A manutenção integral dos dados originais, mesmo com suas imperfeições, permite não apenas cumprir os objetivos pedagógicos da disciplina de Fundamentos de Banco de Dados, mas também simular condições autênticas de trabalho em ambientes corporativos e institucionais.

As análises desenvolvidas abrangem desde consultas básicas de listagem até agregações complexas envolvendo múltiplas tabelas, permitindo extrair informações estratégicas sobre padrões de ocorrências, distribuição geográfica de incidentes, características das aeronaves envolvidas e origem dos voos com maior incidência de eventos adversos. Os resultados obtidos demonstram a viabilidade de aplicação das técnicas de modelagem relacional e linguagem SQL em contextos práticos de análise de dados aeronáuticos, contribuindo para a compreensão dos fundamentos de sistemas gerenciadores de banco de dados e sua aplicabilidade em domínios especializados.

2 PROBLEMAS IDENTIFICADOS

2.1 REDUNDÂNCIA E INCONSISTÊNCIA EM CHAVES DE RELACIONAMENTO

Durante a análise da estrutura relacional da base de dados, identificou-se um problema crítico de integridade referencial envolvendo os códigos de ocorrência utilizados para estabelecer relacionamentos entre tabelas. A tabela principal `tabela_ocorrencia` apresenta uma sequência de códigos que, embora majoritariamente contínua, exibe lacunas não documentadas em sua numeração. Observou-se, por exemplo, a progressão 87402, 87403, 87405, evidenciando a ausência do código 87404 no conjunto de chaves primárias disponíveis.

2.2 AUSENCIA DE REGISTROS

Paradoxalmente, ao examinar a tabela `tabela_tipo_ocorrencia`, constatou-se a presença de registros referenciando justamente códigos ausentes na tabela principal, como o código 87404 mencionado, que aparece duplicado em duas linhas distintas. Esta situação configura uma violação direta do princípio de integridade referencial, fundamental para sistemas de gerenciamento de banco de dados relacionais. A existência de chaves estrangeiras apontando para chaves primárias inexistentes compromete a confiabilidade das consultas que envolvem junções entre estas tabelas e pode resultar em análises incompletas ou estatisticamente enviesadas.

2.3 INCOMPLETUDE DE DADOS E CAMPOS COM VALORES NULOS

A análise exploratória da base de dados revelou um padrão preocupante de incompletude informacional, manifestado através da presença massiva de valores NULL em múltiplas colunas, particularmente nas tabelas que armazenam detalhes técnicos e operacionais das aeronaves envolvidas em ocorrências. Campos como, `aeronave_motor_tipo`, `aeronave_pmd`

(peso máximo de decolagem) e `aeronave_ano_fabricacao` apresentam lacunas significativas que comprometem a capacidade de realizar análises técnicas aprofundadas sobre perfis de risco associados a características específicas de aeronaves.

As tabelas periféricas apresentam situação ainda mais crítica. A `tabela_fator_contribuinte`, responsável por registrar causas e condições que contribuíram para a ocorrência, contém proporção elevada de registros com campos descritivos nulos ou preenchidos com valores genéricos. Considerando que a identificação precisa de fatores contribuintes constitui elemento fundamental para políticas preventivas de segurança operacional, esta incompletude representa limitação substancial para a utilização da base em análises preditivas ou estudos estatísticos de causalidade.

2.4 INADEQUAÇÃO DO FORMATO DE ARMAZENAMENTO DE DATAS

Uma das inconsistências estruturais mais significativas identificadas refere-se ao armazenamento de informações temporais. Notadamente `ocorrencia_dia`, `divulgacao_dia_publicacao` e campos relacionados a horários — encontram-se definidos com tipos de dados textuais (VARCHAR ou TEXT) em vez do tipo nativo DATE ou DATETIME oferecido pelo MySQL. Esta decisão de design, embora tecnicamente funcional, acarreta múltiplas limitações operacionais e analíticas. Primeiramente, impossibilita a utilização direta de funções temporais nativas do SGBD. Consultas que busquem identificar tendências sazonais, calcular intervalos entre eventos ou agrupar ocorrências por períodos específicos requerem conversões explícitas através de funções como `STR_TO_DATE()`, introduzindo complexidade adicional e potencial degradação de desempenho em consultas sobre grandes volumes de dados.

A hipótese mais plausível para justificar esta escolha arquitetural relaciona-se ao público-alvo da base de dados e aos sistemas que a alimentam. Considerando que as informações são acessadas por múltiplos órgãos muitos dos quais não necessariamente técnicos ou especializados em aviação, a apresentação de datas no formato brasileiro dd/mm/aaaa pode ter sido priorizada para facilitar a legibilidade e reduzir erros de interpretação por usuários não familiarizados com formatos internacionais (aaaa-mm-dd). É possível que sistemas legados ou interfaces de entrada de dados mais antigas tenham imposto este formato, e que conversões para tipos nativos não tenham sido implementadas na camada de persistência.

Outra possibilidade envolve questões de compatibilidade com sistemas externos ou requisitos de exportação de dados para planilhas eletrônicas, onde o formato textual brasileiro

pode ser mais imediatamente compreensível para analistas que não trabalham diretamente com bancos de dados. Não obstante estas justificativas pragmáticas, permanece a recomendação técnica de que conversões de tipo sejam implementadas na camada de aplicação, armazenando-se os dados em formato nativo no banco e formatando-os apenas na apresentação ao usuário final. Para os propósitos deste projeto, optou-se por respeitar o formato original.

3 PASSO 3: DEFINIÇÃO DE CHAVES PRIMÁRIAS, ESTRANGEIRAS E CORREÇÃO DE TIPOS

3.1 IDENTIFICAÇÃO E DEFINIÇÃO DE CHAVES PRIMÁRIAS

A definição adequada de chaves primárias constitui fundamento essencial para garantir a unicidade e integridade dos registros em um sistema de gerenciamento de banco de dados relacional. Após análise criteriosa da estrutura fornecida, identificou-se que a tabela `tabela_ocorrencia` deveria funcionar como entidade central do modelo, utilizando `codigo_ocorrencia` como sua chave primária natural.

A implementação foi realizada através do comando:

```
ALTER TABLE tabela_ocorrencia  
ADD PRIMARY KEY (codigo_ocorrencia);
```

Esta definição assegura que cada ocorrência aeronáutica possua um identificador único e imutável, impossibilitando a duplicação de registros e estabelecendo a base para relacionamentos consistentes com as demais tabelas do sistema. A escolha de `codigo_ocorrencia` como chave primária justifica-se por sua natureza sequencial, sua presença obrigatória em todos os registros e sua função de identificador oficial nos processos investigativos da aviação civil brasileira.

3.2 IMPLEMENTAÇÃO DE CONSTRAINTS UNIQUE E NOT NULL

Uma particularidade estrutural significativa da base de dados reside na presença de múltiplos códigos de ocorrência (`codigo_ocorrencia1`, `codigo_ocorrencia2`, `codigo_ocorrencia3`, `codigo_ocorrencia4`) destinados a estabelecer relacionamentos com tabelas periféricas específicas. Embora esta abordagem represente redundância conceitual — dado que um único código deveria ser suficiente para todos os relacionamentos —, optou-se por respeitar a arquitetura original, implementando constraints que assegurassem a integridade destes campos.

Para cada código de relacionamento, foram definidas restrições de unicidade e obrigatoriedade:

```
ALTER TABLE tabela_ocorrenci  
ADD UNIQUE (codigo_ocorrencial);
```

```
ALTER TABLE tabela_ocorrenci  
MODIFY COLUMN codigo_ocorrencial INT NOT NULL;
```

Este processo foi replicado para `codigo_ocorrenci2`, `codigo_ocorrenci3` e `codigo_ocorrenci4`, garantindo que cada um destes campos contenha valores únicos e não nulos. A constraint `UNIQUE` previne duplicações que comprometeriam relacionamentos um-para-muitos, enquanto a restrição `NOT NULL` assegura que nenhum registro principal exista sem a capacidade de estabelecer vínculos com suas tabelas dependentes.

As mesmas modificações foram aplicadas às colunas correspondentes nas tabelas periféricas, assegurando consistência estrutural em ambos os lados dos relacionamentos:

```
ALTER TABLE tabela_tipo_ocorrenci  
MODIFY COLUMN codigo_ocorrencial INT NOT NULL;
```

```
ALTER TABLE tabela_aeronave  
MODIFY COLUMN codigo_ocorrenci2 INT NOT NULL;
```

```
ALTER TABLE tabela_fator_contribuinte  
MODIFY COLUMN codigo_ocorrenci3 INT NOT NULL;
```

```
ALTER TABLE tabela_recomendacao_seguranca  
MODIFY COLUMN codigo_ocorrenci4 INT NOT NULL;
```

3.3 DEFINIÇÃO DE CHAVES ESTRANGEIRAS

A implementação de chaves estrangeiras representa mecanismo fundamental para manutenção da integridade referencial, assegurando que relacionamentos entre tabelas permaneçam consistentes e que operações de modificação ou exclusão não comprometam a coerência do conjunto de dados. Entretanto, conforme discutido anteriormente, a base apresentava inconsistências pré-existentes que inviabilizariam a criação de constraints de integridade referencial sob condições normais de operação.

Para contornar esta limitação sem comprometer a estrutura original dos dados, utilizou-se o comando:

```
SET FOREIGN_KEY_CHECKS = 0;
```

Esta diretiva desabilita temporariamente a verificação de integridade referencial, permitindo a criação de foreign keys mesmo na presença de registros órfãos — situação onde

chaves estrangeiras apontam para chaves primárias inexistentes. Embora esta abordagem seja tecnicamente desaconselhada em ambientes de produção ideais, mostrou-se necessária dada a impossibilidade de correção retroativa dos dados fornecidos.

As chaves estrangeiras foram então estabelecidas conforme o modelo relacional proposto:

```
ALTER TABLE tabela_tipo_ocorrencia
ADD FOREIGN KEY (codigo_ocorrencial)
REFERENCES tabela_ocorrencia(codigo_ocorrencial);

ALTER TABLE tabela_aeronave
ADD FOREIGN KEY (codigo_ocorrencia2)
REFERENCES tabela_ocorrencia(codigo_ocorrencia2);

ALTER TABLE tabela_fator_contribuinte
ADD FOREIGN KEY (codigo_ocorrencia3)
REFERENCES tabela_ocorrencia(codigo_ocorrencia3);

ALTER TABLE tabela_recomendacao_seguranca
ADD FOREIGN KEY (codigo_ocorrencia4)
REFERENCES tabela_ocorrencia(codigo_ocorrencia4);
```

Estes relacionamentos estabelecem que cada registro nas tabelas periféricas deve, idealmente, corresponder a uma ocorrência válida na tabela principal. A partir da definição destas constraints, novas inserções serão validadas pelo SGBD, prevenindo a criação de registros órfãos adicionais, embora os pré-existentis permaneçam tolerados pelo sistema.

3.4 CORREÇÃO DE TIPOS DE DADOS: CONVERSÃO DE DATAS

Uma das intervenções mais significativas realizadas no âmbito deste projeto consistiu na correção do tipo de dados utilizado para armazenamento de informações temporais. Conforme discutido no relatório de inconsistências, os campos `ocorrencia_dia` e `divulgacao_dia_publicacao` encontravam-se definidos como `VARCHAR`, impossibilitando a utilização de funções nativas de manipulação temporal do MySQL.

Adicionalmente, identificou-se que a base continha datas em formatos mistos: parte dos registros utilizava o padrão internacional `yyyy-mm-dd` (2021-08-12), enquanto outros empregavam o formato brasileiro `dd/mm/yyyy` (07/01/2007). Esta inconsistência evidencia processos de migração ou consolidação de dados provenientes de sistemas distintos, configurando desafio adicional para a conversão.

A solução implementada consistiu na criação de uma coluna temporária do tipo `DATE`, seguida de conversões específicas para cada formato identificado:

```
ALTER TABLE tabela_ocorrencia
```

```
ADD COLUMN ocorrencia_dia_temporario DATE;
```

```
UPDATE tabela_ocorrencia  
SET ocorrencia_dia_temporario = STR_TO_DATE(ocorrencia_dia, '%Y-%m-%d')  
WHERE ocorrencia_dia REGEXP '^[0-9]{4}-[0-9]{2}-[0-9]{2}$';
```

```
UPDATE tabela_ocorrencia  
SET ocorrencia_dia_temporario = STR_TO_DATE(ocorrencia_dia, '%d/%m/%Y')  
WHERE ocorrencia_dia REGEXP '^[0-9]{2}/[0-9]{2}/[0-9]{4}$'  
AND ocorrencia_dia_temporario IS NULL;
```

A estratégia empregada utiliza expressões regulares (REGEXP) para identificar o formato de cada registro antes de aplicar a conversão apropriada. O primeiro UPDATE processa datas no formato internacional, enquanto o segundo trata exclusivamente registros ainda não convertidos que correspondam ao padrão brasileiro. A cláusula AND ocorrencia_dia_temporario IS NULL no segundo UPDATE previne sobreposições, assegurando que registros já processados não sejam reavaliados.

A função STR_TO_DATE() realiza a conversão efetiva, interpretando a string de acordo com o formato especificado e gerando um valor do tipo DATE que o MySQL armazena internamente no padrão yyyy-mm-dd. Este tipo nativo permite operações como cálculo de intervalos (DATEDIFF), extração de componentes (YEAR, MONTH, DAY), comparações temporais e agregações por períodos — funcionalidades essenciais para análises estatísticas e investigativas sobre os dados aeronáuticos.

Após validação da conversão e confirmação da ausência de perdas significativas de dados, a coluna original poderia ser removida e a temporária renomeada. Entretanto, mantendo coerência com a diretriz de preservação da estrutura original, optou-se por manter ambas as colunas, permitindo rastreabilidade e comparações futuras entre os formatos.

3.5 RESULTADOS E VALIDAÇÕES

As modificações estruturais implementadas foram validadas através de comandos de inspeção da estrutura das tabelas:

```
DESCRIBE tabela_ocorrencia;  
DESCRIBE tabela_tipo_ocorrencia;  
DESCRIBE tabela_aeronave;  
DESCRIBE tabela_fator_contribuinte;  
DESCRIBE tabela_recomendacao_seguranca;
```

Estes comandos permitiram confirmar que as alterações foram aplicadas corretamente, que os tipos de dados encontram-se adequadamente definidos e que as constraints de chave primária, unicidade e obrigatoriedade estão ativas. A estrutura resultante, embora não ideal devido às limitações impostas pelos dados originais, representa melhoria substancial em relação à configuração inicial, estabelecendo base sólida para consultas, análises e futuras expansões do sistema.

A desativação de `FOREIGN_KEY_CHECKS`, embora necessária neste contexto específico, deve ser compreendida como medida excepcional. Em cenários profissionais onde seja possível intervir na qualidade dos dados de origem, a recomendação técnica consiste em corrigir as inconsistências antes da definição de constraints, assegurando integridade referencial plena desde a concepção do modelo.

4 LINK DO REPOSITÓRIO

<https://github.com/danieloliveira33/Banco-de-Dados.git>