

EEL890 Big Data 2024.1

Tarefa de Avaliação Parte II

Carlos Henrique Ferreira Brito Filho (DRE 120081409)
Bárbara Rodrigues dos Santos Cerqueira (DRE 117198425)
Ramon de Attayde Barros de Souza (DRE 122047728)
Barbara Varela Bonfim (DRE 120130698)

1 Introdução

Este trabalho consistiu em desenvolver uma solução de Data Warehouse (DWH) para a fusão de quatro empresas de aluguel de automóveis, assim como projetar o sistema operacional de uma das empresas, gerar os relatórios gerenciais e realizar algumas análises de dados.

As seções a seguir irão aprofundar sobre o processo de desenvolvimento do projeto e as decisões tomadas ao longo do trabalho.

2 Contexto do Trabalho

Quatro empresas de aluguel de automóveis decidiram se fundir, mantendo suas operações em quatro locais distintos: Aeroporto do Galeão, Santos Dumont, Shopping Rio Sul e Barra Shopping. Cada uma possui sistemas transacionais próprios para gerenciar clientes, frota de veículos, reservas e locações, além de sistemas de RH, controle de pátio, compras e fornecedores.

As empresas continuarão usando seus sistemas existentes, permitindo que os veículos possam ser retirados e devolvidos em qualquer dos quatro pátios. Elas resolveram implementar uma solução de Data Warehouse (DWH) para integrar dados históricos e gerar relatórios gerenciais e análises de dados unificados.

O projeto focou na integração dos dados do subsistema de controle de reservas e locações de veículos, além do desenvolvimento do banco de dados transacional de uma das empresas. O objetivo era criar um DWH que permitisse relatórios sobre controle de pátio, locações, reservas, tipos de veículos mais alugados e análise de movimentação de frota entre os pátios usando a matriz estocástica de Markov.

3 Desenvolvimento do Banco de Dados Transacional

O esquema do banco de dados transacional projetado para uma das empresas de aluguel de automóveis foi projetado para suportar as operações de reserva, locação, e controle de frota, e resultou no desenvolvimento do Modelo Entidade-Relacionamento encontrado na Figura 1.

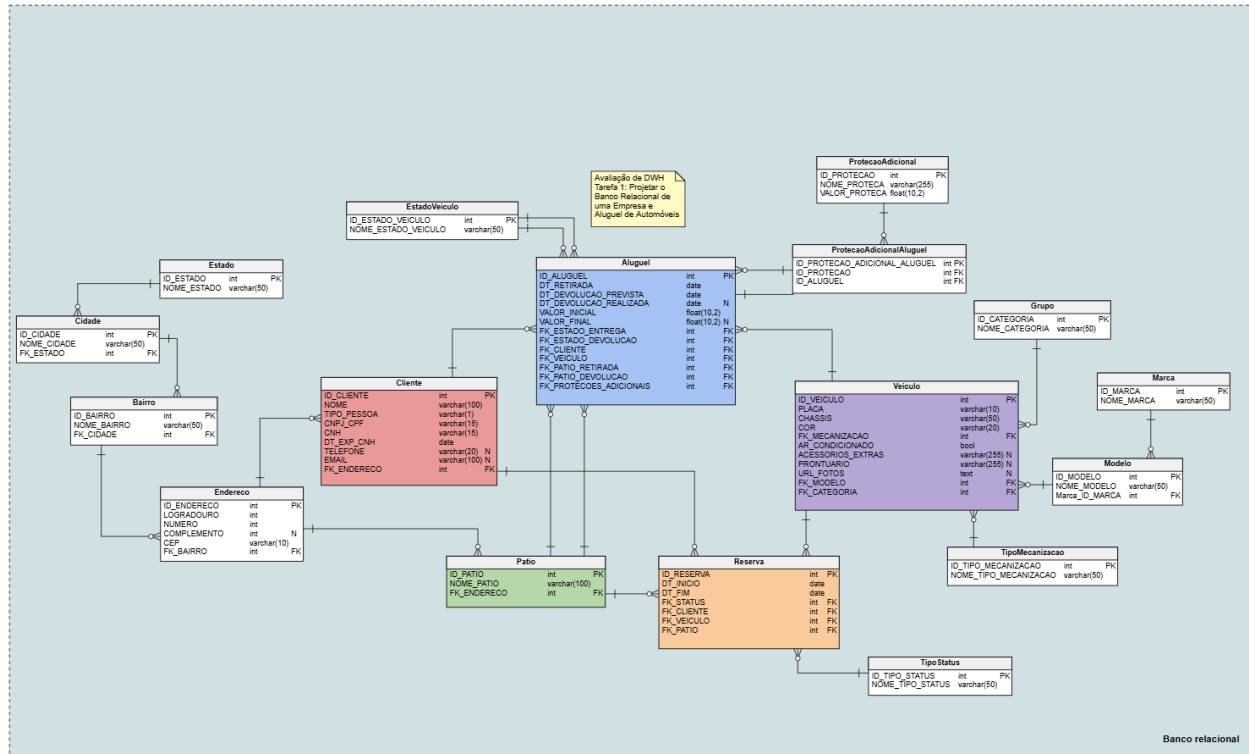


Figura 1: Modelo ER do Banco de Dados Transacional

Foi usado um modelo relacional normalizado que atendesse às expectativas do negócio, além dos relatórios gerenciais e análises de movimentação de veículos entre pátios pretendidos no projeto do Data Warehouse posteriormente.

3.1 Dicionário de Dados

O desenvolvimento do banco de dados foi organizado em torno do que foi considerado as principais entidades do sistema, que seriam Cliente, Veículo, Pátio, Reserva e Aluguel, e posteriormente foi realizada a normalização em cima delas. O script para a criação das tabelas se encontra no GitHub neste link. A seguir está uma breve descrição das tabelas e seus campos.

Tabela: Cliente

Campo	Descrição	Tipo	Restrição
ID_CLIENTE	Identificador do cliente	int	PK, NOT NULL
NOME	Nome do cliente	varchar(100)	NOT NULL
TIPO_PESSOA	Tipo de pessoa (F ou J)	varchar(1)	NOT NULL
CNPJ_CPF	Documento (CNPJ ou CPF)	varchar(15)	NOT NULL
CNH	Número da CNH	varchar(15)	NOT NULL
DT_EXP_CNH	Data de expedição da CNH	date	NOT NULL
TELEFONE	Telefone do cliente	varchar(20)	NULL
EMAIL	Email do cliente	varchar(100)	NULL
FK_ENDERECO	Endereço do cliente	int	NOT NULL, FK

Tabela: Aluguel

Campo	Descrição	Tipo	Restrição
ID_ALUGUEL	Identificador do aluguel	int	PK, NOT NULL
DT_RETIRADA	Data de retirada do veículo	date	NOT NULL
DT_DEVOLUCAO_PREVISTA	Data prevista de devolução do veículo	date	NOT NULL
DT_DEVOLUCAO_REALIZADA	Data real de devolução do veículo	date	NULL
VALOR_INICIAL	Valor inicial do aluguel	float(10,2)	NOT NULL
VALOR_FINAL	Valor final do aluguel	float(10,2)	NULL
FK_ESTADO_ENTREGA	Estado do veículo na entrega	int	NOT NULL, FK
FK_ESTADO_DEVOLUCAO	Estado do veículo na devolução	int	NOT NULL, FK
FK_CLIENTE	Cliente que alugou o veículo	int	NOT NULL, FK
FK_VEICULO	Veículo alugado	int	NOT NULL, FK
FK_PATIO_RETIRADA	Pátio de retirada do veículo	int	NOT NULL, FK
FK_PATIO_DEVOLUCAO	Pátio de devolução do veículo	int	NOT NULL, FK
FK_PROTECOES_ADICIONAIS	Proteções adicionais contratadas	int	NOT NULL, FK

Tabela: Bairro

Campo	Descrição	Tipo	Restrição
ID_BAIRRO	Identificador do bairro	int	PK, NOT NULL
NOME_BAIRRO	Nome do bairro	varchar(50)	NOT NULL
FK_CIDADE	Cidade do bairro	int	NOT NULL, FK

Tabela: Cidade

Campo	Descrição	Tipo	Restrição
ID_CIDADE	Identificador da cidade	int	PK, NOT NULL
NOME_CIDADE	Nome da cidade	varchar(50)	NOT NULL
FK_ESTADO	Estado da cidade	int	NOT NULL, FK

Tabela: Endereco

Campo	Descrição	Tipo	Restrição
ID_ENDERECO	Identificador do endereço	int	PK, NOT NULL
LOGRADOURO	Logradouro do endereço	int	NOT NULL
NUMERO	Número do endereço	int	NOT NULL
COMPLEMENTO	Complemento do endereço	int	NULL
CEP	Código postal do endereço	varchar(10)	NOT NULL
FK_BAIRRO	Bairro do endereço	int	NOT NULL, FK

Tabela: Estado

Campo	Descrição	Tipo	Restrição
ID_ESTADO	Identificador do estado	int	PK, NOT NULL
NOME_ESTADO	Nome do estado	varchar(50)	NOT NULL

Tabela: EstadoVeiculo

Campo	Descrição	Tipo	Restrição
ID_ESTADO_VEICULO	Identificador do estado do veículo	int	PK, NOT NULL
NOME_ESTADO_VEICULO	Nome do estado do veículo	varchar(50)	NOT NULL

Tabela: Grupo

Campo	Descrição	Tipo	Restrição
ID_CATEGORIA	Identificador da categoria	int	PK, NOT NULL
NOME_CATEGORIA	Nome da categoria	varchar(50)	NOT NULL

Tabela: Marca

Campo	Descrição	Tipo	Restrição
ID_MARCA	Identificador da marca	int	PK, NOT NULL
NOME_MARCA	Nome da marca	varchar(50)	NOT NULL

Tabela: Modelo

Campo	Descrição	Tipo	Restrição
ID_MODELO	Identificador do modelo	int	PK, NOT NULL
NOME_MODELO	Nome do modelo	varchar(50)	NOT NULL
Marca_ID_MARCA	Identificador da marca	int	NOT NULL, FK

Tabela: Patio

Campo	Descrição	Tipo	Restrição
ID_PATIO	Identificador do pátio	int	PK, NOT NULL
NOME_PATIO	Nome do pátio	varchar(100)	NOT NULL
FK_ENDERECO	Identificador do endereço	int	NOT NULL, FK

Tabela: ProtecaoAdicional

Campo	Descrição	Tipo	Restrição
ID_PROTECAO	Identificador da proteção	int	PK, NOT NULL
NOME_PROTECAO	Nome da proteção	varchar(255)	NOT NULL
VALOR_PROTECAO	Valor da proteção	float(10,2)	NOT NULL

Tabela: ProtecaoAdicionalAluguel

Campo	Descrição	Tipo	Restrição
ID_PROTECAO_ADICIONAL_ALUGUEL	Identificador da proteção adicional do aluguel	int	PK, NOT NULL
ID_PROTECAO	Identificador da proteção	int	NOT NULL, FK
ID_ALUGUEL	Identificador do aluguel	int	NOT NULL, FK

Tabela: Reserva

Campo	Descrição	Tipo	Restrição
ID_RESERVA	Identificador da reserva	int	PK, NOT NULL
DT_INICIO	Data de início da reserva	date	NOT NULL
DT_FIM	Data de fim da reserva	date	NOT NULL
FK_STATUS	Status da reserva	int	NOT NULL, FK
FK_CLIENTE	Identificador do cliente	int	NOT NULL, FK
FK_VEICULO	Identificador do veículo	int	NOT NULL, FK
FK_PATIO	Identificador do pátio	int	NOT NULL, FK

Tabela: TipoMecanizacao

Campo	Descrição	Tipo	Restrição
ID_TIPO_MECANIZACAO	Identificador do tipo de mecanização	int	PK, NOT NULL
NOME_TIPO_MECANIZACAO	Nome do tipo de mecanização	varchar(50)	NOT NULL

Tabela: TipoStatus

Campo	Descrição	Tipo	Restrição
ID_TIPO_STATUS	Identificador do tipo de status	int	PK, NOT NULL
NOME_TIPO_STATUS	Nome do tipo de status	varchar(50)	NOT NULL

Tabela: Veiculo

Campo	Descrição	Tipo	Restrição
ID_VEICULO	Identificador do veículo	int	PK, NOT NULL
PLACA	Placa do veículo	varchar(10)	NOT NULL
CHASSIS	Chassis do veículo	varchar(50)	NOT NULL
COR	Cor do veículo	varchar(20)	NOT NULL
FK_MECANIZACAO	Identificador do tipo de mecanização	int	NOT NULL, FK
AR_CONDICIONADO	Indica se o veículo possui ar condicionado	bool	NOT NULL
ACESSORIOS_EXTRAS	Acessórios extras do veículo	varchar(255)	NULL
PRONTUARIO	Prontuário do veículo	varchar(255)	NULL
URL_FOTOS	URLs das fotos do veículo	text	NULL
FK_MODELO	Identificador do modelo do veículo	int	NOT NULL, FK
FK_CATEGORIA	Identificador da categoria do veículo	int	NOT NULL, FK

4 Desenvolvimento do Modelo Dimensional

O modelo dimensional para o Data Warehouse (DWH) foi desenvolvido para integrar os dados históricos das quatro empresas que se fundiram. Este modelo visa facilitar a geração de relatórios gerenciais e a realização de análises de dados unificados. As tabelas dimensionais e de fatos foram projetadas para suportar as principais necessidades de relatórios e análises descritas no contexto do trabalho.

O Modelo Entidade-Relacionamento desenvolvido para o Data Warehouse pode ser encontrado na Figura 2.

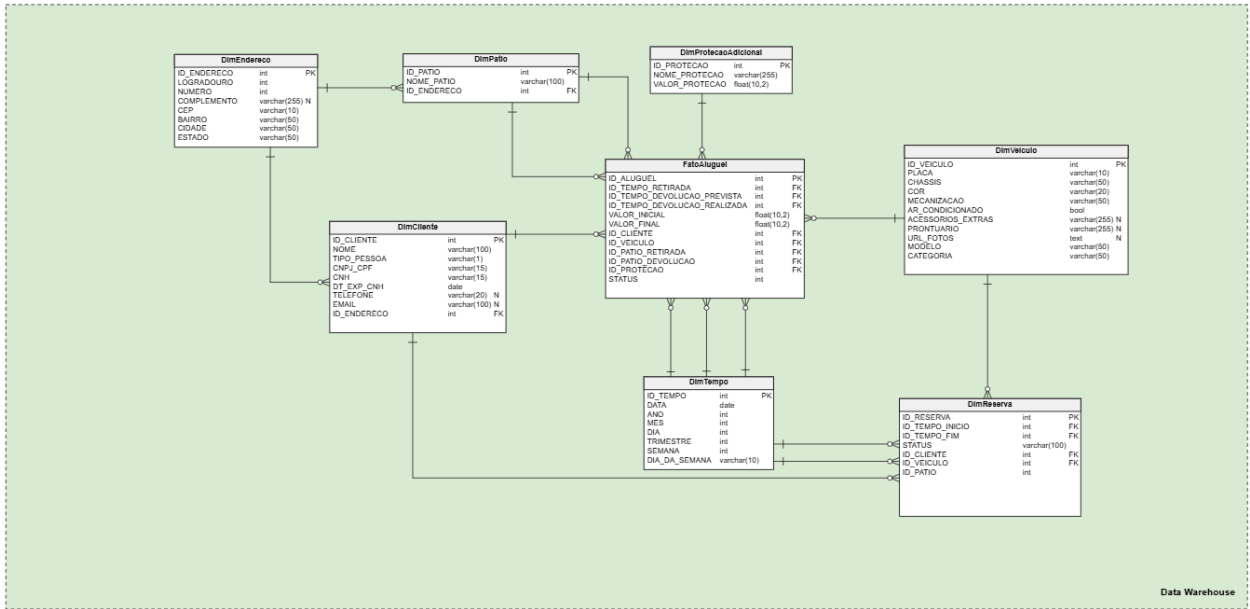


Figura 2: Modelo ER do Data Warehouse

4.1 Dicionário de Dados

A seguir está uma descrição detalhada das tabelas e colunas do modelo.

Tabela: FatoAluguel

Campo	Descrição	Tipo	Restrição
ID_ALUGUEL	Identificador do aluguel	int	PK, NOT NULL
ID_TEMPO_RETIRADA	Identificador do tempo de retirada	int	NOT NULL, FK
ID_TEMPO_DEVOLUCAO_PREVISTA	Identificador do tempo de devolução prevista	int	NOT NULL, FK
ID_TEMPO_DEVOLUCAO_REALIZADA	Identificador do tempo de devolução realizada	int	NOT NULL, FK
VALOR_INICIAL	Valor inicial do aluguel	float(10,2)	NOT NULL
VALOR_FINAL	Valor final do aluguel	float(10,2)	NOT NULL
ID_CLIENTE	Identificador do cliente	int	NOT NULL, FK
ID_VEICULO	Identificador do veículo	int	NOT NULL, FK
ID_PATIO_RETIRADA	Identificador do pátio de retirada	int	NOT NULL, FK
ID_PATIO_DEVOLUCAO	Identificador do pátio de devolução	int	NOT NULL, FK
ID_PROTECAO	Identificador da proteção adicional	int	NOT NULL, FK
STATUS	Status do aluguel	int	NOT NULL

Tabela: DimCliente

Campo	Descrição	Tipo	Restrição
ID_CLIENTE	Identificador do cliente	int	PK, NOT NULL
NOME	Nome do cliente	varchar(100)	NOT NULL
TIPO_PESSOA	Tipo de pessoa (F ou J)	varchar(1)	NOT NULL
CNPJ_CPF	Documento do cliente	varchar(15)	NOT NULL
CNH	CNH do cliente	varchar(15)	NOT NULL
DT_EXP_CNH	Data de expedição da CNH	date	NOT NULL
TELEFONE	Telefone do cliente	varchar(20)	
EMAIL	Email do cliente	varchar(100)	
ID_ENDERECO	Identificador do endereço	int	NOT NULL, FK

Tabela: DimEndereco

Campo	Descrição	Tipo	Restrição
ID_ENDERECO	Identificador do endereço	int	PK, NOT NULL
LOGRADOURO	Logradouro do endereço	int	NOT NULL
NUMERO	Número do endereço	int	NOT NULL
COMPLEMENTO	Complemento do endereço	varchar(255)	
CEP	CEP do endereço	varchar(10)	NOT NULL
BAIRRO	Bairro do endereço	varchar(50)	NOT NULL
CIDADE	Cidade do endereço	varchar(50)	NOT NULL
ESTADO	Estado do endereço	varchar(50)	NOT NULL

Tabela: DimPatio

Campo	Descrição	Tipo	Restrição
ID_PATIO	Identificador do pátio	int	PK, NOT NULL
NOME_PATIO	Nome do pátio	varchar(100)	NOT NULL
ID_ENDERECO	Identificador do endereço do pátio	int	NOT NULL, FK

Tabela: DimProtecaoAdicional

Campo	Descrição	Tipo	Restrição
ID_PROTECAO	Identificador da proteção adicional	int	PK, NOT NULL
NOME_PROTECAO	Nome da proteção adicional	varchar(255)	NOT NULL
VALOR_PROTECAO	Valor da proteção adicional	float(10,2)	NOT NULL

Tabela: DimReserva

Campo	Descrição	Tipo	Restrição
ID_RESERVA	Identificador da reserva	int	PK, NOT NULL
ID_TEMPO_INICIO	Identificador do tempo de início da reserva	int	NOT NULL, FK
ID_TEMPO_FIM	Identificador do tempo de fim da reserva	int	NOT NULL, FK
STATUS	Status da reserva	varchar(100)	NOT NULL
ID_CLIENTE	Identificador do cliente	int	NOT NULL, FK
ID_VEICULO	Identificador do veículo	int	NOT NULL, FK
ID_PATIO	Identificador do pátio	int	NOT NULL, FK

Tabela: DimTempo

Campo	Descrição	Tipo	Restrição
ID_TEMPO	Identificador do tempo	int	PK, NOT NULL
DATA	Data	date	NOT NULL
ANO	Ano	int	NOT NULL
MES	Mês	int	NOT NULL
DIA	Dia	int	NOT NULL
TRIMESTRE	Trimestre	int	NOT NULL
SEMANA	Semana	int	NOT NULL
DIA_DA_SEMANA	Dia da semana	varchar(10)	NOT NULL

Tabela: DimVeiculo

Campo	Descrição	Tipo	Restrição
ID_VEICULO	Identificador do veículo	int	PK, NOT NULL
PLACA	Placa do veículo	varchar(10)	NOT NULL
CHASSIS	Chassis do veículo	varchar(50)	NOT NULL
COR	Cor do veículo	varchar(20)	NOT NULL
MECANIZACAO	Mecanização do veículo	varchar(50)	NOT NULL
AR_CONDICIONADO	Indica se tem ar condicionado	bool	NOT NULL
ACESSORIOS_EXTRAS	Acessórios extras do veículo	varchar(255)	
PRONTUARIO	Prontuário do veículo	varchar(255)	
URL_FOTOS	URL das fotos do veículo	text	
MODELO	Modelo do veículo	varchar(50)	NOT NULL
CATEGORIA	Categoria do veículo	varchar(50)	NOT NULL

5 Desafios Encontrados

A execução deste projeto apresentou diversos desafios significativos. Primeiramente, a construção do modelo relacional e, principalmente, do modelo dimensional (novo para muitos de nós) foi bastante complicada devido à nossa inexperiência. Tivemos algumas dificuldades em entender e implementar os conceitos teóricos de modelagem de dados na prática, o que resultou em várias iterações e revisões dos modelos até chegarmos a uma versão satisfatória.

A maior dificuldade foi sem dúvidas a integração dos dados de outros alunos, referentes as outras três empresas. Cada modelo relacional tinha estruturas de dados e padrões bastante distintos. A harmonização desses dados para construir um Data Warehouse (DWH) único e coeso foi um processo extremamente desafiador. Os modelos de dados eram muito diferentes do modelo criado por nós, o que resultou em muitos dados não conseguindo ser adequadamente integrados. Um exemplo bastante problemático foi a informação do Endereço, que nos outros modelos foi implementado como uma única coluna em formato de string, mas no nosso foi quase totalmente normalizado; não soubemos como adequar um ao outro. Isso

gerou um alto número de valores nulos nas tabelas finais, prejudicando a completude e a qualidade dos dados disponíveis para análise.

6 Conclusão

Em suma, foi possível alcançar os objetivos principais do projeto de DWH. A construção do modelo dimensional permitiu a consolidação dos dados históricos das quatro empresas, simplificando a geração de relatórios gerenciais e análises de dados unificadas com os scripts SQL. O esquema dimensional implementado permite consultas analíticas com mais facilidade, permitindo insights sobre o desempenho e a operação das empresas fundidas.

No entanto, o problema com a dificuldade de integração dos dados provavelmente abalaria consideravelmente a precisão do modelo na prática, devido à alta quantidade de valores nulos ou ausentes.