



CENTRO UNIVERSITÁRIO UNIMETROCAMP WYDEN -PR
Graduate Program in Data Science and Analytical Intelligence

Regilene Mariano
Data Warehouse: arquitetura, requisitos e fontes de dados

RESUMO

Este trabalho apresenta as etapas para a Modelagem de dados para Data Warehouse, estuda sua arquitetura, os levantamentos de requisitos e a modelagem física do Banco de Dados. A partir do enunciado proposto, procuramos compreender os desafios e as metas da empresa com o objetivo de definir as métricas e as dimensões necessárias para análise dos dados. Antes de implementar o Data Warehouse precisamos modelar as tabelas transacionais com base nos dados apresentados e realizar a população dessas tabelas relacionais. A partir disso, elaboramos a definição das tabelas de fatos e dimensões necessárias, bem como as métricas e análises essenciais para monitorar o desempenho da empresa. Ao final propusemos consultas em SQL para extração das métricas. A implementação deste DW ajudará a organização a atingir seus objetivos estratégicos, além de fornecer insights para a tomada de decisões.

Table contents

1. Introdução	5
2. Objetivos	7
2.1 Objetivo Geral	7
2.2 Objetivo Específico	7
3. Justificativa	8
4. Metodologia	8
5. Estudo de caso	8
6. Modelo Físico de Tabelas Relacionais	11
5.1 Modelo de dados Físico -Sistema de Concessão de crédito	11
7. Implementação de Banco de Dados Relacional	13
8. Populando Tabelas Transacionais	14
9. Projeto de banco de Dados para Data Warehouse	15
10. O que é um Data Warehouse?	15
11. Arquitetura de um Data Warehouse	15
10.1 Entendimento do Negócio	15
10.2 Mapeamento de dados	15
10.3 <i>Staging area</i> de dados	16
10.4 Processo de ETL	16
10.5 Análise	16
12. Tabelas Dimensionais, Requisitos de Análise e Métricas	17
13. Matriz de Granularidade	18
14. Tabelas Dimensionais	19
15. Tabela Fato	21
16. Tabela Fonte de dados	22
17. Implementação do Banco de Dados -Data Warehouse	23
18. <i>Starchema</i> - Diagrama Entidade-Relacionamento	24
19. Consultas Analíticas	25
20. Visualização de Dados	26
21. Estrutura do Banco de Dados	27
22. Processo de ETL	27
23. Conclusão	28
Referências	29

Table List

1. Tabela 1: cliente	11
2. Tabela 2: produto	11
3. Tabela 3: endereco	11
4. Tabela 4: pagamento	12
5. Tabela 5: contrato	12
6. Tabela 6: fonte de renda	13
7. Tabela 7: faixa de renda	13
8. Tabela 8: Visão cliente	17
9. Tabela 9: Visão produto	17
10. Tabela 10: Visão pagamento	18
11. Tabela 11: DimCliente	19
12. Tabela 12: DimTempo	20
13. Tabela 13: DimFaixaRenda	20
14. Tabela 14: DimFonteRenda	20
15. Tabela 15: DimPerfilDemografico table	21
16. Tabela 16: Fato FatoTransacaoFinanceira	21

Images List

Figura 1: Quadro de objetivos	6
Figura 2: Diagrama de Processo geral	10
Figura 3: DDL -Tabela Relacional cliente	13
Figura 4: DML - Inserindo dados	14
Figura 5: Matriz de Granularidade	18
Figura 6: Fonte de dados	21
Figura 7: DDL - DimCliente	22
Figura 8: DDL - DimFonteRenda	22
Figura 9: DDL - FaixaRenda	23
Figura 10: DDL - DimTempo	23
Figura 11: DDL - DimPerfilDemografico	23
Figura 12: FatoTransacao	23
Figura 13: DER - Starschema	24
Figura 14: Consulta SQL à tabela Fato	25
Figura 15: Relatório de consulta Fato Resumo de Pagamento	26
Figura 16: Gráfico: Maior alta de inadimplência por mês	27

1.Introdução

Uma série de documentos é necessária antes de iniciar a elaboração de um projeto de Data Warehouse. É preciso compreender o ambiente organizacional, suas regras de negócio, seus processos, e, principalmente, seus desafios e objetivos como organização. Para isso é essencial o levantamento das informações através dos usuários envolvidos nos processos da empresa. Pois, quando pensamos em criar um Modelo de Dados para Data Warehouse temos que ter em mente que o objetivo final da tarefa é criar uma ferramenta estratégica de negócios. É a partir do Data Warehouse que os usuários finais contarão com relatórios, painéis e análises para extrair insights de dados, monitorar o desempenho dos negócios e apoiar a tomada de decisões.

Realizados os levantamentos e a documentação necessária desses levantamentos, seguimos para o ativo mais importante: os dados. Através de processos e técnicas bem definidos, os dados de uma organização podem se transformar em conhecimento e embasar a melhoria de atividades, o crescimento e minimizar riscos, colaborando para a tomada de decisões mais assertivas. Esses dados podem vir das mais diversas fontes, no exemplo estudado, eles fazem parte da rotina diária da empresa. Segundo Silberschatz, Korth e Sudarshan (1999), um banco de dados é uma coleção de dados que contém informações relevantes referentes à organização.

A partir do enunciado elaboramos as tabelas relacionais da empresa 'CrediSimples', inferindo a rotina de operações de uma financeira. Para isso, foi necessário estudar os processos que envolvem o pedido e a concessão de empréstimo como a solicitação, a coleta de dados preliminares, a análise de dados, a aprovação do crédito, a liberação dos fundos e o monitoramento de pagamentos. O levantamento dessas informações se deu por meio de pesquisa sobre os processos de Financeiras de Crédito. Criamos, então, o Modelo conceitual, o Modelo Lógico, o Modelo físico, realizamos a implementação do mesmo e sua população com dados fictícios. O Banco de dados OLTP servirá de base para a criação do Data Warehouse.

Após esse processo, iniciamos a Modelagem de dados para Data Warehouse. O design segue o mesmo padrão da Modelagem Relacional, do conceitual ao físico e ao Diagrama Entidade-Relacionamento. Diferente de um modelo de dados transacional onde os modelos são focados em operações diárias e consistência dos dados, no Data Warehouse os modelos dimensionais são voltados para a análise e a tomada de decisões rápidas, então a principal diferença entre eles está na finalidade. Com a definição dos desafios da empresa em mãos, inferimos algumas métricas e análises capazes de apoiar decisões estratégicas e orientar a criação dos modelos dimensionais.

1.1 Problema

A partir do enunciado ficou claro quais eram os principais desafios e os objetivos estratégicos atuais da empresa. Conforme quadro abaixo:

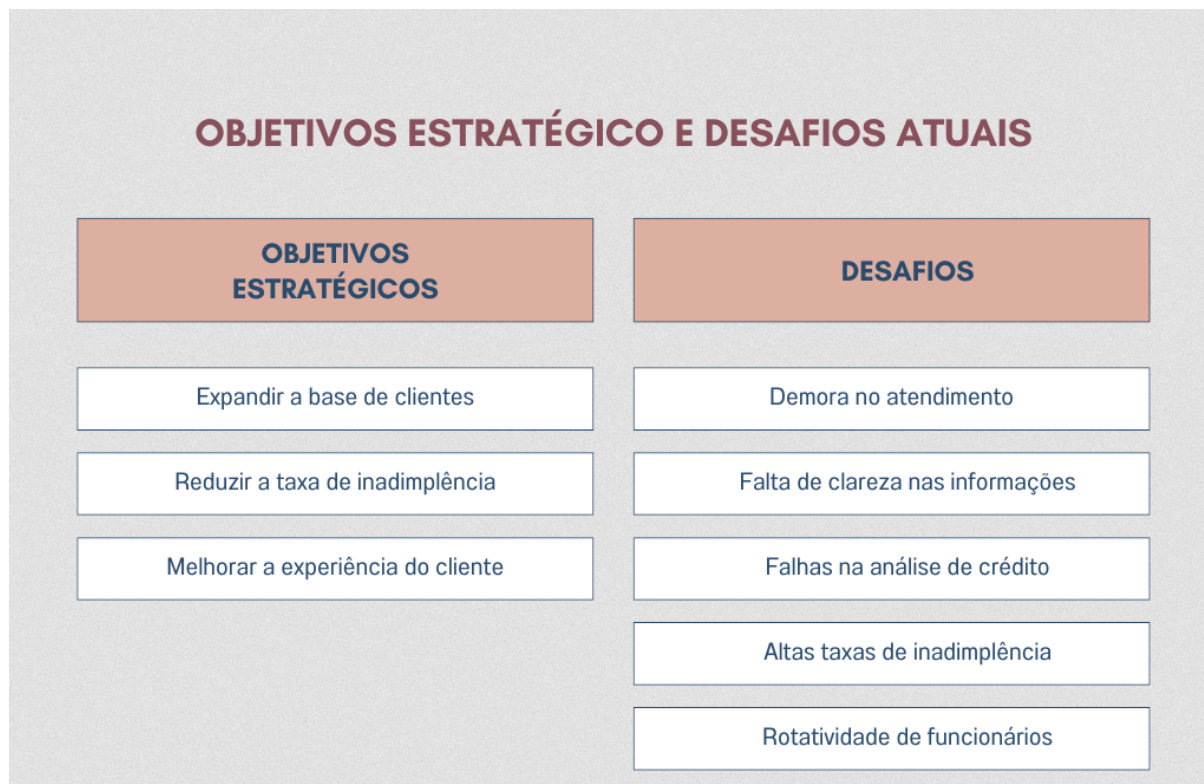


Figura 1: Quadro de objetivos e estratégias

Através da elaboração do quadro de objetivos e desafios um novo ponto pode ser observado: os processos que impactam diretamente a experiência do cliente, a qualidade da análise de crédito e a eficiência operacional são altamente relevantes para o alcance dos objetivos estratégicos da empresa e devem ser priorizados num plano de inovação de processos.

Os objetivos atuais da empresa são: expandir a base de clientes, melhorar a experiência do cliente e reduzir as taxas de inadimplência. Inferimos que não é de conhecimento da empresa dados relevantes sobre a Taxa de inadimplência de seus clientes, produtos e serviços. Esse é um tópico crítico, pois afeta diretamente o fluxo de caixa da empresa e sem um fluxo de caixa adequado e organização poderá enfrentar sérias dificuldades para arcar com suas obrigações.

A elaboração de indicadores poderá contribuir para o monitoramento da taxa de inadimplência por produto, região e até por perfil de cliente. A análise desses indicadores pode ajudar a identificar tendências de inadimplência e permitir ações preventivas por parte

da empresa. A análise das taxas de inadimplência é fundamental para a saúde financeira da CrediSimples e é um dos tópicos mais importantes para o Business Intelligence. Logo, a construção do Data Warehouse é fundamental para lidar com essa questão crítica.

2. Objetivos

2.1 Objetivos geral

Projetar o Data Warehouse para dar suporte à tomada de decisões da empresa CrediSimples.

2.2 Objetivos específicos

- Elaborar o Banco de dados transacional para a organização.
- Implementar o Banco de dados OLTP e realizar sua população com dados fictícios.
- Criar e implementar o Data Warehouse com base na fonte de dados.
- Apresentar consultas para obter insights de tópicos críticos para a empresa.

3 Justificativa

Não é mais novidade que os dados são o novo petróleo (The Economist, 2017), assim como o petróleo se torna valioso depois de ser extraído e refinado, os dados brutos quando coletados e analisados corretamente, podem gerar informações valiosas para uma organização. Atualmente os dados são o motor da economia digital, eles impulsionam inovações na área de inteligência artificial, aprendizado de máquina, big data, internet das coisas (IoT), e-commerce, marketing digital, etc. Para as empresas, uma cultura orientada a dados pode influenciar a tomada de decisões, ajudar a compreender melhor o perfil dos seus clientes e o mercado em que atua. Pode ainda, através dos dados históricos e de tempo real, prever tendências futuras, o que facilitaria o planejamento e a preparação para desafios e oportunidades.

Para uma financeira de crédito, como a CrediSimples, os dados que revelam o comportamento dos usuários de seus produtos, o seu perfil em relação às taxas de inadimplência, encontram oportunidade de otimização de seus processos, tem uma enorme importância. Por isso, a construção do Data Warehouse que oportuniza essa análise e insights é essencial.

4. Metodologia

A metodologia se refere aos procedimentos e as ferramentas utilizadas nesse trabalho. A metodologia compreende o método de pesquisa adotado e as ferramentas às ferramentas tecnológicas utilizadas.

4.1 Ferramentas

- a) Python como linguagem de programação;
- b) Oracle como Sistema de Gerenciamento de Banco de Dados;
- c) Oracle cliente para elaborar o Modelo Entidade-Relacionamento;
- d) Miro app para apresentação visual de processos;
- e) Power BI para visualização dos dados das queries em SQL;

4.2 Metodologia

A metodologia utilizada nessa tarefa foi a Pesquisa Exploratória. A pesquisa exploratória nos ajudou a compreender o contexto de mercado em que as Financeiras de créditos atuam e os problemas iniciais do enunciado. Após a definição dos problemas, consideramos o método mais adequado para obter informações preliminares. Além do próprio enunciado, estudamos relatórios e documentos das principais instituições bancárias do país, também exploramos os documentos que apresentam o Mapa da Inadimplência (Serasa,2024) no Brasil e como ela se distribui por região, faixa etária e fonte de renda. Também fez parte do método da pesquisa, o estudo sobre cálculo de inadimplência (CobreFácil,2024), e a leitura de documentos sobre como funcionam a análise de crédito, os tipos de produtos de crédito e suas condições em bancos como Santander e Banco do Brasil. Essa etapa foi essencial para compreender as Regras de Negócio no *Bureau* de Crédito. Realizamos um estudo sobre Perfil de Cliente (SERASA,2019), o que nos permitiu entender melhor as necessidades, os comportamentos e desafios dos clientes de uma financeira.

5. Estudo de caso

A CrediSimples é uma empresa que oferece empréstimos populares através de contratos bancários. Presente nos estados do Paraná e Santa Catarina, na Região Sul do Brasil, a CrediSimples tem passado por uma reestruturação operacional e organizacional nos últimos anos. Durante esse processo, a empresa identificou pontos que precisam ser melhorados para atingir seus objetivos estratégicos e continuar cumprindo sua missão. Além das altas taxas de inadimplência e da rotatividade de funcionários, a CrediSimples enfrenta outros desafios comuns às financeiras de crédito popular. Esses desafios incluem demora no atendimento ao cliente, falta de clareza nas informações sobre produtos e serviços, e perdas

financeiras devido a falhas na análise de crédito, resultando na concessão de empréstimos a clientes com alto risco de inadimplência.

A visão e missão da CrediSimples são de tornar a principal fornecedora de soluções financeiras acessíveis e inovadoras para clientes de baixa e média renda na região sul do Brasil. Como parte de seu planejamento estratégico, a empresa pretende expandir sua base de clientes, reduzir a taxa de inadimplência e melhorar a experiência do cliente.

Atualmente a empresa conta com um setor de atendimento presencial, um setor para Análise de Crédito, setor de Aprovação e Contratação e Setor de Cobrança, onde funcionários do atendimento são destacados para realizar suporte ao cliente(solicitação de serviços, dúvidas, reclamações e feedbacks) e monitorar os pagamentos. O Setor de Gestão Interna e Suporte Operacional, engloba processo de Gestão em TI, Recursos Humanos, e Gestão Financeira. O setor de Gestão Estratégica, é responsável pelo Planejamento Estratégico com definição de objetivos de longo prazo, análise de mercado e competitividade e desenvolvimento de novas estratégias de negócio.

Para alcançar os objetivos, o escritório de Gestão Estratégica da empresa foi encarregado de elaborar um plano de inovação para estabelecer um novo modelo de negócio conforme definido pela alta administração.

5.1 Processos da empresa

Segundo Davenport(1994) a definição de um processo envolve as atividades estruturadas com a finalidade de agregar valor resultando em um produto para os clientes. São atividades de trabalho com ordenação específica no tempo e no espaço, com começo e fim, com input e outputs bem definidos.

5.2 Processo geral

O processo geral da empresa ocorre da maneira como é explicitado pelo diagrama de processos abaixo:

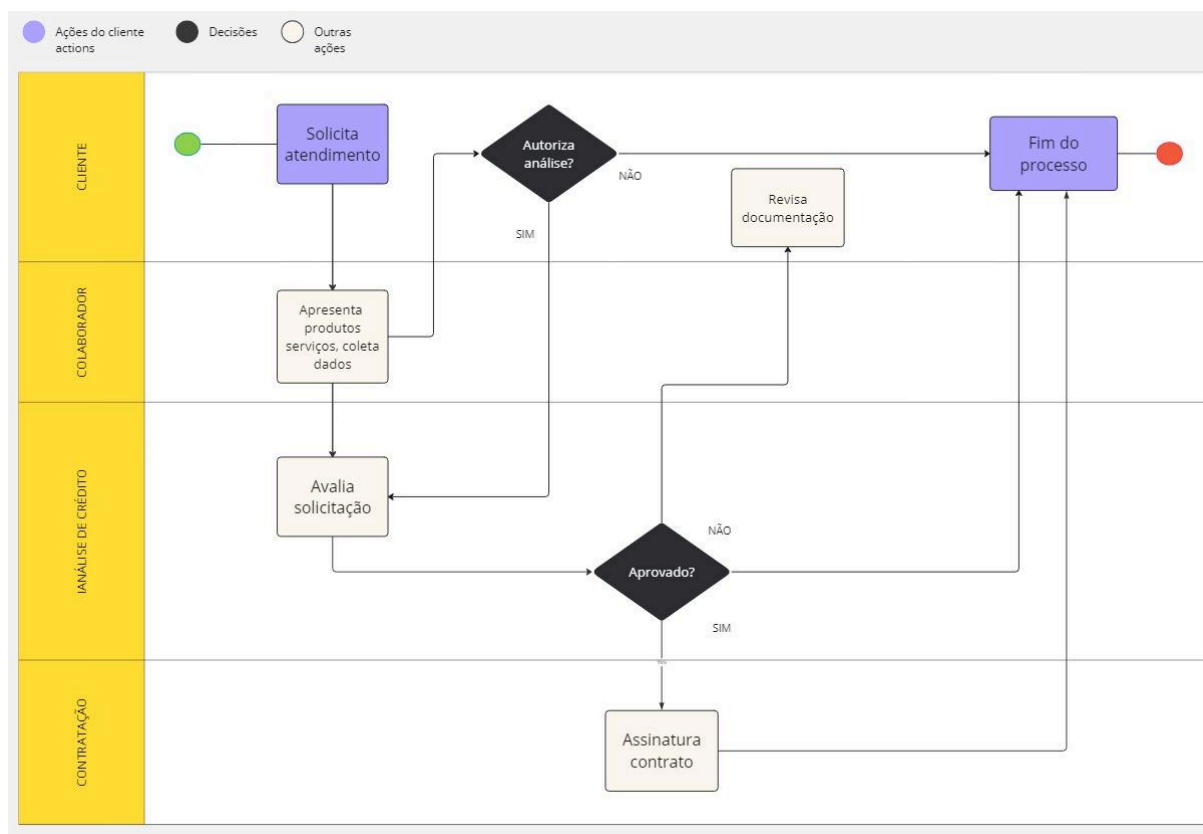


Figura 2: Diagrama de processo

Pensar no processo geral da empresa foi essencial para conhecer suas regras de negócios, inferir sobre documentação e dados produzidos durante as atividades operacionais.

6. Modelo Físico de tabelas Relacional

6.1 Modelo de Dados Físico para Sistema de Concessão de Crédito

No modelo relacional o banco de dados deve representar com precisão os negócios da organização.

6.1.1 Tabela Cliente

TABELA CLIENTE	
Field name	Data Type
idcliente	IDENTITY PRIMARY KEY NN
nome	VARCHAR(100)
cpf	VARCHAR(11) NOT NULL
idade	INT NOT NULL
data_nascimento	DATE NOT NULL
sexo	VARCHAR (1)
email	VARCHAR(100)
telefone	VARCHAR(20)
profissao	VARCHAR (20)
endereco_id	INT
fonte_renda_id	INT
faixa_renda_id	INT
FOREING KEY (endereco_id)	REFERENCES endereco (idendereco)
FOREIGN KEY (fonte_renda_id)	REFERENCES fonte_renda(idfonte_renda)
FOREIGN KEY (faixa_renda_id)	REFERENCES faixa_renda(idfaixa_renda)

Tabela 1: Modelo Físico - Tabela cliente

6.1.2Tabela Produto

TABELA PRODUTO	
Field name	Data Type
idproduto (PK)	INT IDENTITY PRIMARY KEY NOT NULL
nome_produto	VARCHAR(255) NOT NULL
descricao	TEXT
tipo_produto	VARCHAR(100) NOT NULL
valor_minimo	DECIMAL(10, 2) NOT NULL
valor_maximo	DECIMAL(10, 2) NOT NULL
numero_parcelas_maximo	INT NOT NULL
condicoes_especiais	TEXT

Tabela 2: Modelo Físico - Tabela produto

5.1.3 Tabela Endereco

TABELA ENDERECO	
Field name	Data Type
idendereco	INT IDENTITY PRIMARY KEY NOT NULL
cep	VARCHAR(15)
rua	VARCHAR(100) NOT NULL
bairro	VARCHAR(100) NOT NULL
cidade	VARCHAR(100) NOT NULL
estado	VARCHAR(2) NOT NULL

Tabela 3: Modelo Físico - Tabela endereco

6.1.4 Tabela Pagamento

TABELA PAGAMENTO	
Field name	Data Type
idpagamento	INT IDENTITY PRIMARY KEY NOT NULL
numero_parcela	INT NOT NULL,
valor_parcela	DECIMAL(10, 2) NOT NULL,
data_vencimento	DATE NOT NULL,
data_pagamento	DATE
status_pagamento	VARCHAR(20) NOT NULL,
contrato_id (FK)	INT NOT NULL
FOREIGN KEY (contrato_id)	REFERENCES contrato(idcontrato)

Tabela 4: Modelo Físico - Tabela pagamento

6.1.5 Tabela Contrato

TABELA CONTRATO	
Field name	Data type
idcontrato (PK)	INT IDENTITY PRIMARY KEY NOT NULL
codigo_contrato	VARCHAR2(200) UNIQUE NOT NULL
valor_emprestimo	DECIMAL(10, 2) NOT NULL
numero_parcelas	INT NOT NULL
taxa_juros	DECIMAL(5, 2) NOT NULL
data_contratacao	DATE NOT NULL
status_contrato	VARCHAR(20) NOT NULL
cliente_id(FK)	INT NOT NULL
produto_id(FK)	INT NOT NULL
FOREIGN KEY (cliente_id)	REFERENCES cliente(idcliente)
FOREIGN KEY (produto_id)	REFERENCES produto(idproduto)

Tabela 5: Modelo Físico - Tabela contrato

6.1.6 Tabela Fonte de renda

TABELA RELACIONAL FONTE DE RENDA	
Nome do campo	Tipo de dado
idfonte_renda	INT IDENTITY PRIMARY KEY NOT NULL
fonte_renda	VARCHAR(100) NOT NULL

Tabela 6: Modelo Físico - Tabela faixa de renda

6.1.7 Tabela Faixa de renda

Nome do campo	Tipo de dado
idfaixa_renda	INT IDENTITY PRIMARY KEY NOT NULL
faixa_renda	VARCHAR(100) NOT NULL

Tabela 7: Modelo Físico - Tabela fonte de renda

No Modelo físico de dados definimos os atributos, os tipos de dados e os tamanhos de cada campo. O Modelo Lógico de Dados é uma extensão do Modelo Conceitual de Dados, logo o Modelo Físico para o Banco de dados refina ainda mais elementos para construção do Banco de Dados em si.

7. Implementação do Banco de Dados Relacional

O arquivo completo com o SQL para implementação do Banco de Dados encontra-se no Diretório /DataWarehouse/OLTP/DDL_OLTP.

```

idfaixa_renda NUMBER GENERATED AS IDENTITY PRIMARY KEY,
descricao VARCHAR2(100) NOT NULL
);

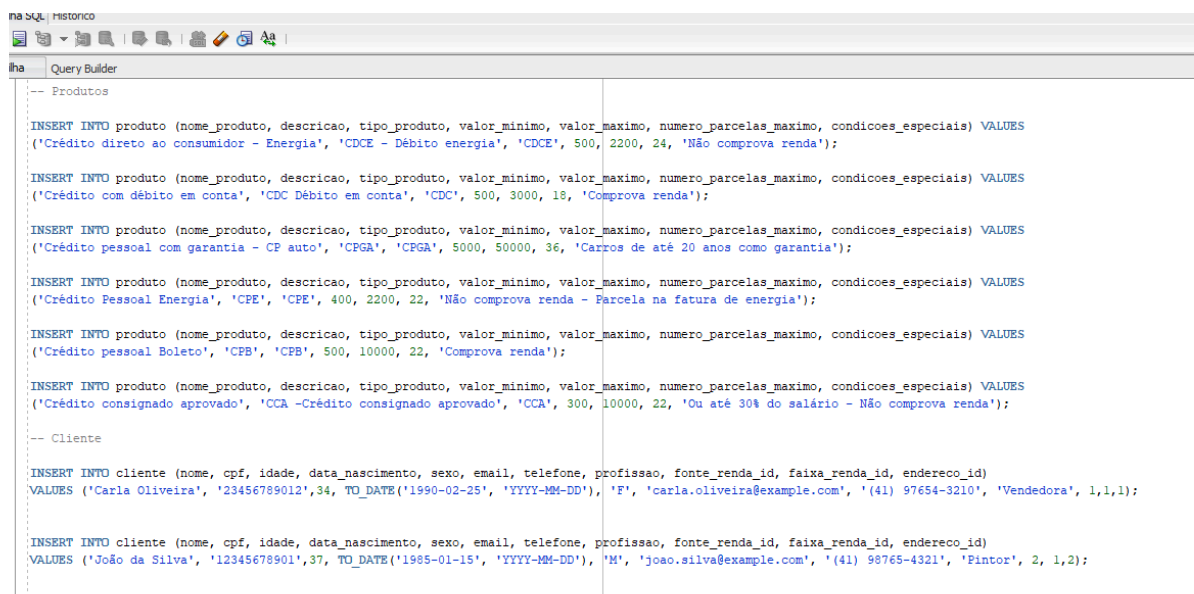
CREATE TABLE cliente (
  idcliente NUMBER GENERATED BY DEFAULT AS IDENTITY PRIMARY KEY,
  nome VARCHAR2(100),
  cpf VARCHAR2(11) UNIQUE NOT NULL,
  idade NUMBER NOT NULL,
  data_nascimento DATE NOT NULL,
  sexo VARCHAR2(1),
  email VARCHAR2(100),
  telefone VARCHAR2(20),
  profissao VARCHAR2(50),
  fonte_renda_id NUMBER,
  faixa_renda_id NUMBER,
  endereco_id NUMBER,
  CONSTRAINT fk_endereco FOREIGN KEY (endereco_id) REFERENCES endereco (idendereco),
  CONSTRAINT fk_fonte_renda FOREIGN KEY (fonte_renda_id) REFERENCES fonte_renda (idfonte_renda),
  CONSTRAINT fk_faixa_renda FOREIGN KEY (faixa_renda_id) REFERENCES faixa_renda (idfaixa_renda)
);

```

Figura 3: DDL - Tabelas relacionais: tabela cliente

8. Populando as tabelas do Banco de Dados Transacional

Depois de criar as tabelas para as transações de rotina da empresa, “populamos” as tabelas com os dados fictícios necessários.



```

na SQL | historico
Query Builder

-- Produtos

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito direto ao consumidor - Energia', 'CDCE - Débito energia', 'CDCE', 500, 2200, 24, 'Não comprova renda');

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito com débito em conta', 'CDC Débito em conta', 'CDC', 500, 3000, 18, 'Comprova renda');

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito pessoal com garantia - CP auto', 'CPGA', 'CPGA', 5000, 50000, 36, 'Carros de até 20 anos como garantia');

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito Pessoal Energia', 'CPE', 'CPE', 400, 2200, 22, 'Não comprova renda - Parcela na fatura de energia');

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito pessoal Boletão', 'CPB', 'CPB', 500, 10000, 22, 'Comprova renda');

INSERT INTO produto (nome_produto, descricao, tipo_produto, valor_minimo, valor_maximo, numero_parcelas_maximo, condicoes_especiais) VALUES
('Crédito consignado aprovado', 'CCA -Crédito consignado aprovado', 'CCA', 300, 10000, 22, 'Ou até 30% do salário - Não comprova renda');

-- Cliente

INSERT INTO cliente (nome, cpf, idade, data_nascimento, sexo, email, telefone, profissao, fonte_renda_id, faixa_renda_id, endereco_id)
VALUES ('Carla Oliveira', '23456789012', 34, TO_DATE('1990-02-25', 'YYYY-MM-DD'), 'F', 'carla.oliveira@example.com', '(41) 97654-3210', 'Vendedora', 1, 1, 1);

INSERT INTO cliente (nome, cpf, idade, data_nascimento, sexo, email, telefone, profissao, fonte_renda_id, faixa_renda_id, endereco_id)
VALUES ('João da Silva', '12345678901', 37, TO_DATE('1985-01-15', 'YYYY-MM-DD'), 'M', 'joao.silva@example.com', '(41) 98765-4321', 'Pintor', 2, 1, 2);

```

Figura 4: DML - inserindo dados no Database

O script completo pode ser encontrado no Diretório /Datawarehouse/OLTP/DML_OLTP.sql

9. Projeto de Data Warehouse

10. O que é um data warehouse?

O Data Warehouse é um sistema de informação focado em apoiar a tomada de decisões. O conceito de Data Warehouse (DW) ou armazém de dados surgiu entre os anos 1980 e 1990, com o trabalho desenvolvido pelos pesquisadores Devlin e Murphy (1988), sob o nome Business Data Warehouse (BDW), que visava integrar dados para apoiar as análises dos dados de uma organização (Monteiro, 2023).

11. Arquitetura de Data Warehouse

A arquitetura de data warehouse é um design intencional de serviços e subsistemas de dados que consolida diferentes fontes de dados em um único repositório para inteligência de negócios (BI), IA/ML e análise. A arquitetura em si é uma maneira estruturada e coerente de armazenar, gerenciar e recuperar grandes quantidades de dados.

Um projeto de construção de DW/DM consiste em várias etapas importantes. São elas:

11.1 Entendimento do Negócio

Levantar requisitos para entender as necessidades da organização é um passo fundamental no início de um projeto de DW/DM. O escopo definido deve incluir as análises desejadas pela organização, as perspectivas de análise e os indicadores a serem analisados. É necessário definir o grão a ser analisado no ambiente e entender como o tempo deve ser tratado no ambiente criado.

11.2 Mapeamento dos Dados

Este passo envolve a verificação da disponibilidade e viabilidade dos dados necessários para construir as análises.

11.3 Construção da Área de Staging de Dados

Uma área onde os dados são armazenados temporariamente para processamento.

11.4 Construção do Processo ETL (Extract, Transform, Load)

O processo de extração de dados dos sistemas de origem, transformação dos dados para atender aos requisitos de análise e carga dos dados no DW/DM.

11.5 Construção das Análises

Especificação e desenvolvimento de consultas, relatórios, aplicativos de análise e outros componentes das aplicações de BI.

De acordo com Kimball e Ross (2013), a arquitetura de um DW/DM consiste em quatro componentes distintos no ambiente de BI:

- Transações de origem
- Sistema ETL
- Área de apresentação
- Aplicações de BI

Para Kimbal (2016) a chave para uma Modelagem Dimensional eficiente consistem em 4 etapas:

1. Selecione o processo de negócios.
2. Declare o grão.
3. Identifique as dimensões.
4. Identifique os fatos.

Para ele, a escolha do processo de negócio é importante porque define uma meta de design específica e permite que o grão, as dimensões e os fatos sejam declarados. Enquanto o grão é a etapa essencial num processo de Modelagem dimensional. Para esse autor, o grão deve ser declarado antes de escolher dimensões ou fatos porque cada dimensão ou fato candidato deve ser consistente com o grão. O grão pressupõem as perguntas comuns do negócio e cada grão de tabela de fatos proposto resulta em uma tabela física separada; grãos diferentes não devem ser misturados na mesma tabela de fatos.

As dimensões fornecem o contexto “quem, o quê, onde, quando, por que e como” em torno de um evento de processo de negócios. As tabelas de dimensões contêm os atributos descritivos usados por aplicativos de BI para filtrar e agrupar os fatos. Com a granulação de uma tabela de fatos firmemente em mente, todas as dimensões possíveis podem ser identificadas

As tabelas fatos correspondem a um evento físico observável, são quase sempre numéricas e resultam de eventos do processo de negócios.

12. Visão, Análise e Métricas

Com o Banco de Dados implementado, voltamos nossa atenção para a construção do projeto de Data Warehouse. A partir das tabelas relacionais destacamos as visões/ dimensões que consideramos essenciais para responder aos principais desafios da empresa. Inferimos algumas análises e elaboramos métricas correspondentes.

12.1.Visão Cliente

Dimensão: cliente	
Análises:	Métrica
Qual a faixa etária dos clientes inadimplentes?	Percentual de inadimplência por faixa etária.
Qual é a distribuição de inadimplência por região geográfica?	Percentual de inadimplência por região.
Quais fontes de renda estão mais associadas à inadimplência?	Percentual de inadimplência por fonte de renda.
Qual a relação entre a faixa de renda dos clientes e a inadimplência?	Percentual de inadimplência por faixa de renda.

Tabela 8: Visão Cliente

12.2.Visão Produto

Dimensão: produto	
Análises:	Métricas:
Quais tipos de produtos financeiros têm maior taxa de inadimplência?	Percentual de inadimplência por tipo de produto
Qual é a relação entre o valor do empréstimo e a inadimplência?	Percentual de inadimplência por faixa de valor do empréstimo.

Table 9: Visão Produto

12.3 Visão Pagamento

Dimensão: pagamento	
Análises:	Métricas:
Qual é o percentual de inadimplência ao longo do tempo?	Taxa de inadimplência mensal/trimestral/anual.
Qual o impacto de diferentes métodos de pagamento na inadimplência?	Percentual de inadimplência por método de pagamento.
Quais são os padrões de pagamento dos clientes inadimplentes?	Análise de atraso de pagamentos (dias de atraso médio).
Existem épocas do ano com maior incidência de inadimplência?	Métrica: Taxa de inadimplência por período do ano

Tabela 10: Visão Pagamento

13. Matriz de Granulidade

A Matriz de granularidade diz respeito ao nível de detalhes dos dados armazenados no Data Warehouse. O grão de uma tabela fato em um Data Warehouse é a definição do nível mais baixo de detalhe que a tabela vai armazenar. Definir o grão corretamente garante que os dados armazenados sejam precisos e detalhados o suficiente para responder às perguntas de negócios.

A Matriz de granularidade apresenta os indicadores, as métricas e os assuntos relacionados aos processos em questão. O grau de detalhes pode ser indicado pelos gestores no momento do Levantamento de requisitos.

MÉTRICAS		ASSUNTO							
Grain Matriz		INDICADORES				Desempenho Financeiro			
DADOS LEVANTADOS						Perfil Demográfico			
		Taxa de aceitação	Taxa de inadimplência	Taxa de aprovação	Valor Médio dos Empréstimos	Taxas de Juros e Condições	Idade Média	Distribuição de Idade	Distribuição de Sexo
Produtos									
Idproduto			TRUE						
nome do produto			TRUE						
tipo de produto			TRUE						
Condições específicas do produto			TRUE						
Clientes									

Figura 5: Matriz de ganulidade

A Matriz de granularidade pode ser encontrada no Diretório DataWarehouse/OLAP/matriz_granulidade.xlsx

14. Modelo Físico de Tabela Dimensional

As tabelas com as dimensões do produto nos ajudou a definir os possíveis campos de análise e as métricas preliminares. Para calcular a relação entre as dimensões nós construímos tabelas fatos, elas conterão os dados numéricos que poderão ser agregados e analisados. Por exemplo, para calcular a relação entre os tipos de produtos, a fonte de renda e a taxa de inadimplência, poderemos modelar uma tabela fato chamada FatoEmprestimo com os dados das tabelas DimCliente, DimProduto, DimPagamento, DimContrato.

14.1. Tabelas dimensionais

Consideremos as seguintes tabelas dimensionais:

14.1.1 DimCliente

DimCliente	
Nome do campo	Tipo de dado
sk_cliente	INT PRIMARY KEY NN
idcliente	INT NN
nome	VARCHAR(100)
cpf	VARCHAR(15) NOT NULL
idade	INT NOT NULL
data_nascimento	DATE NOT NULL
sexo	VARCHAR (1)
email	VARCHAR(100)
telefone	VARCHAR(20)
sk_endereco	INT
fonte_renda_id	INT
faixa_renda_id	INT
FOREIGN KEY (sk_endereco)	REFERENCES DimEndereco(sk_endereco)
FOREIGN KEY (fonte_renda_id)	REFERENCES fonte_renda(idfonte_renda)
FOREIGN KEY (faixa_renda_id)	REFERENCES faixa_renda (idfaixa_renda)

Tabela 11: Dimensão cliente

14.1.2 DimTempo

TABELA DimTempo		
Nome do Campo	Tipo de Dado	Descrição
sk_tempo	DATE	Data completa (chave primária)
ano	INT	Ano da data
trimestre	INT	Trimestre da data (1, 2, 3, 4)
mes	INT	Mês da data (1 a 12)
dia	INT	Dia do mês
dia_semana	INT	Dia da semana (1 = Segunda, ..., 7 = Domingo)
nome_mes	VARCHAR(50)	Nome completo do mês (Janeiro, Fevereiro)

Tabela 12: Dimensão Tempo

14.1.3 DimFaixaRenda

DimFaixaRenda	
Nome do campo	Decrição
sk_faixa_renda	INT IDENTITY (3000,1) NN
idfaixa_renda	INT NN
descricao	VARCHAR (100)

Tabela 13: Dimensão Faixa de Renda

14.1.4. DimFonteRenda

DimFonteRenda	
Nome do campo	Decrição
sk_fonte_renda	INT INDENTITY (4000, 1)NN
idfonte_renda	INT NN
descricao	VARCHAR (100)

Tabela 14: Dimensão Fonte de renda

14.1.5. DimPerfilDemografico

DimPerfilDemografico	
Nome do campo	Tipo de dado
sk_PerfilDemografico	INT IDENTITY PRIMARY KEY NOT NULL
idade_media	INT NOT NULL,
distribuicao_idade	VARCHAR(255) NOT NULL,
distribuicao_sexo	VARCHAR(255) NOT NULL,
distribuicao_geografica	VARCHAR(255) NOT NULL,

Tabela 15: Dimensão Perfil Demográfico

15. Tabela Fato

15.1 Tabela Fato FatoResumoPagamentosAtrasados

FatoTransacao	
Atributo	Descrição
sk_pagamento_atrasad	INT IDENTITY (7000,1) NN
ano	
mes	
total_pagamentos_atrasados	
valor_total_atrasado	
sk_tempo	INT NN
sk_cliente	INT NN
sk_contrato	INT NN
sk_produto	INT NN
Inadimplente	INT NOT NULL
FOREIGN KEY (sk_cliente)	REFERENCES DimCliente(sk_cliente)
FOREIGN KEY (sk_contrato)	REFERENCES DimContrato(skcontrato)
FOREIGN KEY (sk_produto)	REFERENCES DimProduto(sk_produto)
FOREIGN KEY (sk_tempo)	REFERENCES DimTempo(sk_tempo)

Tabela 16: Fato Resumo Pagamentos Atrasados

No exemplo, a tabela fato 'FatoResumoPagamentosAtrasado' pode responder a análise relacionada ao período de maior inadimplência.

16. Tabela Fonte de dados:

A matriz de Fonte de dados descrevemos de onde cada campo nas tabelas dimensionais e de fato origina-se, e como esses campos serão transformados e carregados no Data Warehouse.

Nome da Tabela de Destino	Nome do Campo de Destino	Tipo de Dado de Destino	Nome da Tabela de Origem	Nome do Campo de Origem	Tipo de Dado de Origem
Dimcliente	sk_cliente	INT PRIMARY KEY NN	cliente	id_cliente	INT
Dimcliente	idcliente	INT NN	cliente	id_cliente	INT
Dimcliente	nome	VARCHAR(100)	cliente	nome	VARCHAR(100)
Dimcliente	cpf	VARCHAR(15) NOT NULL	cliente	cpf	VARCHAR(15)
Dimcliente	idade	INT NOT NULL	cliente	data_nascimento	DATE
Dimcliente	data_nascimento	DATE NOT NULL	cliente	data_nascimento	DATE
Dimcliente	sexo	VARCHAR(1)	cliente	sexo	CHAR(1)
Dimcliente	email	VARCHAR(100)	cliente	email	VARCHAR(100)
Dimcliente	telefone	VARCHAR(20)	cliente	telefone	VARCHAR(20)
Dimcliente	fonte_renda_id	INT	cliente	fonte_renda_id	INT
Dimcliente	faixa_renda_id	INT	cliente	faixa_renda_id	INT
DimTempo	sk_tempo	DATE	Data	data	DATE
DimTempo	ano	INT	Data	data	DATE
DimTempo	trimestre	INT	Data	data	DATE
DimTempo	mes	INT	Data	data	DATE

Figura 6: Fonte de dados

17.Implementação de Modelos de tabelas Dimensionais e tabela Fato

O script SQL para implementação do Data Warehouse está no diretório DataWarehouse/OLAP/DML_DW.sql

17.1 Criando a tabela DimCliente:

```
CREATE TABLE DimCliente (
    sk_cliente INT GENERATED BY DEFAULT AS IDENTITY(START WITH 1000) PRIMARY KEY,
    idcliente INT NOT NULL,
    nome VARCHAR2(100),
    cpf VARCHAR2(15) NOT NULL,
    idade INT NOT NULL,
    data_nascimento DATE NOT NULL,
    sexo VARCHAR2(1),
    email VARCHAR2(100),
    telefone VARCHAR2(20),
    sk_endereco INT,
    sk_fonte_renda INT,
    sk_faixa_renda INT,
    CONSTRAINT fk_endereco FOREIGN KEY (sk_endereco) REFERENCES DimEndereco(sk_endereco),
    CONSTRAINT fk_fonte_renda FOREIGN KEY (sk_fonte_renda) REFERENCES DimFonteRenda(sk_fonte_renda),
    CONSTRAINT fk_faixa_renda FOREIGN KEY (sk_faixa_renda) REFERENCES DimFaixaRenda(sk_faixa_renda)
);
```

Figura7:DimCliente

17.2 Criando a tabela DimFonteRenda:

```
CREATE TABLE DimFonteRenda (
    sk_fonte_renda INT GENERATED BY DEFAULT AS IDENTITY (START WITH 4000) PRIMARY KEY,
    idfonte_renda INT NOT NULL,
    descricao VARCHAR2(100)
);
```

Figura 8:DimFonteRenda

17.3 Criando a Tabela DimFaixaRenda:

```
CREATE TABLE DimFaixaRenda (
    sk_faixa_renda INT GENERATED BY DEFAULT AS IDENTITY (START WITH 3000) PRIMARY KEY,
    idfaixa_renda INT NOT NULL,
    descricao VARCHAR2(100)
);
```

Figura 9:DimFaixaRenda

17.4 Criando a tabela DimTempo:

```
CREATE TABLE DimTempo (
    sk_tempo INT GENERATED BY DEFAULT AS IDENTITY PRIMARY KEY,
    data DATE NOT NULL,
    ano INT NOT NULL,
    trimestre INT NOT NULL,
    mes INT NOT NULL,
    dia INT NOT NULL,
    dia_semana INT NOT NULL,
    nome_mes VARCHAR2(50) NOT NULL,
    CONSTRAINT uq_data UNIQUE (data)
);
```

Figura 10:DimTempo

17.6 Criando a tabela Fato 'FatoResumoPagamentosAtrasados':

```
CREATE TABLE FatoResumoPagamentosAtrasados (
    sk_pagamento_atrasado NUMBER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
    ano NUMBER,
    mes NUMBER,
    total_pagamentos_atrasados NUMBER,
    valor_total_atrasado NUMBER,
    tempo_id NUMBER,
    cliente_id NUMBER,
    contrato_id NUMBER,
    produto_id NUMBER,
    CONSTRAINT fk_fato_tempo FOREIGN KEY (tempo_id) REFERENCES DimTempo (sk_tempo),
    CONSTRAINT fk_fato_cliente FOREIGN KEY (cliente_id) REFERENCES DimCliente (sk_cliente),
    CONSTRAINT fk_fato_contrato FOREIGN KEY (contrato_id) REFERENCES DimContrato (sk_contrato),
    CONSTRAINT fk_fato_produto FOREIGN KEY (produto_id) REFERENCES DimProduto (sk_produto)
);
```

Figura 12: FatoResumo Pagamentos Atrasados

18. DataWarehouse: schema estrela (Starchema)

Um esquema estrela (Starchema) apresenta os dados no formato de uma estrela. No centro do Starchema se contra a tabela de fato, com chaves estrangeiras para as tabelas dimensões. As tabelas dimensões refletem a necessidade do negócio.

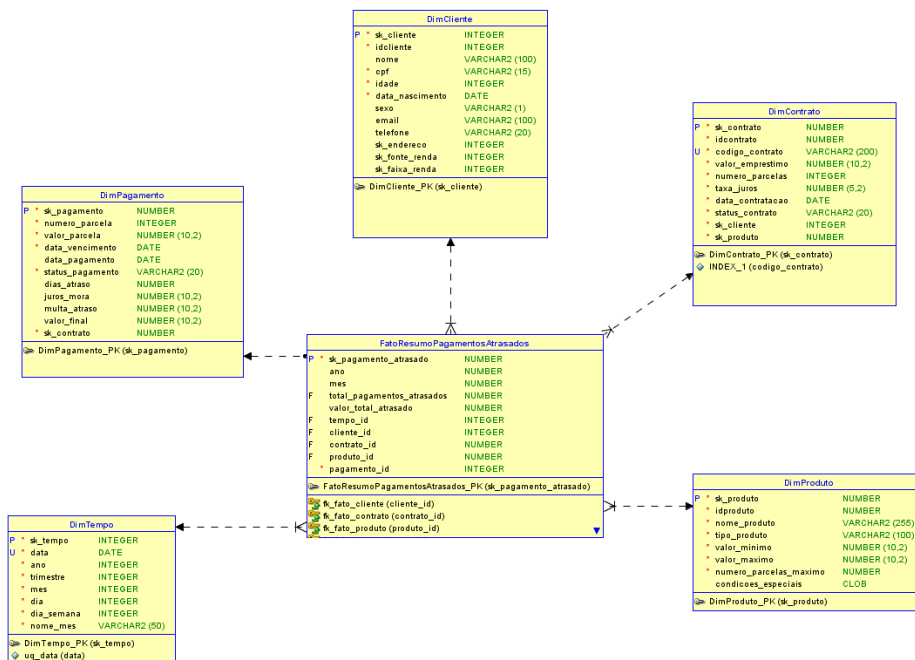


Figura 13: DER -DIAGRAMA ENTIDADE RELACIONAMENTO

19. Consultas analíticas

A criação da tabela Fato deve compreender os fatos / medidas que desejamos armazenar e a identificação das chaves estrangeiras que referenciam as tabelas dimensão. Em seguida criamos a consulta que vai 'popular' a Tabela Fato com os dados necessários e inserimos os dados. Num processo de extração e carregamentos com a linguagem SQL.

A consulta `SELECT * FROM FatoResumoPagamentosAtrasados` :

	SK_PAGAMENTO_ATRASADO	ANO	MES	TOTAL_PAGAMENTOS_ATRASADOS	VALOR_TOTAL_ATRASADO	TEMPO_ID	CLIENTE_ID	CONTRAT...	PRODUTO...	
1		1	2024	4	1	83,33	23	1019	1	4
2		2	2024	3	1	83,33	8	1019	1	4
3		3	2024	3	1	183,36	8	1021	4	4
4		4	2024	3	1	165,02	8	1022	6	4
5		5	2024	3	1	83,33	14	1019	1	4
6		6	2024	3	1	183,36	14	1021	4	4
7		7	2024	3	1	165,02	14	1022	6	4
8		8	2024	3	1	83,33	22	1019	1	4
9		9	2024	3	1	183,36	22	1021	4	4
10		10	2024	3	1	165,02	22	1022	6	4
11		11	2024	2	1	83,33	5	1019	1	4
12		12	2024	2	1	183,36	5	1021	4	4
13		13	2024	2	1	165,02	5	1022	6	4
14		14	2024	2	1	132,72	5	1023	7	4
15		15	2024	2	1	83,33	10	1019	1	4
16		16	2024	2	1	183,36	10	1021	4	4
17		17	2024	2	1	165,02	10	1022	6	4
18		18	2024	2	1	132,72	10	1023	7	4
19		19	2024	2	1	83,33	16	1019	1	4
20		20	2024	2	1	183,36	16	1021	4	4
21		21	2024	2	1	165,02	16	1022	6	4

Figura 14: Resultado consulta Fato Resumo de pagamentos

As consultas na mesma tabela são capazes de mostrar períodos do ano com maior número de parcelas inadimplentes, bem como os meses em que a inadimplência soma o maior valor.

Figura 15: Relatório da tabela Fato Resumo de Pagamentos Atrasados

20. Visualização dos dados

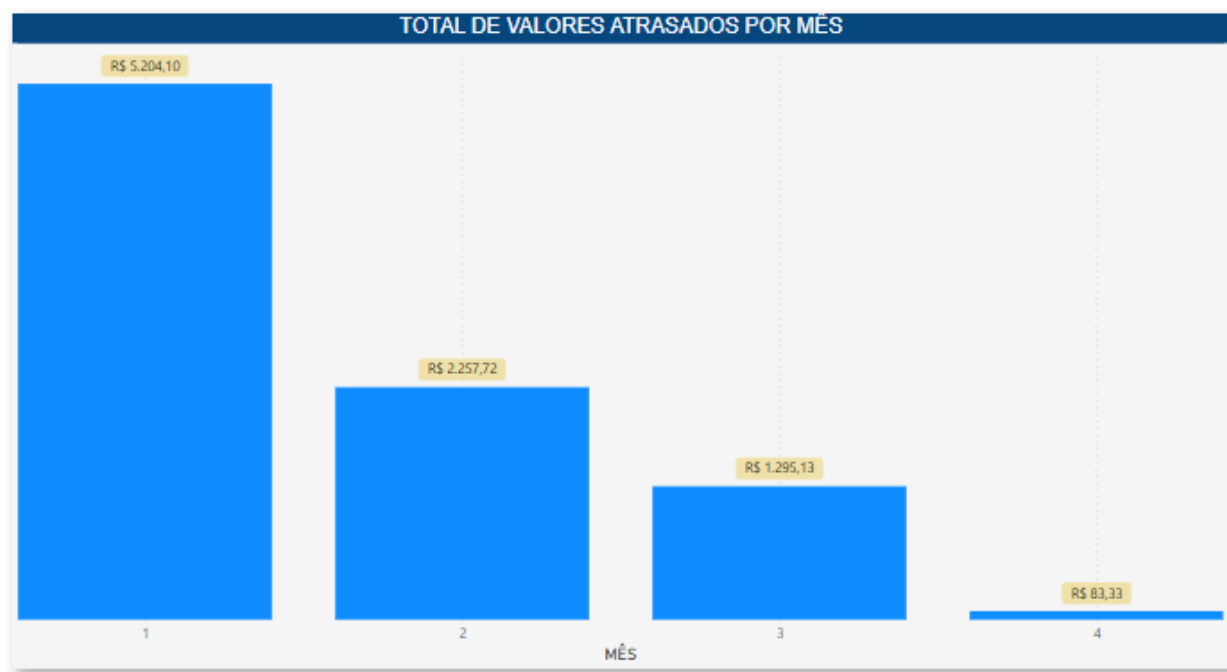


Figura 16: Gráfico: Maior alta de inadimplência por mês

Analisar quais meses do ano tem mais pagamentos atrasados ajuda a identificar padrões sazonais. A organização pode procurar compreender quais fatores afetam esses meses e planejar ações como estratégias de mitigação de risco mais eficazes ou ajustar os critérios de concessão de crédito. A empresa pode preparar campanhas de cobrança ou ofertas de negociação para esse período. Pode oferecer promoções ou incentivos para clientes que pagam em dia durante os meses mais problemáticos. A principal vantagem de compreender os picos de inadimplência seria a Previsão Financeira. Com os dados históricos sobre pagamentos atrasados a empresa pode ajustar suas previsões de fluxo de caixa e garantir que tenha liquidez suficiente para cobrir os períodos de maior inadimplência.

21. Estruturação dos Bancos de Dados

Para essa tarefa estruturamos do zero um Banco de dados Transacional e um ambiente de Data Warehouse, onde:

1. Banco de Dados Transacional (OLTP):
 - Contém as tabelas operacionais que suportam as aplicações do dia-a-dia.
 - Estrutura normalizada para evitar redundância e garantir integridade dos dados.
2. Data Warehouse (DW):
 - Contém as tabelas dimensionais e de fatos usadas para análise e relatórios.
 - Estrutura desnormalizada para facilitar consultas rápidas e eficientes.

22. Processos de ETL

O processo de ETL consistiria em transferir os dados das tabelas Transacionais para o ambiente de Data Warehouse.

Extract:

- Extrair dados das tabelas transacionais (OLTP).
- Feito através de consultas SQL ou ferramentas de ETL.

Transform (Transformar):

- Limpar, transformar e preparar os dados para o formato desejado no DW.
- Pode incluir agregações, cálculos de novas métricas, padronização de formatos, etc.

Load (Carregar):

- Carregar os dados transformados nas tabelas do DW.
- Pode ser feito de forma incremental (somente os novos dados) ou completa (toda a tabela).

Conclusão

Com o desenvolvimento das tabelas dimensionais e a tabela fato, podemos analisar profundamente as causas da inadimplência e outros indicadores financeiros importantes. A modelagem dessas estruturas e a transformação dos dados por meio de processos ETL serão cruciais para garantir a precisão e a relevância das informações armazenadas nos processos transacionais da empresa.

O que se espera é , entre outras soluções, uma melhor compreensão dos perfis de clientes inadimplentes, identificação de produtos com maiores riscos, e períodos de maior incidência de inadimplência. Essas análises permitirão à CrediSimples adotar medidas preventivas, melhorar a experiência do cliente e reduzir significativamente as taxas de inadimplência. A implementação do Data Warehouse é uma estratégia essencial para promover uma cultura orientada a dados na organização.

Referências

Banco do Brasil: Empréstimo Pessoal. Disponível em <<https://www.bb.com.br/site/campanhas/emprestimo/emprestimo-bb>> Acesso em: 11 de jul. 2024.

Índice de inadimplência. CobreFácil, São Paulo, 2024. Disponível em <<https://www.cobrefacil.com.br/blog/indice-inadimplencia>> Acesso em: 11 de jul. 2024

Mapa da inadimplência. Serasa, 2024. Disponível em <<https://cdn.builder.io/o/assets%2Fb212bb18f00a40869a6cd42f77cbeefc%2F76fcc664e2da4b4a9ca3757100cb030e?alt=media&token=1b85bc4d-e047-4391-936f-b01071ec535d&apiKey=b212bb18f00a40869a6cd42f77cbeefc>> Acesso em: 15 jul. 2024

Santander: Como funciona Análise de crédito. Disponível em <<https://www.santander.com.br/blog/analise-de-credito>> Acesso em: 11 de ju;2024.

The world's most valuable resource is no longer oil, but data. The Economist, New York, 6 de maio de 2017. Disponível em <<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>> Acesso em: 05 de jul. 2024

The Kimbal Group. Dimensional Modeling Tecnicas. 2016. Disponível em <<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/>> Acesso em: 10 jul. 2024.