

CASE TÉCNICO



DATA
WAREHOUSE

1. Modelagem dos Data Warehouse

1.1. Descrição do Modelo

Modelo Conceitual e Lógico (Star Schema)

O modelo de Data Warehouse para uma empresa de seguros segue a abordagem **Star Schema**, pois proporciona melhor desempenho em consultas analíticas. Ele possui uma **Tabela Principal** (Sinistro e Prêmio) central conectada a outras **Tabelas Dimensão**.

MODELO CONCEITUAL



MODELO LÓGICO

CLIENTES	SEGURADORA	APOLICE	TEMPO	SINISTRO E PREMIO
ID_Cliente (PK)	ID_Seguradora (PK)	ID_Apolice (PK)	ID_Tempo (PK)	ID_Sinistro (PK)
ID_Apolice (FK - Dim_Apolice)	ID_Tempo (FK - Dim_Tempo)	ID_Seguradora (FK - Dim_Seguradora)	ID_Cliente (FK - Dim_Cliente)	ID_Apolice (FK - Dim_Apolice)
Nome	Nome_Seguradora	Tipo_Apolice	ID_Apolice (FK - Dim_Apolice)	ID_Cliente (FK - Dim_Cliente)
Idade	CNPJ	Valor_Cobertura	Data	ID_Seguradora (FK - Dim_Seguradora)
Sexo	Endereço	Data_Inicio	Ano	ID_Tempo (FK - Dim_Tempo)
Cidade		Data_Fim	Mês	Valor_Sinistro
Estado		Valor_Premio	Trimestre	Quantidade_Sinistros
Tipo_Cliente				Valor_Indenizacao
				Status_Sinistro

Justificativa das Entidades e Relacionamentos

A **Tabela Principal centraliza os eventos financeiros**, permitindo análise de indicadores de rentabilidade e risco.

As **Dimensões fornecem contexto** para os dados financeiros, como características da apólice, perfil do cliente, informações da seguradora e o período de ocorrência dos sinistros.

O **modelo star schema** facilita consultas agregadas e segmentação de análises por tipo de apólice, perfil de cliente, seguradora e períodos históricos.

Exemplos de Métricas-Chave e Fórmulas

Índice de Sinistralidade

Índice de Sinistralidade (%) = $(\text{Valor_Sinistro} / \text{Valor_Premio}) * 100$

→ Mede a relação entre o total de sinistros pagos e os prêmios arrecadados.

Frequência de Sinistros = $\text{Quantidade_Sinistros} / \text{Número Total de Registros}$

→ Indica a média de sinistros por contrato no período.

Valor Médio de Sinistro = $\text{Valor_Sinistro} / \text{Quantidade_Sinistros}$

→ Calcula o custo médio de um sinistro.

Valor Médio da Indenização = $\text{Valor_Indenizacao} / \text{Quantidade_Sinistros}$

→ Mostra o valor médio pago ao cliente por sinistro indenizado.

Distribuição de Sinistros por Tipo de Apólice → Identifica quais categorias de seguro (ex: Automóvel, Vida, Residencial) possuem maior incidência de sinistros.

Evolução Temporal da Sinistralidade → Permite identificar padrões sazonais de sinistros ao longo dos meses, trimestres e anos.

Taxa de Retenção de Clientes

Taxa de Sinistros Pagos (%) = $(\text{Quantidade de Registros com Status_Sinistro} = \text{"Pago"}) / \text{Quantidade_Sinistros} * 100$ → Mede a eficiência na liquidação de sinistros.

Taxa de Sinistros Negados (%) = $(\text{Quantidade de Registros com Status_Sinistro} = \text{"Negado"}) / \text{Quantidade_Sinistros} * 100$ → Analisa a proporção de sinistros recusados.

Comparação da Sinistralidade entre Clientes Pessoa Física e Jurídica

→ Identifica se há diferenças na taxa de sinistros entre os tipos de cliente.

Distribuição de Clientes por Idade e Sexo

→ Permite avaliar perfis predominantes de clientes e possíveis correlações com o risco de sinistro.

Desempenho das Apólices

Rentabilidade por Tipo de Apólice = $[(\text{Valor_Premio} - \text{Valor_Sinistro}) / \text{Valor_Premio}] * 100$

→ Mede a lucratividade de cada categoria de seguro.

Ranking das Apólices mais Rentáveis

→ Classifica os tipos de apólice com melhor desempenho financeiro com base no saldo entre prêmios e sinistros.

Valor Total de Cobertura por Tipo de Apólice

→ Identifica qual modalidade de seguro oferece a maior proteção financeira.

Distribuição de Prêmios Arrecadados por Seguradora

→ Avalia o volume financeiro movimentado por cada seguradora.

Segmentação de Riscos

Mapa de Concentração de Sinistros por Estado e Cidade

→ Com base na localização da Dim_Cliente, identifica regiões com maior incidência de sinistros.

Comparação da Sinistralidade Entre Seguradoras

→ Mede a performance de cada seguradora no pagamento de sinistros.

Taxa de Indenização (%) = $(\text{Valor_Indenizacao} / \text{Valor_Sinistro}) * 100$

→ Indica a proporção de sinistros que resultaram em pagamento ao cliente.

Correlação entre Valor_Cobertura e Sinistralidade

→ Analisa se apólices com maior cobertura tendem a apresentar maior índice de sinistros.

Eficiência Operacional

Comparação de Sinistros Pagos x Sinistros Negados

→ Permite avaliar possíveis gargalos ou inconsistências no pagamento de indenizações.

Evolução do Volume de Sinistros ao Longo do Tempo

→ Utilizando a Dim_Tempo, analisa tendências de crescimento ou redução dos sinistros.

Desempenho das Seguradoras na Liquidação de Sinistros

→ Compara seguradoras quanto à frequência e valores pagos em indenizações.

1.2. Comunicação com Engenharia de Dados

Demanda de Negócio

Problemas que o Data Warehouse (DW) irá resolver

O DW precisa proporcionar uma estrutura consolidada e eficiente para resolver os seguintes problemas da empresa de seguros:

- 1. Análises fragmentadas e demoradas:** A empresa enfrenta dificuldades em realizar análises detalhadas e cruzar dados de diferentes fontes (sinistros, apólices, clientes, seguradoras) de maneira rápida e eficaz. O DW centraliza e integra dados de múltiplos sistemas, permitindo análises de longo prazo com alta performance.
- 2. Falta de visão consolidada do desempenho:** A análise do desempenho das apólices, sinistros e seguradoras é descentralizada. O DW permite que a empresa obtenha uma visão consolidada dos dados, facilitando a comparação entre diferentes tipos de apólices, sinistros e comportamentos dos clientes.
- 3. Dificuldade de identificar padrões de comportamento e riscos:** Com a integração de dados no DW, será possível identificar padrões de comportamento de clientes e sinistros, o que permitirá à seguradora adotar uma estratégia mais assertiva para precificação, mitigação de riscos e prevenção de fraudes.
- 4. Gerenciamento ineficaz da informação temporal:** Sem uma análise temporal eficiente, a empresa encontra dificuldades em acompanhar a evolução de sinistros e apólices ao longo do tempo. O DW resolve essa questão ao incluir a Dim_Tempo, permitindo análises históricas detalhadas e o monitoramento do impacto de decisões ao longo de períodos diferentes.
- 5. Inconsistência na qualidade dos dados:** A empresa enfrenta problemas de dados inconsistentes entre os sistemas de apólices, sinistros e clientes. O DW assegura a qualidade e integridade dos dados, com um processo de ETL (Extração, Transformação e Carga) que padroniza e valida os dados antes de serem carregados, minimizando erros.

Relatórios e Insights Esperados

O Data Warehouse foi projetado para fornecer uma ampla gama de relatórios e insights, com foco na melhoria do desempenho e na redução de riscos. Abaixo estão os principais relatórios e insights esperados:

Tipos de Relatórios

1. Desempenho de Sinistros:

- Relatório de sinistros pagos, em andamento e negados.
- Comparação entre os valores pagos de sinistros versus os prêmios recebidos.
- Identificação de sinistros com valores fora da média ou padrões anormais.

2. Análise de Rentabilidade por Apólice:

- Relatório de rentabilidade por tipo de apólice (Automóvel, Vida, Residencial, etc.).
- Comparação entre o valor do prêmio pago e o valor da indenização paga por tipo de apólice.
- Identificação de apólices que estão gerando mais sinistros e como isso impacta a rentabilidade da seguradora.

3. Análise de Perfil de Clientes:

- Relatório segmentado por **Tipo_Cliente** (Pessoa Física ou Jurídica) e dados sociodemográficos (idade, sexo, cidade, estado).
- Identificação de grupos de clientes com maior número de sinistros ou maior frequência de pedidos de indenização.
- Análise de comportamento de clientes em relação a diferentes tipos de apólices e seus padrões de sinistro.

4. Análise Temporal de Sinistros:

- Relatórios de tendências de sinistros ao longo do tempo (ano, mês, trimestre).
- Identificação de períodos com maior volume de sinistros.
- Análise de sazonalidade e de padrões anuais de sinistros e prêmios.

5. Análise de Desempenho de Seguradoras:

- Relatórios comparando o desempenho de diferentes seguradoras, com base no valor de prêmios recebidos e valores de sinistros pagos.
- Análise de rentabilidade por seguradora, destacando seguradoras com maiores ou menores custos com sinistros.

6. Análise de Cancelamento de Apólices:

- Relatórios de apólices canceladas, com motivos e correlacionados com sinistros registrados.
- Análise de risco e perda de clientes ao longo do tempo.

Insights Esperados

1. Identificação de Riscos e Oportunidades de Mercado:

- O DW permitirá identificar regiões ou tipos de apólices com maior incidência de sinistros, permitindo que a seguradora adapte seus produtos, ajuste precificação ou concentre esforços de marketing em nichos menos arriscados.

2. Previsão de Sinistros e Projeções de Prêmios:

- Através de análises históricas (por exemplo, tendências de sinistros por ano, mês e tipo de apólice), a empresa poderá prever futuros sinistros e projetar prêmios adequados a cada segmento de clientes.

3. Estratégias de Redução de Fraude:

- Com dados completos sobre os sinistros, a seguradora poderá desenvolver modelos analíticos para detectar comportamentos anômalos e prevenir fraudes, com base na combinação de dados históricos de clientes e sinistros.

4. Segmentação de Clientes para Marketing e Vendas:

- O DW permitirá segmentar a base de clientes de acordo com o tipo de apólice, o histórico de sinistros e os comportamentos demográficos, possibilitando campanhas de marketing mais personalizadas e eficazes.

Requisitos Técnicos

Fontes de Dados e Integração

O Data Warehouse deve integrar informações dos diferentes sistemas operacionais da empresa, incluindo:

- **Sistema de Gestão de Apólices:** Contém informações sobre contratos de seguro, valores de cobertura e status das apólices.
- **Sistema de Sinistros:** Registra ocorrências de sinistros, valores pagos e status das indenizações.
- **CRM e Base de Clientes:** Inclui dados sobre perfis de clientes e histórico de relacionamento.
- **Financeiro e Contábil:** Fornece informações sobre pagamentos, recebíveis e fluxo de caixa.

Para isso, pode usar ferramentas como:

- **Fontes de Dados:** Sistemas como HubSpot CRM, QuickBooks, Mitchell e soluções de apólices como Sapiens.
- **ETL e APIs:** Fivetran ou Talend para integração simples de dados.
- **Data Warehouse:** Soluções em nuvem como Google BigQuery ou Amazon Redshift.
- **BI:** Power BI ou Tableau para visualização de dados.

Volume Estimado de Dados e Política de Retenção

- O volume inicial estimado é de **5TB**, com um crescimento anual de aproximadamente **15%**, dependendo do fluxo de novos contratos e eventos. Esse volume reflete aproximadamente 10 bilhões de registros na tabela de fato e 5 bilhões de registros nas tabelas de dimensões.
- Dados transacionais devem ser mantidos no DW por **10 anos** para análise histórica e conformidade regulatória.
- Dados consolidados podem ser armazenados por tempo indeterminado, desde que sejam anonimizados e sigam as diretrizes de governança de dados.

Estratégia de Atualização

A combinação de ambas as estratégias, Batch e Streaming, me parece ideal no caso de uma empresa de seguros.

O uso de **batch** para relatórios regulares e análises históricas e **streaming** para a detecção imediata de eventos críticos, como indícios de fraude ou mudanças no comportamento do segurado, possibilitando respostas rápidas e tomadas de decisão ágeis.

Recomendações de Segurança e Governança de Dados

- **Controle de Acesso Baseado em Função (RBAC):** Permitir acesso diferenciado para analistas, gerentes e times técnicos.
- **Criptografia de Dados Sensíveis:** Aplicar criptografia em dados financeiros e de clientes para conformidade com LGPD.
- **Auditoria e Logs de Acesso:** Manter trilhas de auditoria detalhadas para rastreabilidade.
- **Backup e Recuperação:** Implementar backup com acesso limitado e reduzido.