**Documentação processos de extração de dados SQL**

**Nome extração: Tabela dimensão – Quantidade de advogados ativos**

**1.Objetivo da Extração**

O objetivo desta extração é criar uma tabela dimensão para ser utilizada no projeto de BI do setor Jurídico Imobiliário do escritório. Os dados foram extraídos da tabela de providências históricas (T00076) com o intuito de complementar as informações das tabelas fato já existentes nos projetos de BI do setor imobiliário, contabilizando o número de advogados ativos em cada mês. Essas informações serão utilizadas para o cálculo da produtividade individual de cada advogado, aprimorando as análises de desempenho do setor.

**2. Descrição do Ambiente**

* **Banco de Dados:** SQL Server - ramaprod.
* **Gateway:** Power BI Gateway conectado ao Power BI Service.
* **Ferramenta de Extração:** Fluxo de Dados do Power BI.
* **Permissões:** A consulta SQL tem acesso apenas às colunas e tabelas liberadas dentro do gateway.

**3. Fontes de Dados**

* **Sistema de Extração:** Sistema Espaider Rama
* **Periodicidade da Extração/Atualização:** Extração diária agendada para às 18:30

A consulta faz uso das seguintes tabelas e colunas do banco de dados:

**Tabela: [ramaprod].[dbo].T00076 (Alias: a)**

* Colunas Selecionadas e Chaves:
  + F06337: Utilizada para calcular MIN(a.F06337), YEAR(a.F06337), e MONTH(a.F06337). Filtrada na cláusula WHERE e utilizada na cláusula GROUP BY.
  + F00453: Utilizada na cláusula WHERE para filtrar registros com YEAR(a.F00453) >= 2022.
  + F06982: Chave utilizada em JOIN com a tabela [ramaprod].[dbo].T00041.
  + F00447: Chave utilizada em JOIN com a tabela [ramaprod].[dbo].T00077.
  + F05341: Chave utilizada em JOIN com a tabela [ramaprod].[dbo].T00003.
  + F05633: Chave utilizada em JOIN com a tabela [ramaprod].[dbo].T00557.

**Tabela: [ramaprod].[dbo].T00077 (Alias: b)**

* Colunas Selecionadas e Chaves:
  + ID: Chave primária da tabela [ramaprod].[dbo].T00077, utilizada em JOIN com a coluna F00447 da tabela [ramaprod].[dbo].T00076.

**Tabela: [ramaprod].[dbo].T00041 (Alias: c)**

* Colunas Selecionadas e Chaves:
  + F01187: Utilizada na cláusula WHERE para filtrar registros com valores específicos.
  + ID: Chave primária da tabela [ramaprod].[dbo].T00041, utilizada em JOIN com a coluna F06982 da tabela [ramaprod].[dbo].T00076.

**Tabela: [ramaprod].[dbo].T00003 (Alias: d)**

* Colunas Selecionadas e Chaves:
  + ID: Chave primária da tabela [ramaprod].[dbo].T00003, utilizada em JOIN com a coluna F05341 da tabela [ramaprod].[dbo].T00076.

**Tabela: [ramaprod].[dbo].T00557 (Alias: e)**

* Colunas Selecionadas e Chaves:
  + F05200: Chave utilizada em JOIN com a tabela [ramaprod].[dbo].T00030.
  + ID: Chave primária da tabela [ramaprod].[dbo].T00557, utilizada em JOIN com a coluna F05633 da tabela [ramaprod].[dbo].T00076.

**Tabela: [ramaprod].[dbo].T00030 (Alias: f)**

* Colunas Selecionadas e Chaves:
  + F00091: Utilizada na cláusula WHERE para filtrar registros com valores específicos e para calcular COUNT(DISTINCT f.F00091).
  + ID: Chave primária da tabela [ramaprod].[dbo].T00030, utilizada em JOIN com a coluna F05200 da tabela [ramaprod].[dbo].T00557.

**4. Código SQL Completo**

*SELECT*

*MIN(a.F06337) AS cumpri\_em,*

*YEAR(a.F06337) AS cumprido\_em\_ano,*

*MONTH(a.F06337) AS cumprido\_em\_mes,*

*COUNT(DISTINCT f.F00091) AS qtde\_advogados*

*FROM [ramaprod].[dbo].T00076 AS a*

*LEFT JOIN [ramaprod].[dbo].T00077 AS b ON a.F00447 = b.ID*

*LEFT JOIN [ramaprod].[dbo].T00041 AS c ON a.F06982 = c.ID*

*LEFT JOIN [ramaprod].[dbo].T00003 AS d ON a.F05341 = d.ID*

*LEFT JOIN [ramaprod].[dbo].T00557 AS e ON a.F05633 = e.ID*

*LEFT JOIN [ramaprod].[dbo].T00030 AS f ON e.F05200 = f.ID*

*WHERE*

*YEAR(a.F06337) IS NOT NULL AND*

*YEAR(a.F00453) >= 2022 AND*

*c.F01187 IN (32, 34, 14, 36, 35) AND*

*f.F00091 IN ('Ana Carolina Bressan da Silva', 'Gustavo Araujo Tavares', 'Erick Damin Bitencourt', 'Darlei Jacoby Kayser', 'Thierri Rech',*

*'Bruno Gonçalves Barrios', 'João Lucas Martins Falcão', 'Sabrina de Jesus Pereira', 'Felipe Machado da Luz', 'Rafaella Rodrigues dos Santos Marques',*

*'Rafael Rama e Silva', 'Thiago Haggstrom dos Santos', 'Thiago dos Santos', 'Danielle Lais da Silva Lutkemeyer','Augusto Almeida Gerhardt','Jenifer da Silva Lopes', 'Matheus Cezar Dias',*

*'Sarah Raquel Lopes Gonçalves')*

*GROUP BY YEAR(a.F06337), MONTH(a.F06337)*

*ORDER BY cumprido\_em\_ano, cumprido\_em\_mes;*

**5. Destinos dos Dados Extraídos**

Os dados extraídos por meio desta consulta SQL são carregados diretamente no Power BI Service através do fluxo de dados configurado no Power BI. Esses dados são armazenados em datasets no Power BI, onde são utilizados em dashboards e relatórios para análise e visualização.

* **Destino:** Projeto de BI Jurídico Imobiliário.
* **Formato dos Dados:** Os dados são carregados em formato tabular no Power BI Service, prontos para visualização em relatórios e dashboards.

**6. Qualidade e Segurança**

Antes que os dados sejam considerados prontos para uso nos dashboards, são aplicados os seguintes critérios de qualidade:

* **Integridade dos Dados**
* A integridade dos dados é verificada e atestada através do banco de dados relacional, com uso das chaves primárias de cada tabela que participa de junções.
* Todo dataset oriundo da extração passa por uma análise exploratória para validação das colunas. Valores nulos e/ou inconsistentes são tratados da maneira que a regra de negócio do escopo permite.
* **Precisão dos Dados:**
* A precisão dos dados é checada através de validações diretamente no sistema Espaider Rama e através de planilhas de cada setor. Cada coluna é validada usando esses comparativos citados.
* **Versionamento:**
* Todos os códigos de extrações e documentações são armazenados localmente na pasta: C:\Users\pedro.cecere\Documents\Projetos\A - Jurídico Imobiliário\A - 06 - Queries BI. Esta pasta é adicionada remotamente ao repositório criado no GitHub: A-04-Queries\_BI.
* Qualquer alteração no código é postada no diretório do GitHub, onde temos acesso ao histórico de versões deste código, mantendo assim uma linha cronológica de todas versões do código desenvolvido.

Sobre segurança e permissões de acessos aos dados, são tomadas as seguintes medidas:

* **Acessos e Permissões:**
  + A extração de dados é realizada de acordo com as permissões estabelecidas pela Smart Suport, empresa terceirizada responsável por hospedar o banco de dados. O acesso é gerenciado através do gateway do Power BI, que controla quais colunas e tabelas podem ser acessadas. Somente usuários autorizados pelo administrador do gateway têm permissão para modificar ou executar essa consulta.
* **Conformidade com Políticas de Privacidade:**
  + Os dados extraídos são restritos a informações profissionais e não incluem dados sensíveis. Todo o processo de extração é realizado em conformidade com as normas e diretrizes da LGPD, bem como as exigências estabelecidas pelo Banco Santander.
* **Logs e Auditoria:**
  + O Power BI Service registra todas as modificações no fluxo de dados, detalhando quem fez as alterações e em que momento. Esses logs são mantidos para fins de auditoria e conformidade. Da mesma forma, o versionamento realizado no GitHub permite rastrear todas as alterações e identificar os responsáveis por cada modificação.
* **Encriptação de Dados:**
  + Os dados são encriptados, por padrão, em trânsito e em repouso no Power BI Service, garantindo que as informações estão protegidas contra acessos não autorizados.