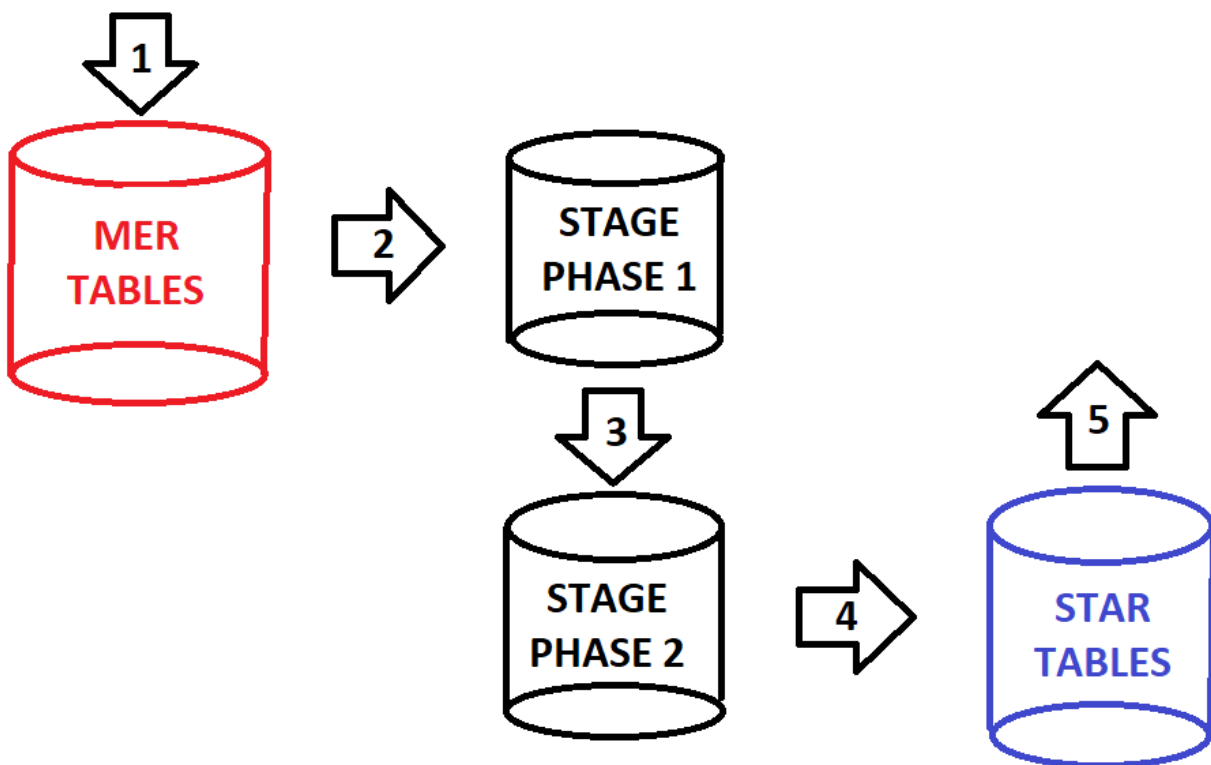


Capítulo 5 - OLAP // FASE 4

Início

O objetivo deste projeto é adaptar nossa tese do **Construmax** para o ambiente OLAP. Para quem não conhece nossa proposta, deixo aqui este [documento explicativo](#). Sendo breve estamos desenvolvendo uma aplicação que dê mais opção para moradores que desejam realizar uma reforma/obra em sua residência, dando mais oportunidades e crescimento para os prestadores de serviço na área de construção civil.

Temos o seguinte esquema abaixo:



Nós queremos transformar as tabelas do nosso modelo de entidade relacional (MER) representadas pelo **elemento vermelho** para o **elemento azul** que vai agrupar as tabelas dimensionais (STAR). Este processo será dividido em **5 etapas**.

- A etapa **1** será a criação das tabelas dos conjuntos ilustrados.
- As etapas **2, 3 e 4** serão o **ETL** do projeto. As tabelas MER vão passar por duas fases de **Staging**. Primeiro vamos extrair os dados para tabelas sem relacionamento (STAGE PHASE 1), depois vamos transforma-las em outras tabelas que prototipam nossas dimensões e fatos (STAGE PHASE 2) e por fim realizar a carga desses dados para as tabelas STAR.
- Todo o processo será feito utilizando a ferramenta do **Pentaho Data Integration**.
- Na etapa **5** vamos criar um cubo utilizando o **Modrian Workbench** e publica-lo no **Pentaho BI Server**.

ETAPA DE CRIAÇÃO

The diagram illustrates the database schema for the T2 system, showing the relationships between various tables. The tables are represented by yellow boxes, and the relationships are indicated by lines with crow's foot notation.

Tables and their attributes:

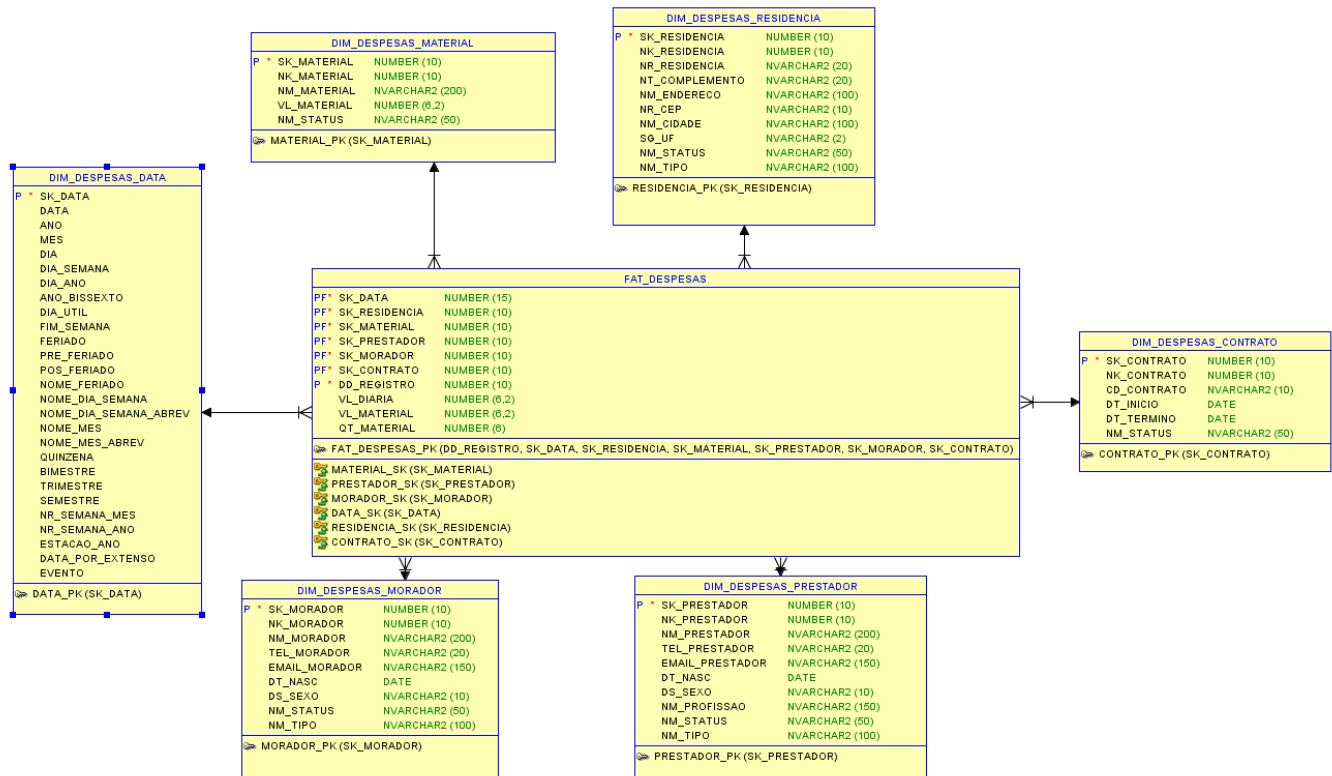
- T2_ST_MOR**: CD_STATUS (NUMBER(10)), NM_STATUS (NVARCHAR2(50)).
- T2_TP_MOR**: CD_TIPO (NUMBER(10)), NM_TIPO (NVARCHAR2(100)).
- T2_MORADOR**: CD_MORADOR (NUMBER(10)), NM_MORADOR (NVARCHAR2(200)), TEL_MORADOR (NVARCHAR2(20)), EMAIL_MORADOR (NVARCHAR2(150)), DT_NASC (DATE), DS_SEXO (NVARCHAR2(10)), CD_TIPO (NUMBER(10)), CD_STATUS (NUMBER(10)).
- T2_RESIDENCIA**: CD_RESIDENCIA (NUMBER(10)), NR_RESIDENCIA (NUMBER(10)), NT_COMPLEMENTO (NVARCHAR2(20)), CD_ENDERECO (NUMBER(10)), CD_STATUS (NUMBER(10)), CD_TIPO (NUMBER(10)), CD_MORADOR (NUMBER(10)).
- T2_ENDERECO**: CD_ENDERECO (NUMBER(10)), NM_ENDERECO (NVARCHAR2(100)), NR_CEP (NVARCHAR2(10)), CD_CIDADE (NUMBER(10)).
- T2_CIDADE**: CD_CIDADE (NUMBER(10)), NM_CIDADE (NVARCHAR2(100)), CD_UF (NVARCHAR2(2)), CD_CIDADE_PK (CD_CIDADE), UF_FK (CD_UF).
- T2_UF**: CD_UF (NVARCHAR2(2)), SG_UF (NVARCHAR2(2)).
- T2_CONTRATO**: NR_CONTRATO (NUMBER(10)), CD_CONTRATO (NVARCHAR2(10)), DT_INICIO (DATE), DT_TERMINO (DATE), CD_STATUS (NUMBER(10)).
- T2_ST_CON**: CD_STATUS (NUMBER(10)), NM_STATUS (NVARCHAR2(50)).
- T2_TP_PREST**: CD_TIPO (NUMBER(10)), NM_TIPO (NVARCHAR2(100)), VL_DIARIA (NUMBER(6,2)).
- T2_PROFISSAO**: CD_PROFISSAO (NUMBER(10)), NM_PROFISSAO (NVARCHAR2(150)).
- T2_PRESTADOR**: CD_PRESTADOR (NUMBER(10)), NM_PRESTADOR (NVARCHAR2(200)), TEL_PRESTADOR (NVARCHAR2(20)), EMAIL_PRESTADOR (NVARCHAR2(150)), DT_NASC (DATE), DS_SEXO (NVARCHAR2(10)), CD_PROFISSAO (NUMBER(10)), CD_TIPO (NUMBER(10)), CD_STATUS (NUMBER(10)).
- T2_REGISTRO**: NR_REGISTRO (NUMBER(10)), DT_REGISTRO (DATE), QT_MATERIAL (NUMBER(6)), CD_RESIDENCIA (NUMBER(10)), CD_MATERIAL (NUMBER(10)), NR_CONTRATO (NUMBER(10)), CD_PRESTADOR (NUMBER(10)).
- T2_MATERIAL**: CD_MATERIAL (NUMBER(10)), NM_MATERIAL (NVARCHAR2(200)), VL_MATERIAL (NUMBER(6,2)), CD_STATUS (NUMBER(10)).
- T2_ST_MAT**: CD_STATUS (NUMBER(10)), NM_STATUS (NVARCHAR2(50)).
- T2_ST_PREST**: CD_STATUS (NUMBER(10)), NM_STATUS (NVARCHAR2(50)).

Relationships:

- T2_MORADOR** is associated with **T2_ST_MOR** (1:M), **T2_TP_MOR** (1:M), and **T2_RESIDENCIA** (1:M).
- T2_RESIDENCIA** is associated with **T2_ENDERECO** (1:M).
- T2_ENDERECO** is associated with **T2_CIDADE** (1:M).
- T2_CIDADE** is associated with **T2_UF** (1:M).
- T2_CONTRATO** is associated with **T2_ST_CON** (1:M).
- T2_CONTRATO** is associated with **T2_TP_PREST** (1:M).
- T2_CONTRATO** is associated with **T2_PROFISSAO** (1:M).
- T2_CONTRATO** is associated with **T2_PRESTADOR** (1:M).
- T2_CONTRATO** is associated with **T2_REGISTRO** (1:M).
- T2_PRESTADOR** is associated with **T2_REGISTRO** (1:M).
- T2_PRESTADOR** is associated with **T2_PROFISSAO** (1:M).
- T2_PRESTADOR** is associated with **T2_TP_PREST** (1:M).
- T2_PRESTADOR** is associated with **T2_ST_PREST** (1:M).
- T2_REGISTRO** is associated with **T2_MATERIAL** (1:M).
- T2_REGISTRO** is associated with **T2_ST_MAT** (1:M).

The diagram uses crow's foot notation to represent the relationships, including cardinalities (1, M) and relationship types (one-to-many, many-to-many).

Para a análise, criamos o seguinte modelo dimensional:



Nós centralizamos os fatos encima do preço e quantidade dos materiais e sobre valor da diária do profissional. Em relação a qualidade dos dados criamos as dimensões data, material, morador, prestador, residência e contrato.

Segue abaixo a lista dos scripts de criação de todas as tabelas do projeto e carga da dimensão data. Execute-os nessa ordem que não terá problemas.

1. Criação tabelas MER
2. Criação tabelas STG para MER
3. Criação tabelas STAR
4. Criação tabelas STG para STAR
5. Carga dim_despesas_data

Sobre o Rollback

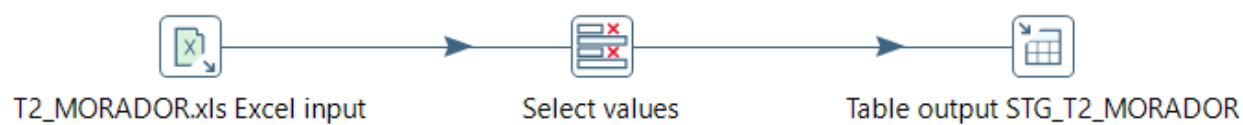
Caso ocorra algum erro no processo você sempre pode reinicia-lo executando nosso script de [rollback](#).

ETAPA ETL

Para que ocorra a extração antes precisamos ter os dados em nossa tabelas, buscando cumprir com a solicitação da atividade, colocamos nossos dados em tabelas de `excel`. Clique [aqui](#) para visualizar os arquivos.

EXTRAÇÃO

Realizamos no total de 18 extrações pegando os dados presentes nas planilhas de excel e incluindo-os dentro das tabelas de início **stg_t2** dentro do banco de dados Oracle. Segue abaixo o fluxo de uma das extrações.



TRANSFORMAÇÃO

Após incluído os dados, iniciamos o processo de transformação do qual juntamos as informações das 18 tabelas stg_t2 em 5 tabelas de início **stg2**. No caso abaixo, STG2_MORADOR será uma junção dos dados que foram armazenados em STG_T2_MORADOR, STG_T2_TP_MOR e STG_T2_ST_MOR.



Letura de Tabela

Nome do Step: Table input STG_T2_MORADOR

Connection: Conn-Cap 5 OLAP

Buttons: Edit... New... Wizard...

SQL: Get SQL select statement...

```

SELECT
  M.CD_MORADOR SK_MORADOR,
  M.CD_MORADOR NK_MORADOR,
  M.NM_MORADOR,
  M.TEL_MORADOR,
  M.EMAIL_MORADOR,
  M.DT_NASC,
  M.DS_SEXO,
  T.NM_TIPO,
  S.NM_STATUS
FROM RM88633.STG_T2_MORADOR M
INNER JOIN RM88633.STG_T2_TP_MOR T ON T.CD_TIPO = M.CD_TIPO
INNER JOIN RM88633.STG_T2_ST_MOR S ON S.CD_STATUS = M.CD_STATUS
  
```

Linha 1 Coluna 0

Store column info in step meta data ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Executar para cada linha? ☐

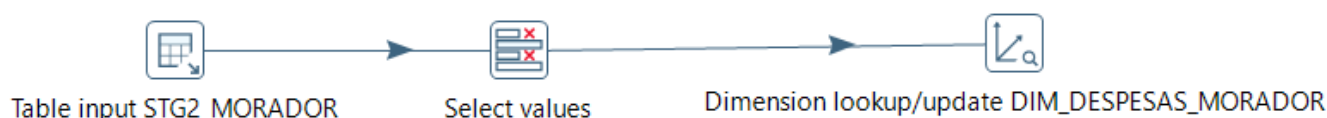
Tamanho limite 0

Buttons: Help OK Preview Cancela

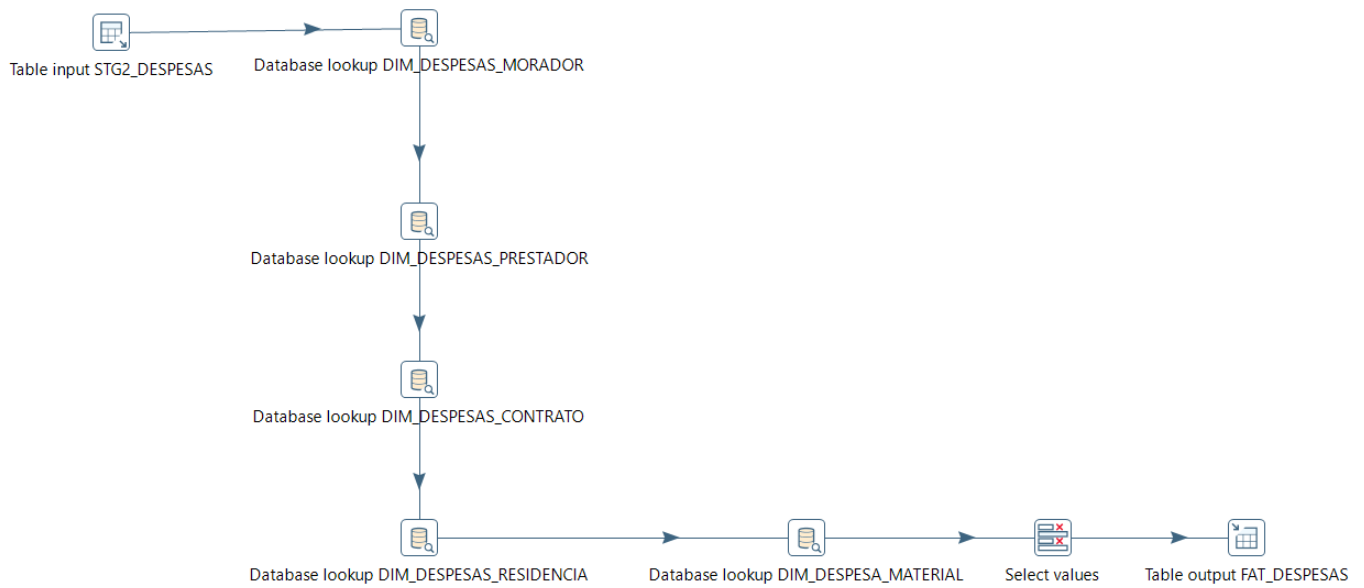
CARGA

Feita a transformação, chegou a hora de efetuarmos a carga dos dados colocados nas tabelas STG2 para nossas dimensões e fato. Abaixo um fluxo que como foi feita a carga das dimensões e fato.

- Dimensão MORADOR



- Fato DESPESAS



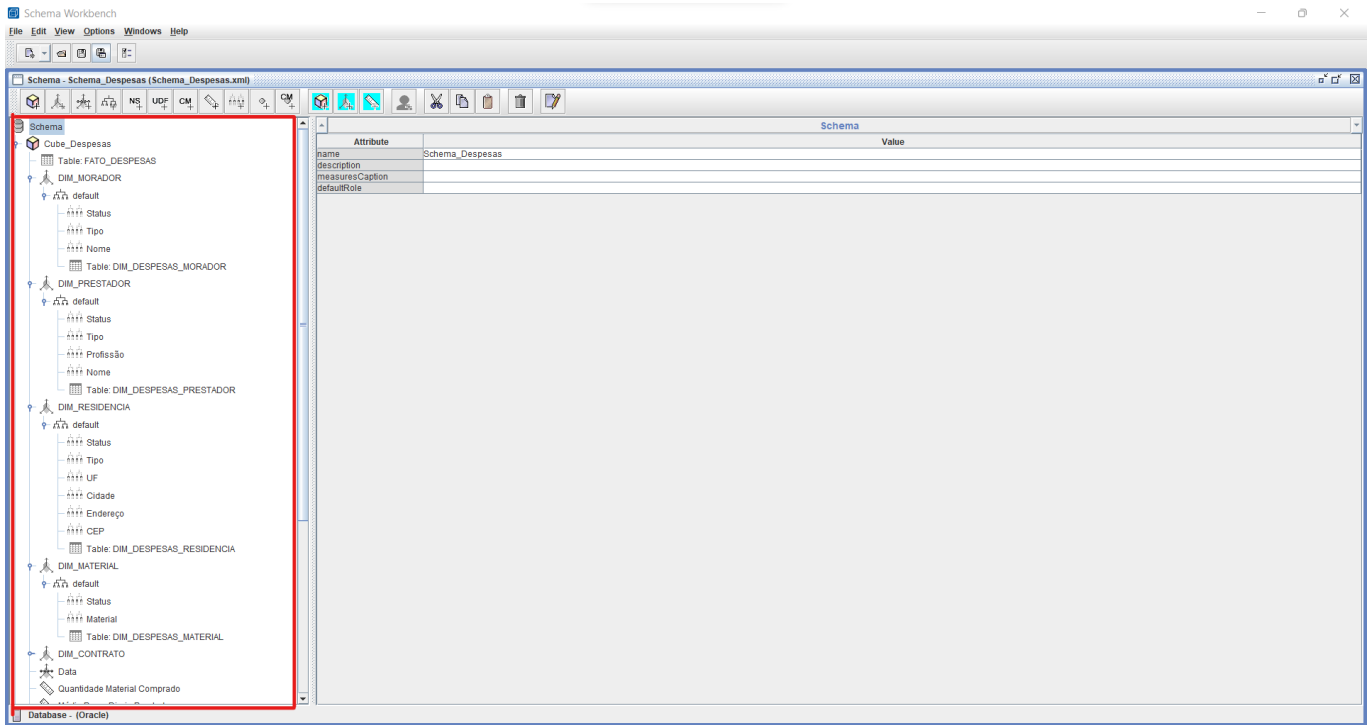
Todo o processo foi automatizado por meio de jobs do Pentaho Data Integration, segue abaixo a esquematização final desta etapa.



Montamos este [vídeo](#) demonstrando o processo. Lembrando que você pode ter acesso a todos arquivos utilizados dentro deste [repositório](#).

ETAPA OLAP

Utilizando o **Modrian**, pegamos os dados inseridos nas tabelas dimensões e fato do banco para modelar o nosso cubo, segue abaixo como ficou a configuração dentro do software.



Cubo montado, vamos publica-lo no nosso server BI local. Utilizando a ferramenta analítica **saiku**, podemos montar diversos relatórios utilizando os dados. No exemplo abaixo, montamos uma tabela com o intuito de observar a quantidade de materiais que prestadores ativos consumiram em obras de apartamentos residenciais ao longo dos anos. Podemos observar que há um detalhamento dos anos em semestres, trimestres e meses assim como uma categorização dos prestadores em 3 níveis (Aprendiz, Pleno e Sênior). Também podemos notar uma separação dos prestadores por estado onde está ocorrendo as obras.

Ano	Semestre	Mes	BA			MG			RJ		
			APRENDIZ	PLENO	SENIOR	APRENDIZ	PLENO	SENIOR	APRENDIZ	PLENO	SENIOR
2020	1	1	-	5	2	12	5	17	25	23	11
		2	-	-	2	9	3	9	-	-	11
		3	-	-	-	-	-	-	-	-	4
		4	-	-	-	-	2	-	-	-	-
		5	-	-	2	5	1	2	3	-	7
		6	-	-	-	-	-	4	-	-	-
	2	7	-	8	-	6	3	11	25	23	-
		8	-	-	-	6	-	-	-	10	-
		9	-	-	-	-	-	4	7	-	-
		10	-	-	-	-	-	-	9	3	-
		11	-	-	-	-	-	-	2	10	-
		12	-	-	-	-	3	7	-	-	-
2021	1	1	-	10	14	41	44	21	5	25	19
		2	-	2	12	19	17	21	5	10	14
		3	-	-	-	9	-	-	-	3	8
		4	-	-	-	8	-	-	5	6	-
		5	-	-	-	2	-	-	-	1	-
		6	-	-	-	-	-	-	-	-	8
	2	7	-	8	2	22	27	-	-	15	5
		8	-	-	2	-	-	-	-	-	-
		9	-	-	-	6	10	-	-	10	-
		10	-	-	-	18	-	-	-	2	-
		11	-	-	-	-	1	-	-	-	5
		12	-	-	-	-	9	-	-	-	-
2022	1	1	-	2	-	21	25	8	5	5	5
		2	-	-	-	21	19	-	5	-	-
		3	-	-	-	-	-	-	3	-	-
		4	-	-	-	7	9	-	-	-	-
		5	-	-	-	14	10	-	-	-	-
		6	-	-	-	-	-	-	-	-	-

Da mesma forma que fizemos na etapa de ETL, gravamos um [vídeo](#) ilustrando o resultado final do trabalho.

Forte Abraço