

Fluxo Estudantil no Ensino Superior

Projeto BI

Histórico de Versões

Data	Versão	Descrição	Autor	Aprovado por
21/02/2021	1.0	Definição Tema – Evasão no Ensino Superior	Todos os Componentes	Todos os Componentes
23/02/2021	2.0	Consolidação dos Dados – Aquisição dos dados da fonte INEP.gov	Todos os Componentes	Todos os Componentes
24/02/2021	3.0	ETL – 1 (EXCEL): Primeira exploração visual da base de dados e ETL via Excel com exclusão de linhas e colunas desnecessárias ao estudo.	Todos os Componentes	Todos os Componentes
25/02/2021	4.0	ETL – 2 (PDI): Carga dos dados no PDI e Operações de formatação e normalização.	Todos os Componentes	Todos os Componentes
25/02/2021	5.0	ETL – 3 Carga dos dados via Power BI e tratamento de categorização pelo Power Query	Todos os Componentes	Todos os Componentes
26/02/2021	6.0	ETL – 4 (PDI inserção de dados dentro da estrutura do banco de dados criado)	Todos os Componentes	Todos os Componentes
02/03/2021	7.0	Preparação do Relatório	Todos os Componentes	Todos os Componentes
03/03/2021	8.0	Necessidade da criação de um DW e para adequar o projeto as características exigidas.	Todos os Componentes	Todos os Componentes
04/03/2021	9.0	Elaboração teórica da Arquitetura do DW e utilização do Power Architect para construção do script SQL e carga via PDI no DW. Resultado das visualizações inconsistentes, NECESSIDADE DA reconstrução da arquitetura do DW.	Todos os Componentes	Todos os Componentes
05/03/2021	10.0	Reestruturação teórica do DW, onde ficou decidido o modelo DW_ENSINO	Todos os Componentes	Todos os Componentes
06/03/2021	11.0	Construção da nova estrutura dos DW via Power Architect, novo tratamento de ETL pelo PDI e carga dos dados nos DWs correspondentes.	Todos os Componentes	Todos os Componentes
08/03/2021	12.0	Exploração da base de dados para construção da visualização pelo Power BI e preparação do relatório final.	Todos os Componentes	Todos os Componentes
09/03/2021	13.0	Adição de comentários extras para entendimento do estudo e dos Dashboards criados no Power BI. Entrega	Todos os Componentes	Todos os Componentes

Sumário

1	INTRODUÇÃO	4
2	ESTUDO DE CASO	5
2.1	DESCRIÇÃO DO ESTUDO DE CASO	5
2.2	PLANO DE AÇÃO 5W2H	6
3	ELABORAÇÃO DO DATA WAREHOUSE	8
3.1	DEFINIÇÃO DO DW	10
4	DESCRIÇÃO DO MODELO MULTIDIMENSIONAL.....	11
5	PROPOSTA DE PROCESSO DE BI.....	14
5.1	ETAPAS PARA CONSTRUÇÃO DA ESTRUTURA DO PROJETO DE BI.	14
6	PROJETO DE ETL.....	16
6.1	DESCRIÇÃO DO PROJETO DE ETL	16
7	DASHBOARD	20
7.1	DESCRIÇÃO DA ELABORAÇÃO	20
7.2	TELAS DO DASHBOARD	20
8	CONCLUSÃO	22

1 Introdução

Este documento tem por finalidade coletar, analisar e definir como as informações serão organizadas a fim de se encontrar respostas as necessidades do Projeto de BI em que o estudo de caso escolhido foi usar o **Fluxo Estudantil no Ensino Superior** e entender quais são as proporções de concluintes, desistentes, permanentes, falecidos e ingressantes, existente nessa esfera da Educação Brasileira.

Através deste estudo, buscamos definir conhecimentos ocultos na base de dados relativa à evasão no Ensino Superior, bem como indicar quantidades e taxas referentes aos Cursos disponibilizados nas instituições tanto públicas quanto as privadas.

2 Estudo de Caso

2.1 Descrição do Estudo de Caso

O estudo de caso é referente ao Fluxo Estudantil existente Ensino Superior com o foco sobre a evasão, onde os dados que foram extraídos para o desenvolvimento do projeto são oriundos de uma pesquisa realizada pelo INEP, e que agora é uma extensão do Ministério da Educação. Essa pesquisa foi realizada entre os anos de 2015 e 2019.

No [site](#) do instituto, é possível, através da seção de dados abertos, encontrar diversas base referentes ao Ensino Superior, e foi nesta que coletamos a base de dados principal do nosso estudo de caso.

A ideia e proposta central do nosso estudo, **é definir as métricas descritas abaixo**, a fim de atender à uma necessidade hipotética criada por nós, onde foi “requerida pelo Ministério da Educação”, no qual **deseja entende** num estudo posterior o porquê da **evasão** em determinadas **áreas de formação, instituição de ensino** e mais profundamente e principalmente quem são os **alunos classificados como desistentes** na pesquisa. E para chegar a esses agentes desistentes, inicialmente e mais precisamente foi entendido que é necessário obter respostas as seguintes questões:

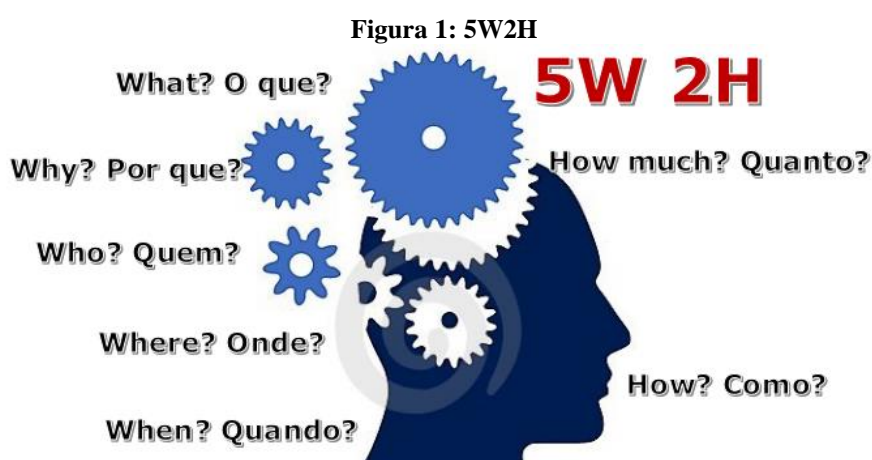
- **Quantidade Total NACIONAL** de alunos Ingressantes, Concluintes, Permanente, Desistentes, Falecidos e Indefinidos.
- **Diferença Proporcional Total NACIONAL** de evasão entre os cursos na Modalidade Online e Presencial por Quantidade de Alunos ingressantes.
- **Quais os 2 Cursos** na Modalidade Online e Presencial que apresenta maior taxa de evasão por: Estado, Região, entre as Privadas e Públicas.
- **Quais as 2 Instituições** na Modalidade Online e Presencial que apresenta maior taxa de evasão por: Estado, Região, entre as Privadas e Públicas.
- **Qual Área de Ensino** na Modalidade Online e Presencial que apresenta maior taxa de evasão por: Estado, Região, entre as Privadas e Públicas.

2.2 Plano de Ação 5W2H

O plano de ação 5W2H foi utilizado para definirmos quem são os agentes principais do estudo, que no nosso caso foi entendido como os **Cursos** que fazem parte das instituições.

Hipoteticamente criamos uma solicitação vinda do Ministério da Educação em parceria com o Ministério de Minas e Energia para remanejamento das verbas direcionada a formação de profissionais para a realidade atual do cenário nacional.

Essa ideia foi definida a fim de que o estudo tivesse um sentido e ficasse dentro de uma realidade para que pudéssemos construir as ideias e soluções do nosso projeto como um todo.



2.2.1 What? (O QUE?)

O projeto busca identificar os cursos de nível superior onde há maior índice de evasão.

2.2.2 Why? (POR QUÊ?)

Para que as verbas gastas com os cursos com o maior índice de evasão sejam direcionadas para ofertar mais vagas em cursos nos quais os índices de conclusão é notável. Acarretando uma possível reestruturação pedagógica desses cursos com alto índice de evasão, mudança no nome, alteração da grade de disciplinas, dentre outras medidas institucionais cabíveis.

2.2.3 Where? (ONDE?)

O projeto tem um âmbito nacional, porém a parceria com o Ministério de Minas e Energia é devido à demanda existente por profissionais para atuar no setor de Energia e Construção, principalmente na região Sudeste do País.

2.2.4 Who?(QUEM?)

O Ministério da Educação e o de Minas e Energia avaliam desenvolver projetos em parcerias com Organizações Acadêmicas Privadas, para ofertar cursos Presenciais, Híbridos ou Semipresenciais com bolsa estudantil.

No Ensino Superior Público, a consequência do projeto seria o aumento do número de vagas nos vestibulares para cursos com maior taxa de conclusão e os relacionados a Energia e Construção.

2.2.5 When? (QUANDO?)

O projeto precisa ficar pronto até antes de março de 2021, que é quando serão distribuídas as verbas pelo Governo Federal para os Ministérios.

2.2.6 How? (COMO?)

A demonstração dos Cursos com maior taxa de evasão será através de uma apresentação, utilizando Dashboards criados a partir dos dados coletados pelo INEP.

2.2.7 How Often? (COM QUAL FREQUÊNCIA)

A cada 6 meses haverá uma revisão por parte do INEP para ver a consistência do projeto indicando novas taxas para os cursos do ensino Superior.

3 Elaboração do Data Warehouse

A elaboração das características ideais do Data Warehouse será a fonte integradora das informações entre o fato do estudo de caso. Entendemos que são os Cursos pertencentes a instituições de ensino superior o fato do estudo.

A tecnologia da construção das bases de dados será utilizada com o intuito de servir para a camada de aplicação de visualização que será responsável por fornecer informação para a tomada de decisão na organização via Dashboards.

A tabela a abaixo demonstra como utilizamos de códigos de cores para definir quais seriam os atributos pertencentes a cada entidade dos DW em questão.

<p>Fazer a relação das colunas com os novos nomes no power Architect: Esta planilha, contem o nome das novas colunas para fazer referência direto com o power architect e pdi para criação do DW e popular o mesmo.</p>					
<p>OBS: QUANDO AS LINHAS ESTIVEREM MARCADAS, SIGNIFICA QUE AS COLUNAS JÁ FORAM DESTINADAS A ARQUITETURA RESPECTIVA NO PA.</p>					
Relação das cores das colunas respectivas a cada estensão da estrela que forma a arquitetura.					dim_codmunicipio (ONDE)
					dim_tempo (QUANDO)
					dim_codinstituicao (O QUE)
					dim_alunos (QUEM)
TABELA FATO			nk_codcurso		
sk_tempo			Colunas EXTRA criadas no PDI		
sk_codmunicipio			Colunas EXTRA criadas no PDI		
sk_alunos			Colunas EXTRA criadas no PDI		
sk_codinstituicao			Colunas EXTRA criadas no PDI		

Figura 2: Autoral (Construção arquitetural DW.1)

A tabela que segue visa demonstrar como os atributos foram definidos e estarão distribuído para criação do modelo de Arquitetura do DW. Acreditamos que os código de cores auxiliam também na utilização do **Power Architect**, para evitar a existência de erros no momento de preencher as características de cada atributo que serão base dos scripts de SQL, antes de serem enviados ao **Postgres** para criação da estrutura dos DWs.

COLUNAS ANTIGAS	COLUNAS NOVAS (POWER ARCHITECT)
Código da Instituição	nk_codinstituicao
Nome da Instituição	nome_instituicao
Categoria Administrativa	cat_adm
Organização Acadêmica	org_academica
Código do Curso de Graduação	nk_codcurso
Nome do Curso de Graduação	nome_curso
Código da Região Geográfica do Curso	cod_regiao
Código da Unidade Federativa do Curso	cod_estado
Código do Município do Curso	nk_codmunicipio
Grau Acadêmico	grau_academico
Nome da Grande Área do Curso segundo a classificação CINE BRASIL	area_curso_CINE
Modalidade de Ensino	mod_ensino
Ano de Referência	anoref
Prazo de Integralização em Anos	quant_anos_expec_termino
Ano de Integralização do Curso	ano_expec_termino
Prazo de Acompanhamento do Curso em anos	quant_anos_jubilo
Ano Máximo de Acompanhamento do Curso	ano_jubilo
Quantidade de Ingressantes no Curso	ingressantes
Quantidade de Permanência no Curso no ano de referência	permanece_ano_ref
Quantidade de Concluintes no Curso no ano de referência	conclui_ano_ref
Quantidade de Desistência no Curso no ano de referência	desiste_ano_ref
Quantidade de Falecimentos no Curso no ano de referência	falece_ano_ref
Taxa de Permanência - TAP	TAP
Taxa de Conclusão Acumulada - TCA	TCACU
Taxa de Desistência Acumulada - TODA	TDACU
Taxa de Conclusão Anual - TCAN	TCANUAL
Taxa de Desistência Anual - TDAN	TDANUAL

Figura 3: Autoral (Construção arquitetural DW.2)

3.1 Definição do *DW*

3.1.1 Arquitetura

A arquitetura definida foi a ***Independente***, pois é vinda de uma necessidade departamental, não tendo conectividade direta com todos os dados no âmbito do Ensino Superior, a base de dados se concentra na evasão por Curso e Instituição de Ensino.

3.1.2 Abordagem de Construção

Foi escolhido o modelo ***Top Down***.

3.1.3 Arquitetura Física

Apesar de o nosso modelo ter sido o ***On-Premises*** para criação do DW. Porém todas as informações como os scripts em SQL para construção da Arquitetura, alimentação do banco de dados de acordo com os atributos, estão salvos em um drive online, caracterizando a base de informações na nuvem.

4 Descrição do Modelo Multidimensional

Esta seção o modelo escolhido, **Estrela – Star Schema** – para construção do DW do estudo de caso Evasão no Ensino Superior.

A figura abaixo, demonstra como foi desenvolvido o esquema da arquitetura através do software **Power Architect**.

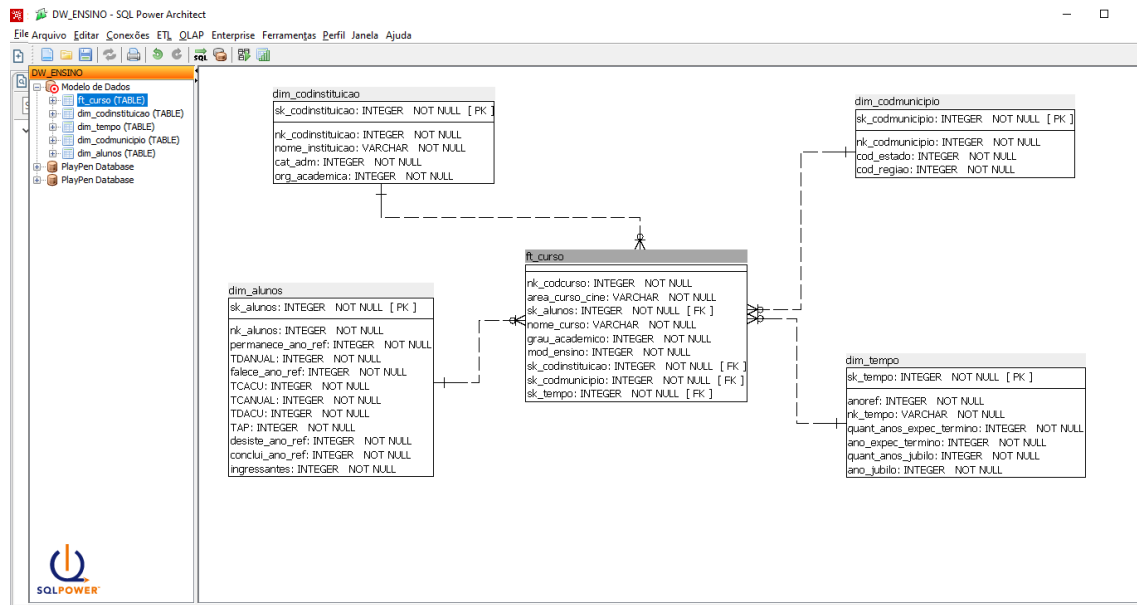


Figura 4: Construção modelo via Power Architect.

Já as imagens que seguem, relatam quais são os atributos pertencentes a cada entidade do modelo proposto.

É possível perceber que a descrição original dos atributos foi modificada para melhor se adequar no momento da criação do DW e posterior alimentação dos dados via PDI(Pentaho Data Integration).

A descrição de cada imagem abaixo indica os atributos de cada entidade, porém, é válido lembrar que todas as entidades possuem suas respectivas nk e sk que foram definidas dentro do Power Architect como incrementais.

- **dim_codinstituicao:** Código da Instituição / Área do Curso / Nome do Curso / Nome da Instituição / Categoria Administrativa / Organização Acadêmica

	A	B	C	D	E	F	G	H	I	J
	cod.instituicao	nome_instituicao	cat.adm	org_academica	cod.curso	nome_curso	cod_regiao	cod_estado	cod.municipio	grau_academico
1	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1
2	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1
3	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1
4	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1
5	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
6	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
7	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
8	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
9	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
10	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
11	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1
12	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1
13	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1
14	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1
15	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1
16	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1
17	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2
18	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2
19	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2
20	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2
21	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2
22	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1
23	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1
24	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1
25	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1
26	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1
27	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1
28	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1
29	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1
30	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1

Figura 5: Tabela Dimensional INSTITUIÇÃO

- **dim_alunos:** Quantidade de Ingressantes / Quantidade de Permanentes / Quantidade de Desistentes / Quantidade de Concluintes / Quantidade de Falecimento / Taxa de Permanência TAP / Taxa de Conclusão Acumulada – TCA / Taxa de Desistência Acumulada – TODA / Taxa de Conclusão Anual – TCAN / Taxa de Desistência Anual – TDAN

	D	E	F	G	H	I	J	K	L	M	N
	org_academica	cod.curso	nome_curso	cod_regiao	cod_estado	cod.municipio	grau_academico	mod_ensino	area_curso_CINE	ingressantes[2]	ano.ref
1	1	1	DIREITO	5	51	5103403	1	1	Negócios, administração e direito	2015	2015
2	1	1	DIREITO	5	51	5103403	1	1	Negócios, administração e direito	2015	2016
3	1	1	DIREITO	5	51	5103403	1	1	Negócios, administração e direito	2015	2017
4	1	1	DIREITO	5	51	5103403	1	1	Negócios, administração e direito	2015	2018
5	1	1	DIREITO	5	51	5103403	1	1	Negócios, administração e direito	2015	2019
6	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1	Ciências sociais, comunicação e informação	2015	2015
7	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1	Ciências sociais, comunicação e informação	2015	2016
8	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1	Ciências sociais, comunicação e informação	2015	2017
9	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1	Ciências sociais, comunicação e informação	2015	2018
10	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1	Ciências sociais, comunicação e informação	2015	2019
11	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1	Engenharia, produção e construção	2015	2015
12	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1	Engenharia, produção e construção	2015	2016
13	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1	Engenharia, produção e construção	2015	2017
14	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1	Engenharia, produção e construção	2015	2018
15	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1	Engenharia, produção e construção	2015	2019
16	1	6	PEDAGOGIA	5	51	5103403	2	1	Educação	2015	2015
17	1	6	PEDAGOGIA	5	51	5103403	2	1	Educação	2015	2016
18	1	6	PEDAGOGIA	5	51	5103403	2	1	Educação	2015	2017
19	1	6	PEDAGOGIA	5	51	5103403	2	1	Educação	2015	2018
20	1	6	PEDAGOGIA	5	51	5103403	2	1	Educação	2015	2019
21	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1	Saúde e bem-estar	2015	2015
22	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1	Saúde e bem-estar	2015	2016
23	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1	Saúde e bem-estar	2015	2017
24	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1	Saúde e bem-estar	2015	2018
25	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1	Saúde e bem-estar	2015	2019
26	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1	Negócios, administração e direito	2015	2015
27	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1	Negócios, administração e direito	2015	2016
28	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1	Negócios, administração e direito	2015	2017
29	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1	Negócios, administração e direito	2015	2018

Figura 6: Tabela Dimensional dim_alunos

- **dim_tempo:** Ano de Ingresso / Ano de Referência / Prazo de Integralização em Anos / Ano de Integralização do Curso / Prazo de Acompanhamento do Curso

N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
ano_ref	anos_expec_tec	ano_expec_term	quant_anos_jubilo	ano_jubilo	gressantes(2015)	manece_ano_ref	oncul_ano_ref	esiste_ano_ref	falece_ano_ref	TAP	TCACU	TDACU	TCANUAL	TDANUAL
2015	6	2020	10	2024	98	97	1	0	0	99,0	1,0	0,0	1,0	0,0
2016	6	2020	10	2024	98	95	0	2	0	96,9	1,0	2,0	0,0	2,0
2017	6	2020	10	2024	98	93	1	1	0	94,9	2,0	3,1	1,0	1,0
2018	6	2020	10	2024	98	85	7	1	0	86,7	9,2	4,1	7,1	1,0
2019	6	2020	10	2024	98	20	5,7	7	1	20,6	68,0	11,3	58,8	7,2
2015	5	2019	8	2022	111	110	0	1	0	99,1	0,0	0,9	0,0	0,9
2016	5	2019	8	2022	111	106	0	4	0	95,5	0,0	4,5	0,0	3,6
2017	5	2019	8	2022	111	100	1	5	0	90,1	0,9	9,0	0,9	4,5
2018	5	2019	8	2022	111	71	9	20	0	64,0	8,0	27,0	8,1	18,0
2019	5	2019	8	2022	111	43	11	17	0	38,7	18,9	42,3	9,9	15,3
2015	6	2020	10	2024	53	52	0	1	0	98,1	0,0	1,9	0,0	1,9
2016	6	2020	10	2024	53	51	0	1	0	96,7	0,0	1,8	0,0	1,8
2017	6	2020	10	2024	53	50	0	1	0	94,3	0,0	5,7	0,0	1,9
2018	6	2020	10	2024	53	47	1	2	0	88,7	1,9	9,4	1,9	3,8
2019	6	2020	10	2024	53	19	2,2	6	0	35,8	43,4	20,8	41,5	11,3
2015	5	2019	8	2022	107	105	0	2	0	98,1	0,0	1,9	0,0	1,9
2016	5	2019	8	2022	107	101	1	3	0	94,4	0,9	4,7	0,9	2,8
2017	5	2019	8	2022	107	99	1	1	0	92,5	1,9	5,0	0,9	0,9
2018	5	2019	8	2022	107	57	4,1	1	0	53,3	40,2	6,5	38,3	0,9
2019	5	2019	8	2022	107	32	4	21	0	29,9	43,9	26,2	3,7	19,6
2015	5	2019	8	2022	81	77	0	4	0	95,1	0,0	4,9	0,0	4,9
2016	5	2019	8	2022	81	72	0	5	0	88,9	0,0	11,1	0,0	6,2
2017	5	2019	8	2022	81	69	0	3	0	85,2	0,0	14,8	0,0	3,7
2018	5	2019	8	2022	81	47	0	22	0	58,0	0,0	42,0	0,0	27,2
2019	5	2019	8	2022	81	29	1	13	0	35,8	13,6	59,6	13,6	8,6
2015	5	2019	8	2022	86	84	0	2	0	97,7	0,0	2,3	0,0	2,3
2016	5	2019	8	2022	86	79	2	3	0	91,9	2,3	5,8	2,3	3,5
2017	5	2019	8	2022	86	78	1	1	0	90,7	3,5	5,8	1,2	0,0
2018	5	2019	8	2022	86	44	3,4	0	0	51,2	43,0	5,8	39,5	0,0

Figura 7: Tabela Dimensional dim_tempo

- **dim_cod.municipio:** Código da Região Geográfica do Curso / Código da Unidade Federativa do Curso / Código do Município do Curso

A	B	C	D	E	F	G	H	I	J	K
	nome_instituicao	catadm	org_academica	cod_curso	nome_curso	cod_regiao	cod_estado	cod_municipio	grau_academico	mod_ensino
1	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
2	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
3	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
4	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
5	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
6	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	1	DIREITO	5	51	5103403	1	1
7	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1
8	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1
9	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1
10	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1
11	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	2	CIÊNCIAS ECONÔMICAS	5	51	5103403	1	1
12	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1
13	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1
14	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1
15	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1
16	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	3	ENGENHARIA CIVIL	5	51	5103403	1	1
17	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2	1
18	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2	1
19	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2	1
20	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2	1
21	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	6	PEDAGOGIA	5	51	5103403	2	1
22	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1
23	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1
24	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1
25	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1
26	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	7	SERVIÇO SOCIAL	5	51	5103403	1	1
27	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1
28	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1
29	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1
30	UNIVERSIDADE FEDERAL DE MATO GROSSO	1	1	8	CIÊNCIAS CONTÁBEIS	5	51	5103403	1	1

Figura 8: Tabela Dimensional dim_codmunicipio

- **ft.curso:** Modalidade de Ensino / Grau Acadêmico / Nome do Curso de Graduação / Área do Curso CINE / Código do Curso e as 4 chaves estrangeiras as outras entidades.

	L	M	N	O	P	Q	R
	area_curso_CINE	ano_ingressantes(2015)	ano_ref	quant_anos_expec_termino	ano_expec_termino	quant_anos_jubilo	ano_jubilo
1	Negócios, administração e direito	2015	2015	6	2020	10	2024
2	Negócios, administração e direito	2015	2016	6	2020	10	2024
3	Negócios, administração e direito	2015	2017	6	2020	10	2024
4	Negócios, administração e direito	2015	2018	6	2020	10	2024
5	Negócios, administração e direito	2015	2019	6	2020	10	2024
6	Negócios, administração e direito	2015	2015	5	2019	8	2022
7	Ciências sociais, comunicação e informação	2015	2016	5	2019	8	2022
8	Ciências sociais, comunicação e informação	2015	2017	5	2019	8	2022
9	Ciências sociais, comunicação e informação	2015	2018	5	2019	8	2022
10	Ciências sociais, comunicação e informação	2015	2019	5	2019	8	2022
11	Engenharia, produção e construção	2015	2015	6	2020	10	2024
12	Engenharia, produção e construção	2015	2016	6	2020	10	2024
13	Engenharia, produção e construção	2015	2017	6	2020	10	2024
14	Engenharia, produção e construção	2015	2018	6	2020	10	2024
15	Engenharia, produção e construção	2015	2019	6	2020	10	2024
16	Educação	2015	2015	5	2019	8	2022
17	Educação	2015	2016	5	2019	8	2022
18	Educação	2015	2017	5	2019	8	2022
19	Educação	2015	2018	5	2019	8	2022
20	Educação	2015	2019	5	2019	8	2022
21	Saúde e bem-estar	2015	2015	5	2019	8	2022
22	Saúde e bem-estar	2015	2016	5	2019	8	2022
23	Saúde e bem-estar	2015	2017	5	2019	8	2022
24	Saúde e bem-estar	2015	2018	5	2019	8	2022
25	Saúde e bem-estar	2015	2019	5	2019	8	2022
26	Negócios, administração e direito	2015	2015	5	2019	8	2022
27	Negócios, administração e direito	2015	2016	5	2019	8	2022
28	Negócios, administração e direito	2015	2017	5	2019	8	2022
29	Negócios, administração e direito	2015	2018	5	2019	8	2022
30	Negócios, administração e direito	2015	2019	5	2019	8	2022

Figura 9: Tabela Dimensional ft_curso

5 Proposta de Processo de BI

5.1 Etapas para construção da estrutura do projeto de BI.

PARTE 1 – CRIAÇÃO d ideia do DW, definindo suas características.

- Decidir quais serão os atributos pertencentes a cada entidade no Data Warehouse, filtrar os atributos por pertinência e em seguida, na base de dados FONTE em Excel preparar e armazenar uma STAGE_AREA no Postgres para segurança do Projeto
- A partir desse momento, a arquitetura do Data Warehouse será desenhada através do Power Architect e via script SQL gerado pelo próprio P.A, a estrutura do DW será desenhada diretamente no Postgres via conexão entre os dois softwares.

PARTE 2 – DESENVOLVIMENTO DAS OPERAÇÕES DE ETL e TRANSFERÊNCIA DOS DADOS TRATADOS PARA A ESTRUTURA DE DW CRIADA.

- Um vez que a arquitetura do Data Warehouse já esteja presente no Postgres, é através do Pentaho Data Integration (PDI) que serão feitas algumas operações de ETL e também a alimentação do DW com os dados oriundos da base separada e que já passou por algumas transformações iniciais.

PARTE 3 – CARGA DOS DADOS ARMAZENADOS NO DW PARA ELABORAÇÃO DOS DASHBOARDS VIA POWER BI.

- Confirmada a carga dos dados no Postgres, é o momento de executar a construção das visualizações via Power BI e finalizar a proposta do projeto, demonstrando através dos diversos recursos da ferramenta como o dados podem gerar, informação e conhecimento.

OBSERVAÇÕES IMPORTANTES NO ESTUDO DE CASO

A grande necessidade de se fazer a correta estruturação do DW, é que na modalidade online dispensa o viés localização, pois a base de dados está relacionada diretamente com o fluxo por instituições e não quem são os indivíduos que fazem parte do fluxo, ou seja, os alunos, a base de dados referida faz sim uma relação direta com quais são as instituições que apresentam as taxas e números de INGRESSANTES, PERMANENTES, CONCLUINTES, DESISTENTES e FALECIDOS.

Um outro ponto importante no estudo, é fazer a comparação proporcional dos estudantes que fluem nos cursos atentando para a modalidade de ensino, já que quando essa é a distância, geograficamente não se tem uma referência na base de dados, essas estão cadastradas como online.

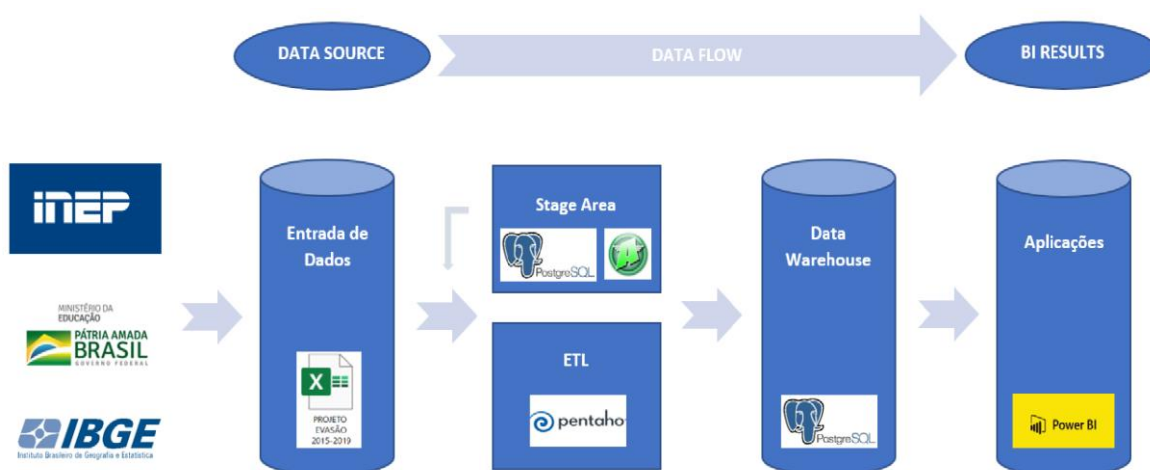


Figura 10: Figura – Processo de elaboração de Projeto de BI.

6 Projeto de ETL

6.1 Descrição do Projeto de ETL

As informações contidas no quadro a baixo são referentes a todas as etapas de ETL desenvolvidas durante todo o projeto.

Parte – 1

Primeira exploração visual da base de dados via Excel e execução do primeiro tratamento da Base, onde foram excluídas linhas desnecessárias e que provavelmente iriam dificultar a construção da base de dados no Postgres.

Parte – 2

Exclusão de diversos atributos da base original a fim de se ter uma melhor precisão na entrega dos resultados através da visualização em Dashboards.

Parte – 3

Utilização do PDI para alterar o formato de alguns atributos, conectar o PDI com o Postgres.

Parte – 4

Utilização do PDI para fazer a inserção dos dados dentro da estrutura DW_ENSINO armazenada no SGBD Postgres.

Parte – 5

(Power BI): Normalização, mudança de formato e categorização – Obs.: A maioria das tarefas de ETL foram realizadas através do Power Query no Power BI.

As imagens que seguem são referentes as operações desenvolvidas no Pentaho Data Integration para transferência da base de dados que já recebeu iniciais tratamentos (ETL), para em seguida ser armazenada no Postgres.

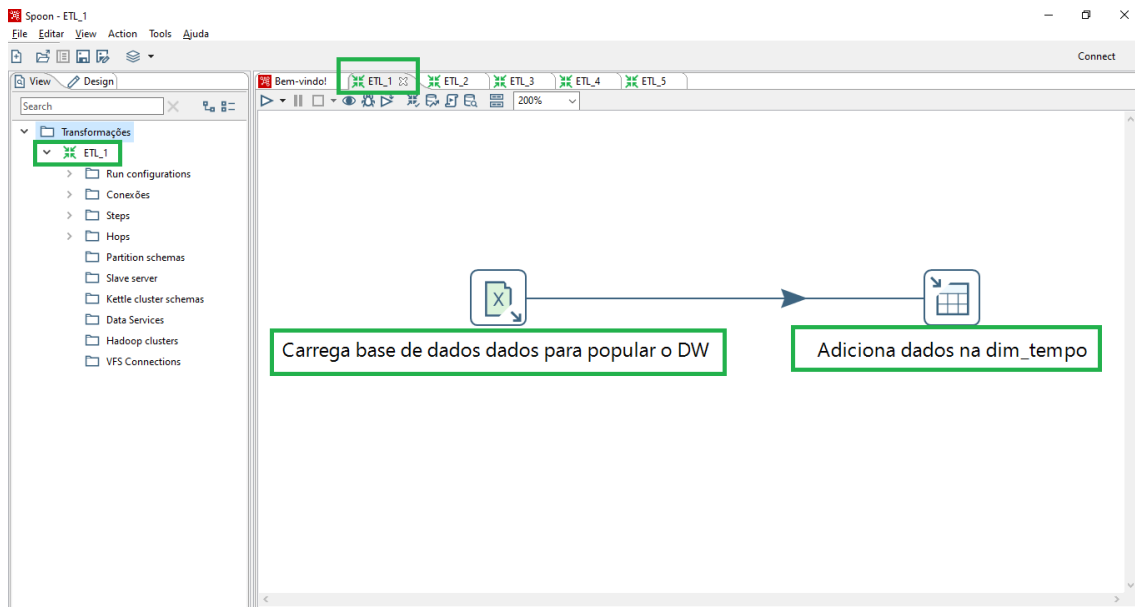


Figura 11: ETL 1 populando DW_ENSINO (dim_tempo)

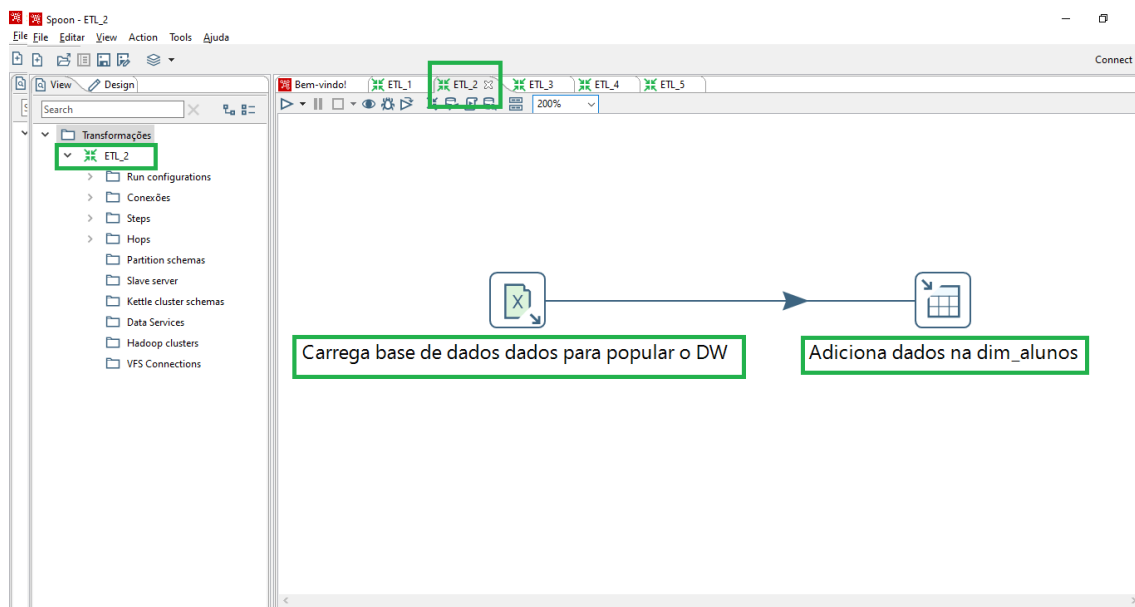


Figura 12: ETL 2 populando DW_ENSINO (dim_alunos)

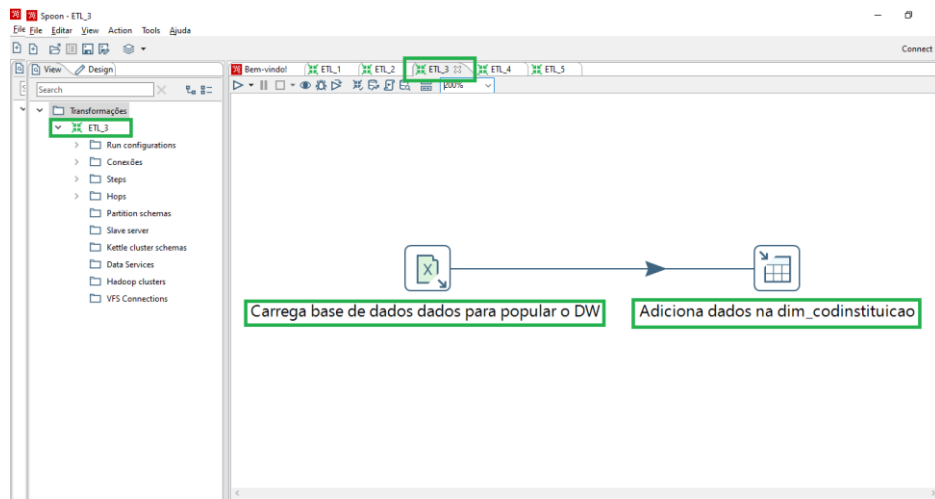


Figura 13: ETL 3 populando DW_ENSINO (dim_codinstituicao)

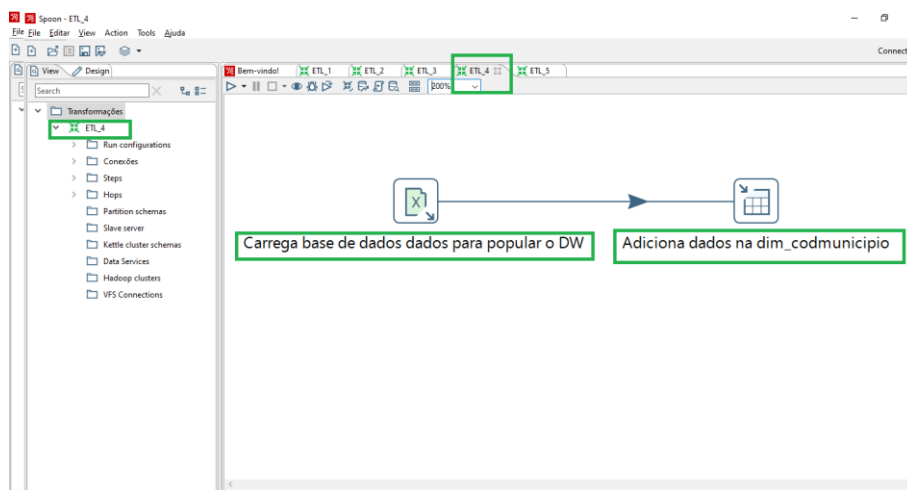


Figura 14: ETL 4 populando DW_ENSINO (dim_codmunicipio)

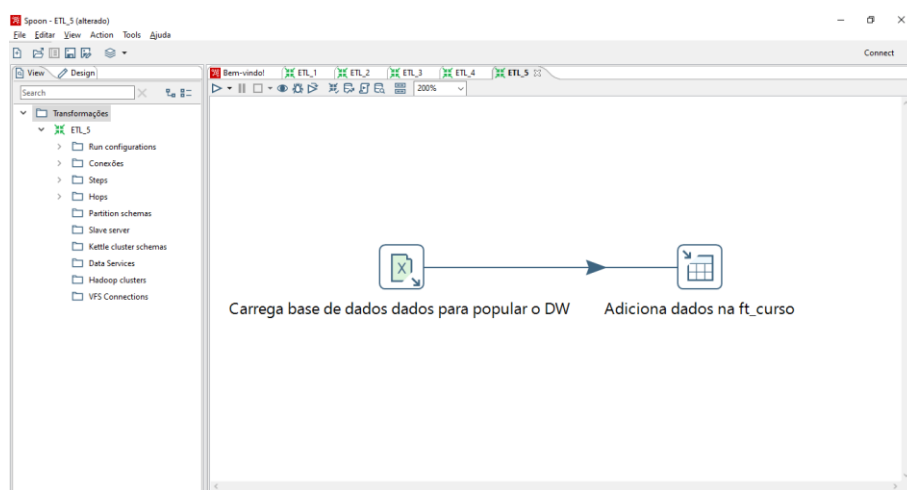


Figura 15: ETL 5 populando DW_ENSINO (ft_curso)

OBSERVAÇÕES IMPORTANTES NO ESTUDO DE CASO

As etapas demonstradas anteriormente através das imagens foram desenvolvidas para a inserir os dados no DW_ENSINO que estavam em formato XLSX e que foram coletados no site do INEP.

As imagens que seguem demonstram o layout do DW_ENSINO que foi criado pelo Power Architect e Populado através do PDI.

The screenshot displays the pgAdmin 4 web interface. On the left, the 'Servers' tree shows 'PostgreSQL 12' expanded, with 'Databases (11)' and 'DW_ENSINO' highlighted. Under 'DW_ENSINO', the 'Tables (5)' folder is expanded, showing 'dim_alunos', 'dim_codinstituicao', 'dim_codmunicipio', 'dim_tempo', and 'ft_curso'. The 'Query Editor' on the right contains the following SQL query:

```
1 SELECT * FROM public.dim_alunos
2 ORDER BY sk_alunos ASC
```

Below the query editor, the 'Data Output' tab shows the results of the query. The table has 5 columns: 'sk_alunos [PK] integer', 'nk_alunos integer', 'permanece_ano_ref integer', 'tdanual integer', and 'falece_ano integer'. The results are displayed in a table with 30 rows.

sk_alunos [PK] integer	nk_alunos integer	permanece_ano_ref integer	tdanual integer	falece_ano integer
1	1	1	97	0
2	2	2	95	2
3	3	3	93	1
4	4	4	85	1
5	5	5	20	7
6	6	6	110	1
7	7	7	106	4
8	8	8	100	5
9	9	9	71	18
10	10	10	43	15
11	11	11	52	2
12	12	12	51	2
13	13	13	50	2
14	14	14	47	4
15	15	15	19	11
16	16	16	105	2
17	17	17	101	3
18	18	18	99	1
19	19	19	57	1
20	20	20	32	20
21	21	21	77	5
22	22	22	72	6
23	23	23	69	4
24	24	24	47	27
25	25	25	29	9
26	26	26	84	2
27	27	27	79	3
28	28	28	78	0
29	29	29	44	0
30	30	30	25	17

Figura 16: DW_ONLINE no Postgres

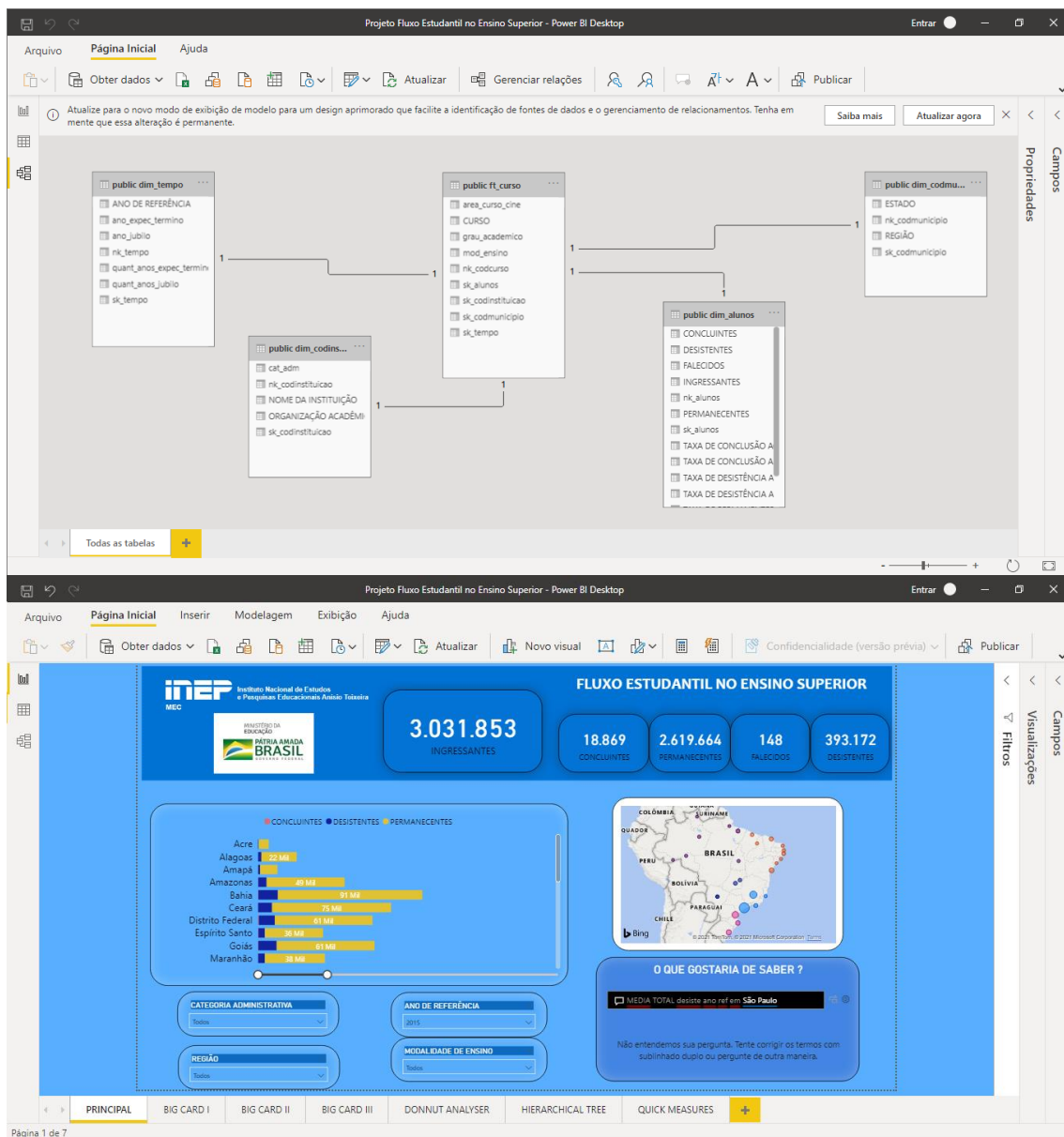
7 Dashboard

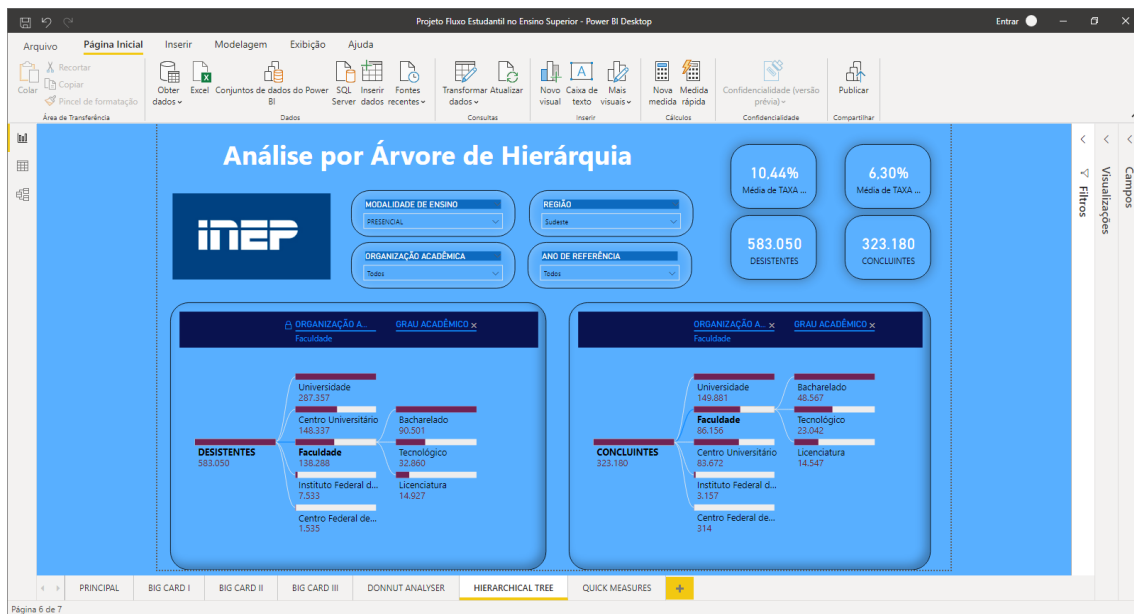
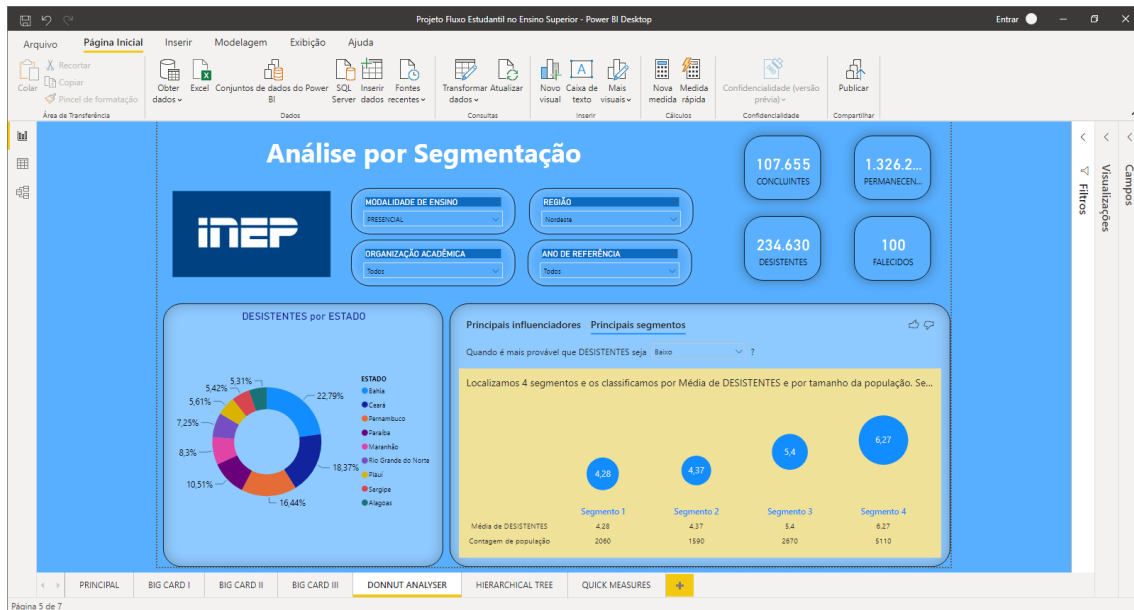
7.1 Descrição da Elaboração

A ferramenta escolhida para criação foi o Power BI e no mesmo, além de criar as telas, foram feitas algumas operações de ETL para categorização via Power Query.

Os dados para construção dos Dashboards foram carregados diretamente do SGBD, no qual continha dois DWs para cada Modalidade de Ensino do nosso estudo de caso.

7.2 Telas do Dashboard





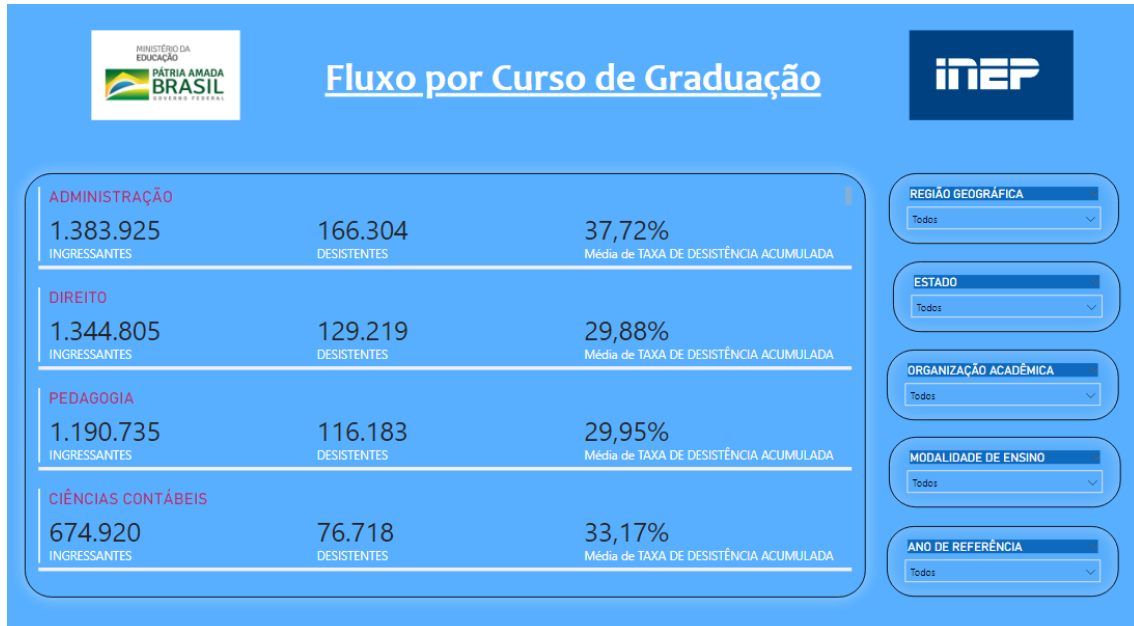
8 Conclusão

Como Conclusão do Projeto, acreditamos que a partir da possibilidade de demonstrar as respostas as questões expressas na introdução desse trabalho através das visualizações dos painéis via Power BI que criamos, é a consolidação do poder dessa ferramenta para distribuir informação e conhecimento dentro de qualquer organização e em qualquer âmbito social de produção de serviço e produtos.

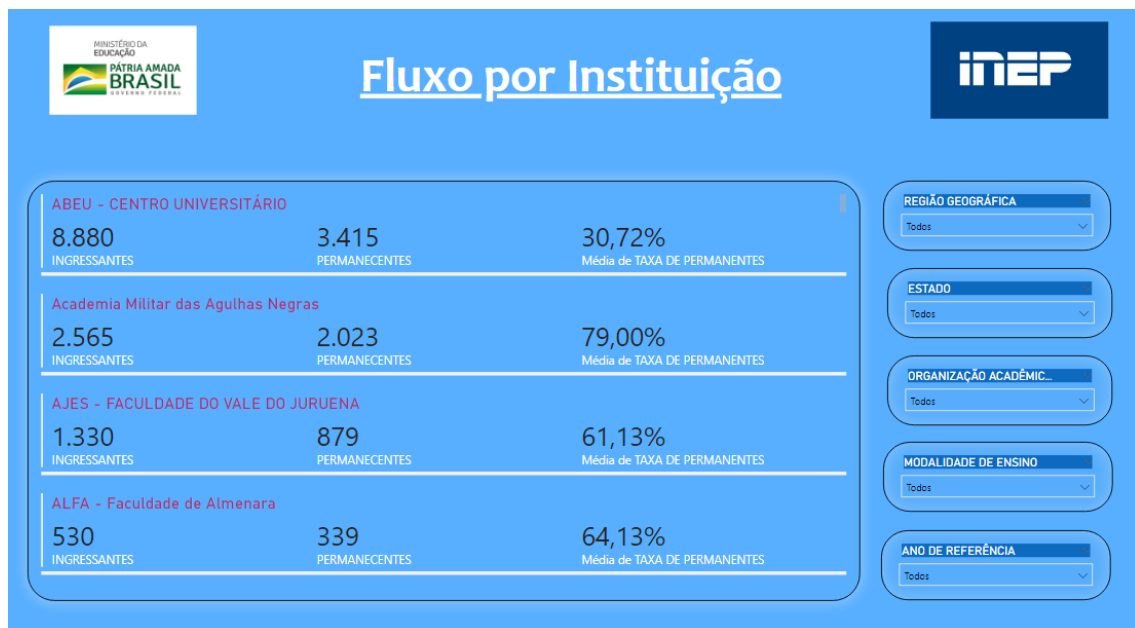
- **Quantidade Total NACIONAL** de alunos Ingressantes, Concluintes, Permanente, Desistentes, Falecidos e Indefinidos.



- **Quais os 2 Cursos** na Modalidade Online e Presencial que apresenta maior taxa de evasão (Desistentes) por: Estado, Região, entre as Privadas e Públicas.



- **Quais as 2 Instituições** na Modalidade Online e Presencial que apresenta maior taxa de permanentes por: Estado, Região, entre as Privadas e Públicas.



- **Qual Área de Ensino** na Modalidade Online e Presencial que apresenta maior taxa de concluintes por: Estado, Região, entre as Privadas e Públicas.

