

UNIVERSIDADE DE BRASÍLIA  
Faculdade do Gama

Sistemas de Banco de Dados 2

**Trabalho Final (TF)**

**ETL**

Matheus Gabriel Alves Rodrigues 18/0106970

Durval Carvalho de Souza 16/0005191

Brasília, DF

2021

## DESCRIÇÃO

Nosso projeto possui como tema ETL (Extract Transform Load) e escolhemos como base de dados os dados para o trabalho final a base do ENEM (2019), trabalhada pelo membro Matheus Gabriel anteriormente, em que existem dados referentes ao desempenho de candidatos do exame neste ano, essa base foi escolhida pelo fato de ser bastante robusta e possuir dados que possibilitam a criação de consultas e análises por parte dos membros da dupla.

Na parte prática do trabalho foi criada uma pipeline de dados utilizando debezium, kafka, elasticsearch e kibana; mais explicações na apresentação e no repositório

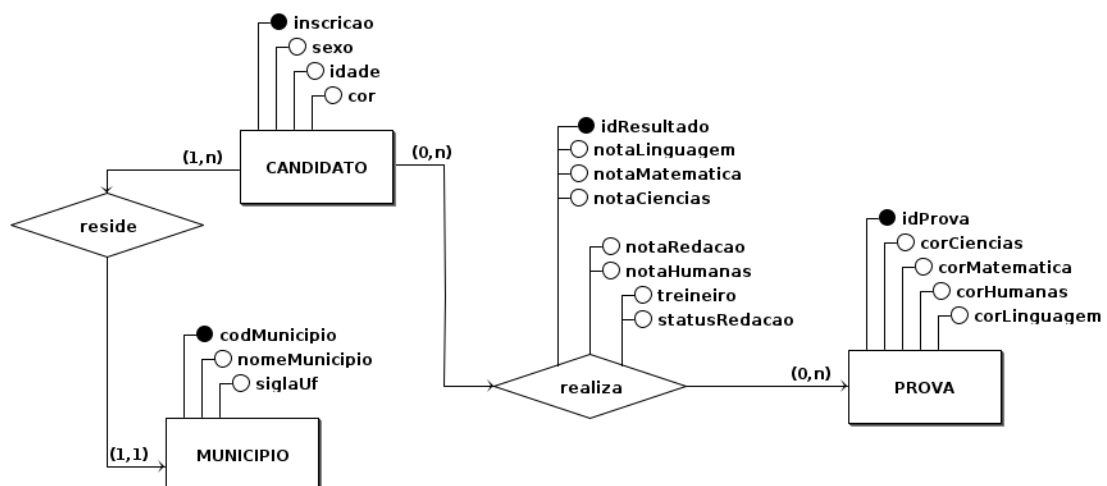
### Prática:

<https://github.com/durvalcarvalho/bd2-etl>

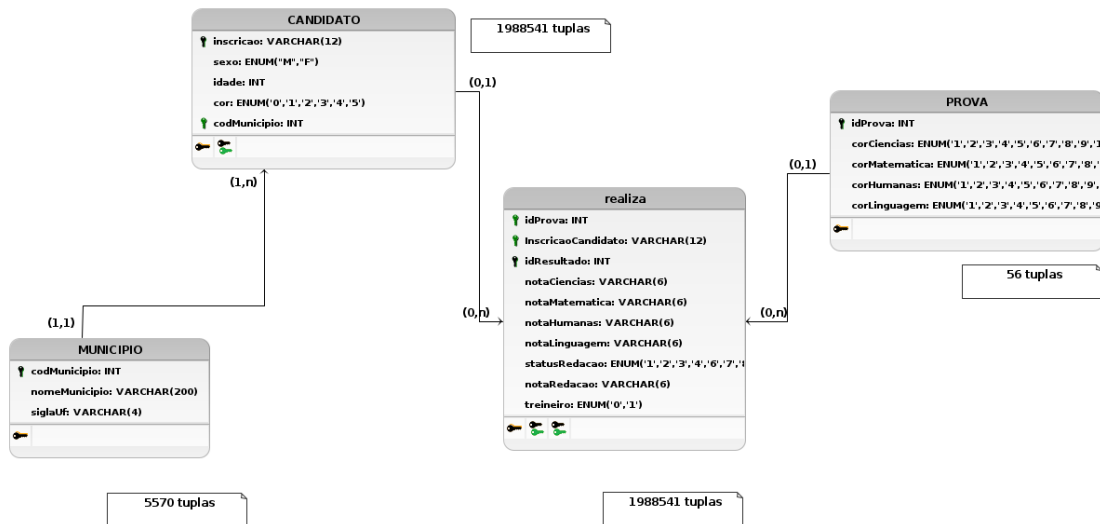
### Fonte:

<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>

## DER



## DLD



## BASE DE DADOS

Nossa dupla decidiu optar pela escolha da base de dados do ENEM (2019) por ser uma base de dados que possui maior quantidade de dados, possibilitando também uma maior quantidade de colunas para análise da dupla.

## DICIONÁRIO DE DADOS

Entidade: MUNICIPIO				
<b>Descrição:</b> Responsável por armazenar os municípios da qual os candidatos residem.				
Atributo	Propriedade atributo	Tipo de dado	Tamanho	Descrição
codMunicipio	Obrigatório, chave primária	INT	-	Código do município
nomeMunicipio	não nulo	VARCHAR	200	nome do município
siglaUf	não nulo	VARCHAR	4	Sigla do estado que esse município faz parte

Entidade: CANDIDATO				
<b>Descrição:</b> Responsável por armazenar os candidatos que realizaram as provas.				
Atributo	Propriedade atributo	Tipo de dado	Tamanho	Descrição
inscricao	Obrigatório, chave primária	VARCHAR(12)	-	Inscrição do candidato na prova
sexo	não nulo	ENUM	("M","F")	Sexo no qual o candidato se identifica
idade	-	INT	-	Idade do candidato
cor	-	ENUM	('0','1','2','3','4','5')	Código da cor na qual o candidato se identifica
codMunicipio	Não nulo, Chave estrangeira	INT	-	Código do município em que o candidato

				reside
--	--	--	--	--------

<b>Entidade: PROVA</b>				
<b>Descrição:</b> Responsável por armazenar os tipos de provas possíveis de serem pegos.				
Atributo	Propriedade atributo	Tipo de dado	Tamanho	Descrição
idProva	Obrigatório, chave primária	INT	-	Índice da tabela
corCiencias	Unico	ENUM	('1','2','3','4','5','6','7','8','9','10','11','12','13')	Código da Cor da prova de ciencias naturais
corMatematica	Unico	ENUM	('1','2','3','4','5','6','7','8','9','10','11','12','13')	Código da Cor da prova de Matemática
corHumanas	Unico	ENUM	('1','2','3','4','5','6','7','8','9','10','11','12','13'))	Código da Cor da prova de Humanas
corLinguagem	Unico	ENUM	('1','2','3','4','5','6','7','8','9','10','11','12','13')	Código da Cor da prova de Linguagem

<b>Entidade: realiza</b>				
<b>Descrição:</b> Tabela de relacionamento responsável por armazenar os resultados dos candidatos em uma prova específica.				
Atributo	Propriedade atributo	Tipo de dado	Tamanho	Descrição
idResultado	Obrigatório, chave primária	INT		Código do município
inscricaoCandidato	Chave estrangeira, não nulo	INT	-	inscrição do candidato

idProva	Chave estrangeira, não nulo	INT	-	Id da prova feita
notaCiencias	-	VARCHAR(6)	-	Nota na parte de ciencias
notaMatematica	-	VARCHAR(6)	-	Nota na parte de matematica
notaHumanas	-	VARCHAR(6)	-	Nota na parte de Humanas
notaLinguagem	-	VARCHAR(6)	-	Nota na parte de Linguagem
statusRedacao	-	ENUM	('1','2','3','4','6','7','8','9')	Código do Status da redação realizada pelo candidato
notaRedacao	-	VARCHAR(6)	-	nota obtida na redação
treineiro	não nulo	ENUM	('0','1')	Se candidato é ou não treineiro

Visão: CANDIDATOS_DF				
Descrição: Visão que unifica dados relevantes sobre os candidatos do distrito federal.				
Atributo	Propriedade atributo	Tipo de dado	Tamanho	Descrição
inscricaoCandidato	Chave estrangeira, não nulo	INT	-	inscrição do candidato
notaCiencias	-	VARCHAR(6)	-	Nota na parte de ciencias
notaMatematica	-	VARCHAR(6)	-	Nota na parte de matematica

notaHumanas	-	VARCHAR(6)	-	Nota na parte de Humanas
notaLinguagem	-	VARCHAR(6)	-	Nota na parte de Linguagem
notaRedacao	-	VARCHAR(6)	-	nota obtida na redação
sexo	não nulo	ENUM	("M","F")	Sexo no qual o candidato se identifica