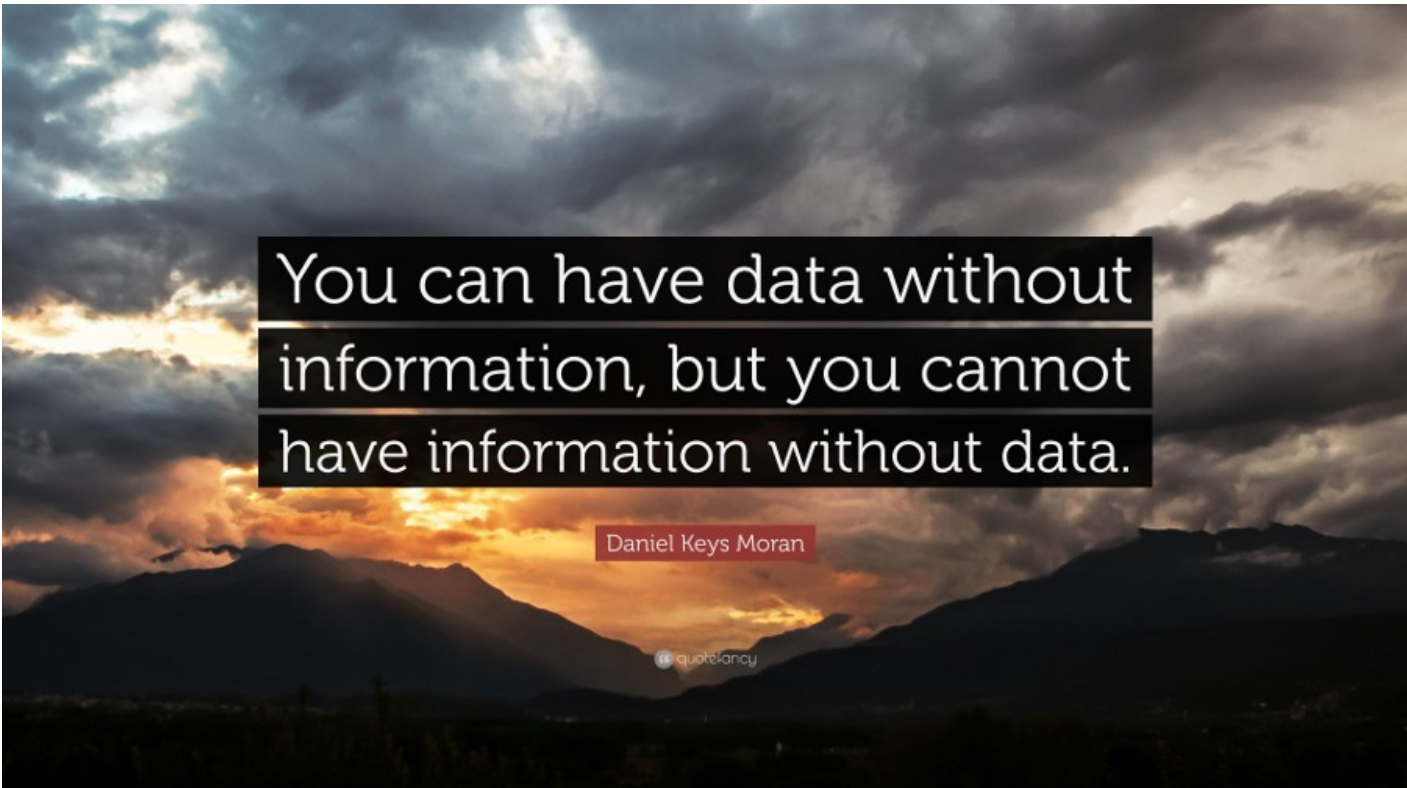


PROJETO PRÁTICO BIMESTRAL
PROJETO DE SOFTWARE DE DATAWAREHOUSE

“Quando a organização acumula informações e as registra em lugares diferentes com objetivos distintos, acaba tornando a análise confusa e escondendo dados muito relevantes para o seu negócio.”



You can have data without
information, but you cannot
have information without data.

Daniel Keys Moran

quoteancy

DATA DE ENTREGA E APRESENTAÇÃO: 06/12/2021
PONTUAÇÃO: 35 PONTOS

CASE:

Considere uma organização da área de educação que possui 2 sistemas distintos:

- **Sistema Acadêmico:** contém os dados dos alunos vinculados aos seus respectivos cursos e suas situações nestes cursos.
- **Sistema de Vestibular:** contém os dados dos candidatos interessados em participar dos processos seletivos para cursos da instituição.

Problema:

Em uma auditoria governamental, foram solicitados os seguintes dados:

- Quantitativos de **alunos** por:
 - Etnia
 - Gênero
 - Escola Origem do Aluno (Rede Pública / Rede Particular)
 - Renda Familiar
 - Naturalidade (Estado - Cidade)
 - Faixa Etária
 - Situação de Matrícula

A equipe de TI se deparou com o seguinte problema: No **sistema acadêmico** não há a informação de **Etnia**, nem **da Escola de Origem** e nem de **Renda Familiar** de cada aluno. No entanto, todo aluno teve que passar pelo sistema de vestibular. No sistema de vestibular há um questionário dinâmico em que umas das questões a serem respondidas são estas.

Solução:

A organização, visando obter os dados estatísticos em uma plataforma consolidada no futuro, optou por construir uma solução de **Datawarehouse** que inicialmente **integre** os sistemas acadêmicos e de vestibular para atender esta demanda específica e possibilitar possível escalabilidade para integração de mais sistemas.

O QUE DEVE SER APRESENTADO PELO ALUNO NESTE TRABALHO:

O objetivo final do trabalho é entregar uma plataforma WEB que apresente o referido indicador pelas dimensões especificadas. Possível exemplo de apresentação dos dados:

	Etnia Branca	Etnia Preta	Etnia Parda	Etnia Indígena	Etnia Amarela	Não declarada
Quantidade de alunos						

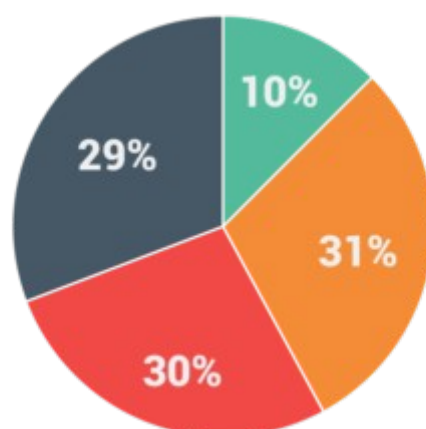
	Rede Pública	Rede Particular	Não Informado
Quantidade de alunos			

	Sexo Feminino	Sexo Masculino	Não Informado
Quantidade de alunos			

... Demais Dimensões...

Possibilidade de cruzamento de dimensões e especificação de filtros para cada dimensão.

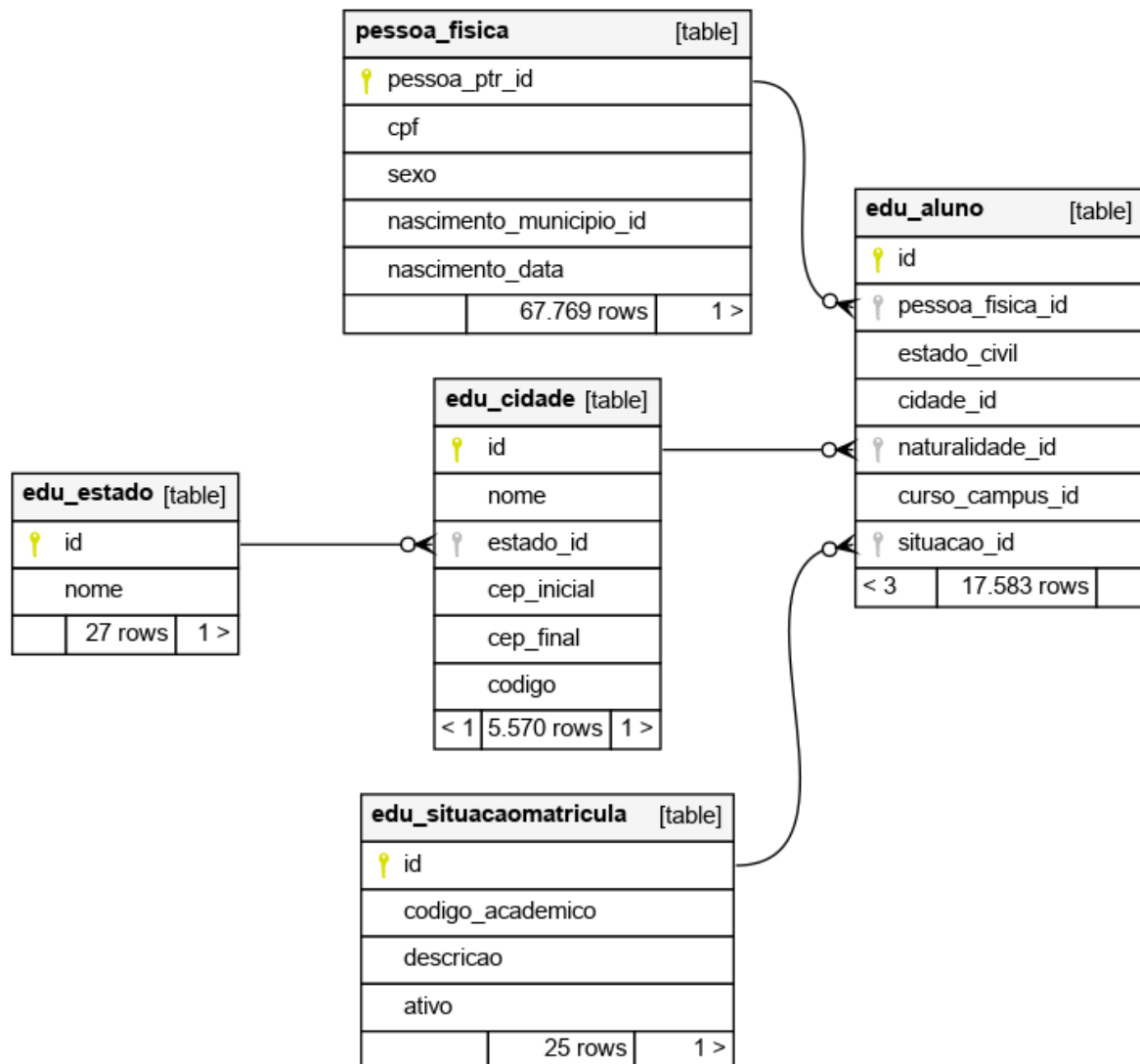
Devem ser apresentados gráficos para melhor análise e interpretação dos dados.



PROCESSO DE TRABALHO

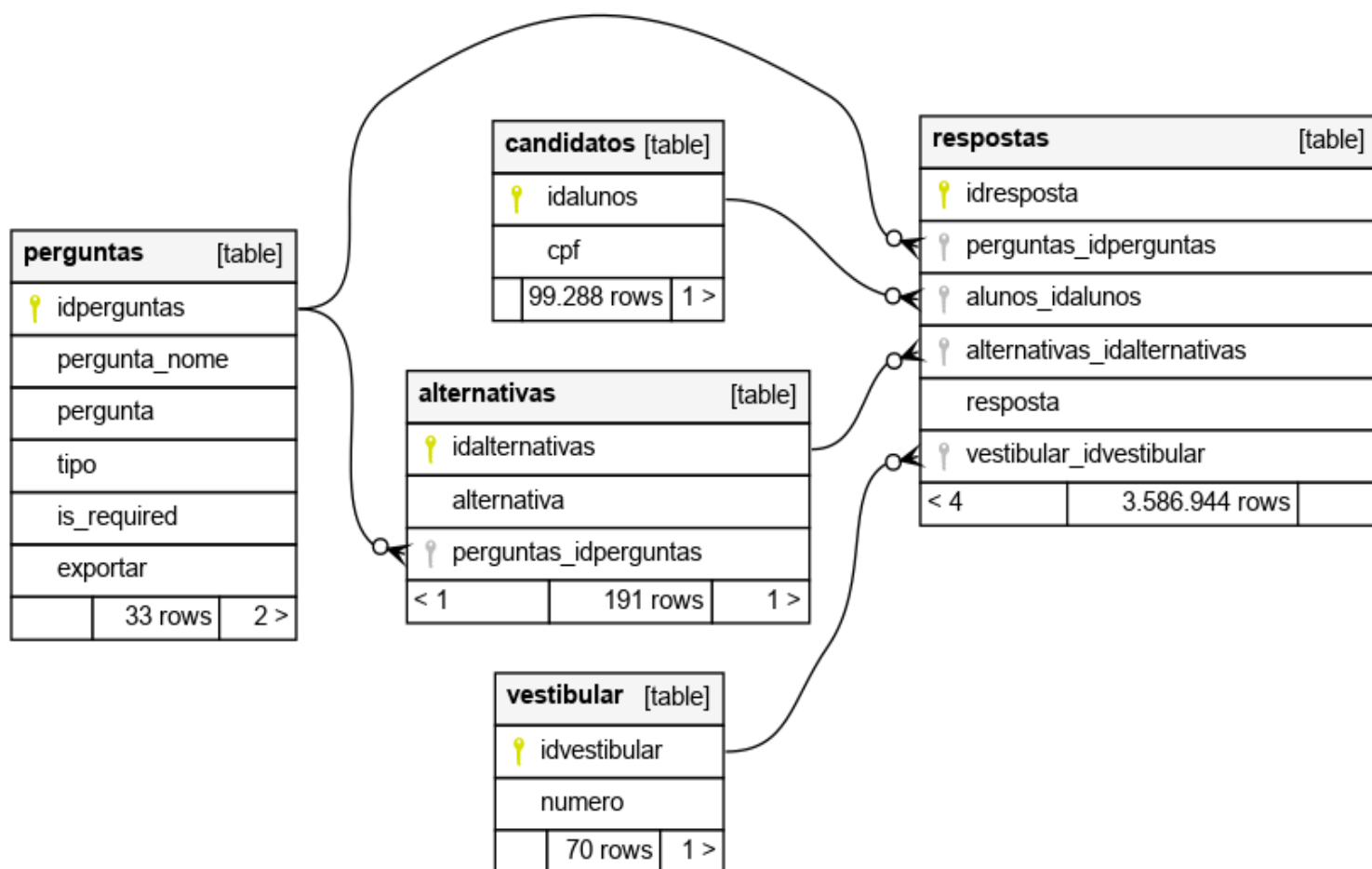
1) Primeiramente, cada base de dados deve ser estudada detalhadamente.

ACADÊMICO: A base de dados do sistema acadêmico é em PostgreSQL e o modelo de dados compacto está a seguir:



Generated by SchemaSpy

VESTIBULAR: A base de dados do sistema de vestibular é em MySQL e o modelo de dados compacto está a seguir:



Generated by SchemaSpy

Base de dados Acadêmica:

- A tabela **edu_aluno** contém os dados de alunos. Existe referência para a tabela **pessoa_fisica**.
- A tabela **pessoa_fisica** contém o gênero do aluno, CPF do aluno e data de nascimento.

Base de dados de Vestibular:

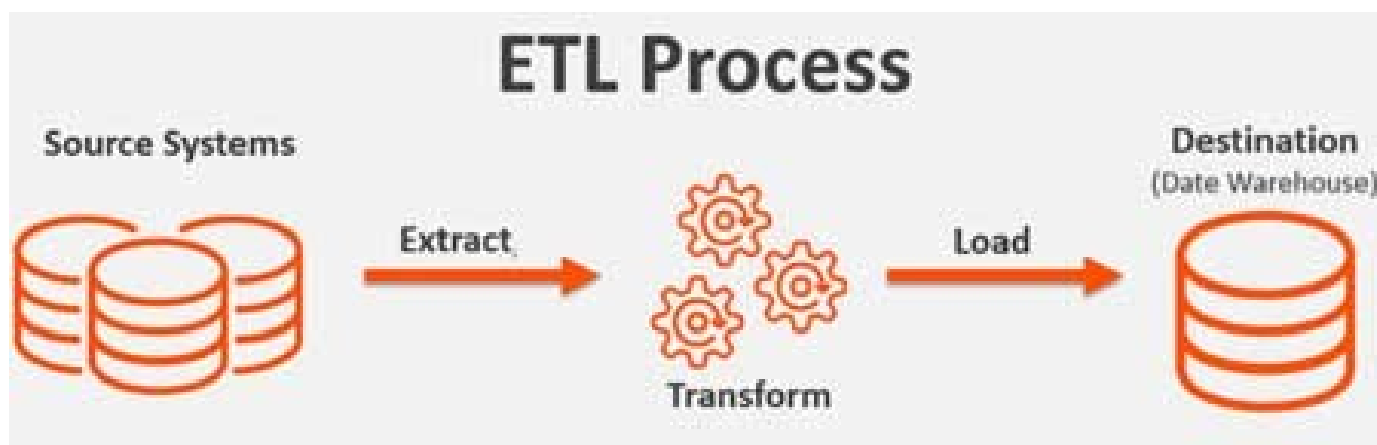
- A tabela **candidatos** possui o CPF do candidato.
- Para cada vestibular que um candidato participa, ele responde a um questionário socioeconômico. Neste questionário, o candidato responde a uma pergunta de etnia e de escola de origem. A tabela respostas armazena as respostas dos candidatos. As tabelas envolvidas são **respostas**, **alternativas**, **perguntas** e **candidatos**.
 - Pergunta de **Etnia**: ID 31
 - Pergunta de **Escola Origem**: ID 3
 - Perguntas de **Renda Familiar**: Ids 4, 33 e 34.

2) Modelar as tabelas do Banco de Dados do DATAWAREHOUSE.

3) Definição da arquitetura de sistema e escolha das tecnologias:

- Escolha do SGBD do Banco de Dados do Datawarehouse;
- Definição da arquitetura de sistema:
 - ETLs ficarão acoplados à aplicação de DW ou não?
 - Ferramentas de visualização dos dados do DW serão acopladas aos ETLs e à aplicação principal do DW ou não?
- Escolha da tecnologia para desenvolvimento do ETL e aplicação de conexão e visualização de dados do DW:
 - Java, Python, PHP, Ruby, etc...

4) Desenvolvimento.



Etapas do desenvolvimento:

- Após modelado o banco de dados do Datawarehouse, deve-se partir para o desenvolvimento do **Módulo ETL e posteriormente para o módulo WEB do relatório a ser exibido.**

MÓDULO ETL

- O módulo ETL da aplicação será responsável por extrair os dados das bases origem.
- Como as bases de dados estão disponíveis, é possível que o módulo se conecte direto na base de dados e efetue **queries** diretamente nelas.
- Os dados de alunos como curso/câmpus e sexo poderão ser extraídos da base acadêmica, transformados se necessários e injetados pelo ETL na base do Datawarehouse.
- **OS DADOS DE ETNIA, ESCOLA ORIGEM E RENDA FAMILIAR DEVERÃO SER MINERADOS E EXTRAÍDOS DA BASE DE VESTIBULAR, TRANSFORMADOS E PREPARADOS PARA SEREM INJETADOS NOS ALUNOS MIGRADOS DO SISTEMA ACADÊMICO.**

MÓDULO WEB

- O módulo WEB da aplicação será responsável por exibir os relatórios solicitados pelo cliente.
- Deverá ser apresentado um único relatório nos moldes da seção **O QUE DEVE SER APRESENTADO PELO ALUNO NESTE TRABALHO.**
 - **Geração de gráficos.**
 - **Aplicação de filtros por dimensão.**
 - **Cruzamento de dimensões.**

Observações:

- O aluno pode desenvolver o módulo ETL, bem como os relatórios em qualquer tecnologia, linguagem de programação de sua preferência.
- O código do projeto deve ser subido no **GITHUB** e encaminhado o link ao professor.