**Levantamento da série histórica das bolsas de estudos concedidas pela CAPES no período entre 2013 e 2019**

**Projeto da Disciplina de BI**

Prof. Anderson Nascimento

prof.anderson@ica.ele.puc-rio.br

**Componentes do Projeto:**

Alisson Souza

Carlos Eduardo Figueiredo

Felipe Alves

Marcelo Costa

Thayná Santos

|  |
| --- |
| **Histórico de Versões** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data** | **Versão** | **Descrição** | **Autor** | **Aprovado por** |
| <10/03/2021> | <0> | Revisão original | <nome> | <nome> |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Sumário**

[1 Introdução 5](#_Toc66321162)

[2 Estudo de Caso 6](#_Toc66321163)

[2.1. Descrição do Estudo de Caso 6](#_Toc66321164)

[3 Descrição do Modelo Transacional 7](#_Toc66321165)

[3.1. Fonte 1 - base de dados “2017-a-2020-bolsas-ativas-em-programas-de-mobilidade-internacional” do site “dados.gov.br” (Portal Brasileiro de Dados Abertos) do governo federal. 7](#_Toc66321166)

[4 Proposta de Processo de BI 8](#_Toc66321167)

[5 Modelo Multidimensional 9](#_Toc66321168)

[6 Elaboração do Data Warehouse 10](#_Toc66321169)

[6.1 Definição do DW 10](#_Toc66321170)

[7 Projeto de ETL 11](#_Toc66321171)

[7.1 Descrição do Projeto de ETL 11](#_Toc66321172)

[8 Dashboard 13](#_Toc66321173)

[8.1 Descrição da Elaboração 13](#_Toc66321174)

[8.2 Telas do Dashboard 13](#_Toc66321175)

[9 Conclusão 15](#_Toc66321176)

[10 Anexos 16](#_Toc66321177)

[11 Arquivos 17](#_Toc66321178)

# Introdução

Este documento tem por finalidade coletar, analisar e definir as principais necessidades do projeto do estudo de caso “**Levantamento da série histórica das bolsas de estudos concedidas pela CAPES no período entre 2013 e 2019”**. O documento procura demonstrar os principais problemas atuais e o foco investigativo desejado pela CAPES.

# Estudo de Caso

## 2.1. Descrição do Estudo de Caso

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), é uma fundação do Ministério da Educação (MEC), e tem como função expandir e consolidar a pós-graduação stricto sensu (mestrado e doutorado) em todos os estados da Federação. Em 2007, passou também a atuar na formação de professores da educação básica ampliando o alcance de suas ações na formação de pessoal qualificado no Brasil e no exterior.

A necessidade de um diagnóstico completo do quadro de bolsas de estudos da pós-graduação: a falta de detalhamento quanto à distribuição das bolsas por áreas do conhecimento, ou por regiões e valores num espaço temporal, foi a motivação principal desse projeto.

Dessa forma é possível por exemplo, criar um panorama das bolsas concedidas pela Capes, possibilitando aos pós-graduandos e universidades o entendimento sobre quais são as áreas do conhecimento mais ofertadas, de quais universidades, valores concedidos num espaço temporal além de servir como diretriz para o governo federal quanto a melhor distribuição dos recursos destinados ao Ministério da Educação, mais precisamente à conceção de bolsas pela Capes.

O projeto tratará a concessão de bolsas de estudos como fato ativo a ser estudado, levando em consideração as dimensões de bolsa, local e aluno. A proposta será cobrir todas as informações a respeito da junção das dimensões com o fato e consequentemente abstrair informações que possam servir de apoio para uma estratégia de investimento mínimo necessário e total eficiência que corrobore com a principal função da instituição.

É desejável que ao final do projeto tenhamos relatórios como: quantidade de bolsas por região geográfica (Brasil e outros países), quantidade de bolsas por área do conhecimento, quantidade de bolsas por área do conhecimento por região geográfica, valor total das bolsas por área de conhecimento e valor total das bolsas por região geográfica.

Ao final do projeto a instituição, além da construção do *Data Warehouse*, espera que seja desenvolvido um *dashboard* com os dados mais relevantes sobre as concessões de bolsas, de forma que os gestores possam ter informações rápidas a qualquer momento.

# Descrição do Modelo Transacional

Originalmente a base da Capes continha no total 49 campos e 11 diferentes categorias para o campo “Moeda funcional de Pagamento” (CD\_MOEDA).

Foram efetuadas algumas alterações de forma que atendesse ao escopo do projeto. Para isso foram suprimidos da base original 31 campos correspondentes a valores de benefícios concedidos aos estudantes pela Capes, valores esses que compunham o valor total da bolsa concedida aos estudantes. Foi descartado o campo “Area de avaliação” (NM\_AREA\_AVALIACAO). No campo “Moeda funcional de Pagamento” (CD\_MOEDA) mantivemos apenas a moeda brasileira (BRL).

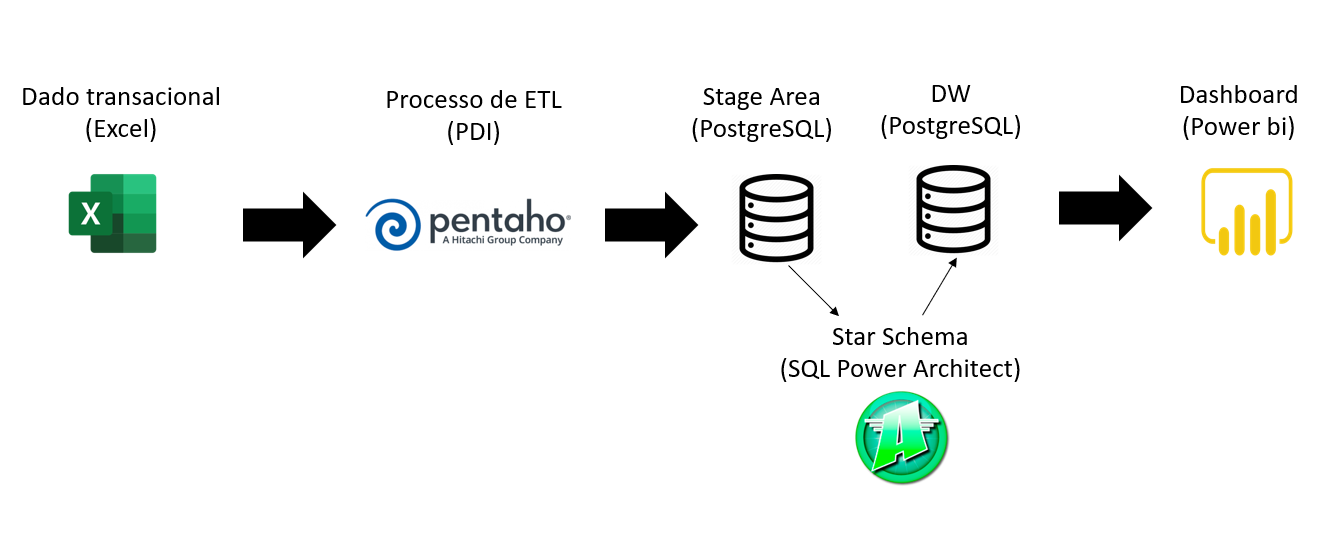
Sendo assim a nova base passou a ser mais enxuta com um total de 17 campos.

## 3.1. Fonte 1 - base de dados “2017-a-2020-bolsas-ativas-em-programas-de-mobilidade-internacional” do site “dados.gov.br” (Portal Brasileiro de Dados Abertos) do governo federal.

# Proposta de Processo de BI

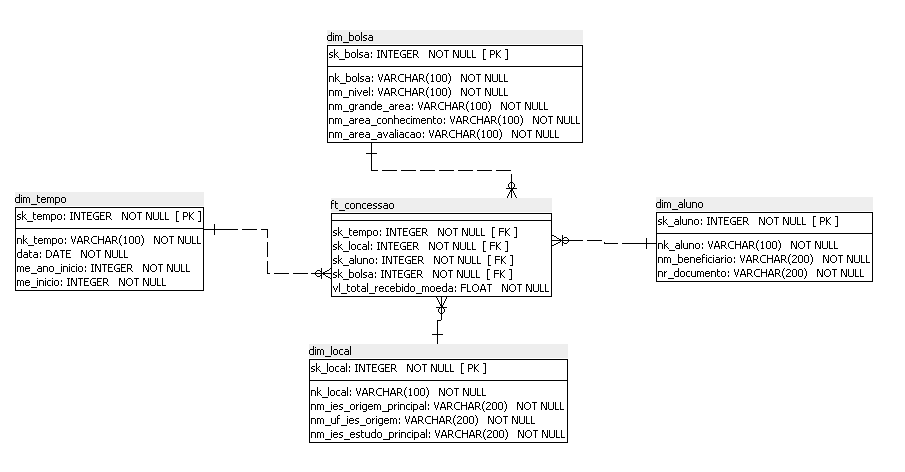
No site de dados do governo será extraída a base de dados de bolsas de estudos concedidas pela CAPES, esse material aponta uma série histórica de 2013 até 2019 com valores pagos no formato de Excel. A princípio, será executado no pdi um tratamento da base de dados para que seja possível carregar no PostgreSQL. Com a etapa anterior concluída, passaremos para o carregamento da base tratada no PostgreSQL. O passo seguinte será o desenvolvimento do modelo Star Schema utilizando o Power Architect e depois será carregado o modelo no DW e seus respectivos dados. Por fim, vamos desenvolver as visualizações no Power Bi.

Proposta:



# Modelo Multidimensional

Esta seção apresenta o *star schema* do programa de concessão de bolsas do governo. O modelo foi desenhado para aprofundar mais o entendimento sobre o perfil do bolsista e o montante gasto com essas bolsas.



# Elaboração do Data Warehouse

O Data Warehouse será a fonte integradora de informações da empresa, a tecnologia será utilizada com o intuito de servir de base para a camada de aplicação que será responsável por fornecer dados para a tomada de decisão na organização.

Os dados serão extraídos a partir das informações imputados pelo sistema transacional da instituição e consolidados em um arquivo Excel. Em uma área temporária (stage) os dados, deste do arquivo, serão selecionados, tratados e devidamente organizados. Na etapa seguinte serão armazenados em *Data Marts(DM)* independentes controlados por um grupo específico de usuários (*stand alone*). As informações consolidadas para apoio à tomada de decisão estarão no painel de BI desenhado para este fim e disponível para os gestores.

## 6.1 Definição do DW

* + 1. Arquitetura

A arquitetura do DW será independente. Esta arquitetura implica em DM independentes controlados por um grupo específico de usuários e que atende as necessidades específicas destes usuários, sem conexão com outros DM. Foi escolhido pela facilidade de implementação e os recursos e pessoal operacional podem ser administrados pelo próprio departamento.

* + 1. Abordagem de Construção

A construção será *Bottom Up*. A escolha deste modelo se dá pois requer menor custo e politicamente mais fácil de ser desenvolvida. Trazendo maior agilidade, pois a construção é altamente direcionada. O propósito dessa implementação é que permite a construção de um DW incremental a partir de DM independentes

* + 1. Arquitetura Física

O DW estará na nuvem. A escolha desta arquitetura foi definida devido ao baixo custo operacional, segurança, facilidade de escalonamento e a responsabilidade da manutenção ser da empresa contratada.

# Projeto de ETL

## 7.1 Descrição do *Projeto de ETL*

Essa análise de business intelligence utilizou dados disponibilizados no portal do governo federal dados.gov.br. O processo de ETL foi realizado no Pentaho Data Integration (PDI).

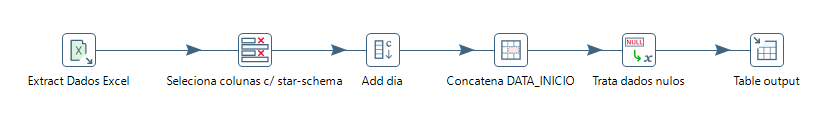
Na primeira etapa, os dados foram extraídos de uma planilha excel disponibilizada no portal do governo, tratados no PDI e carregados em uma stage area no SGBD PostgreSQL (schema: “cargastage”).

Na segunda etapa, os dados foram extraídos da stage area para serem carregados no DW (schema: “dwstagearea”). O PDI foi utilizado para a criação das tabelas dimensões e fato e seus relacionamentos.

A etapa de tratamento dos dados contemplou quatro passos, sendo eles:

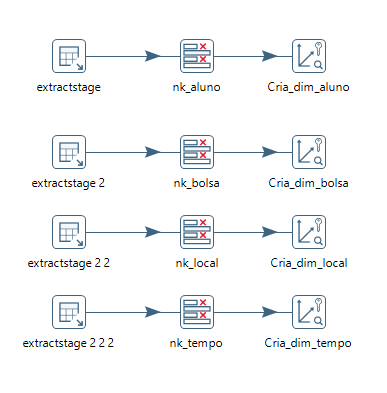
1. Seleção das colunas interessantes para a análise, de acordo com o modelo multidimensional criado no Power Architect;
2. Criação de uma coluna “dia”, com valor fixo de 01, para composição da data de início da bolsa, visto que o transacional registrava apenas o mês e o ano de início.
3. Concatenação das colunas: dia, me\_inicio e an\_inicio para criação da coluna data\_inicio. A informação da data do início de vigência da bolsa de estudos é que é relevante para a análise.
4. Preenchimento de campos nulos com a abreviatura “N.D.”, visto que o modelo multidimensional não permite campos nulos e não é desejável desconsiderar os registros que não tenham algum campo não preenchido.

A primeira etapa de carga foi para armazenamento dos dados já tratados na stage area criada no SGBD PostgreSQL.

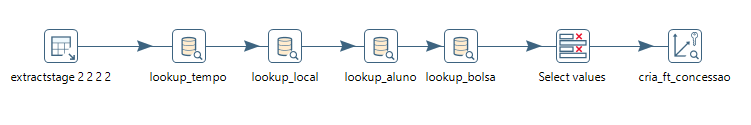


A imagem abaixo mostra as etapas de criação das dimensões e da tabela fato. Para cada dimensão foram executadas as seguintes etapas:

1. Extração dos dados da stage area do PostgreSQL;
2. Criação da chave natural;
3. Criação da dimensão e da chave estrangeira.



Foi utilizada a operação lookup para percorrer os registros de cada dimensão em busca da chave estrangeira e carregá-los na tabela fato.



Devido à erros de formato, quando da criação da tabela fato, o grupo escolheu não levar a coluna “dia” para o “dwbolsascapes”, visto que esse parâmetro não era fundamental para a análise.

# Dashboard

## 8.1 Descrição da Elaboração

Foi utilizado o Power BI para elaboração do dashboard: Análise da série histórica das bolsas de estudos CAPES.

No data warehouse importado do banco de dados Postgres temos apenas uma métrica na tabela fato, contendo os valores em reais concedidos para cada bolsa de estudo.

Para apresentar tanto a quantidade quanto o valor das bolsas foram criadas duas medidas.

1. Bolsas de Estudo = COUNTROWS(‘dwbolsascapes ft\_concessao’)
2. Valor Bolsas de Estudo = SUM(‘dwbolsascapes ft\_concessao‘[vl\_total\_recebido\_moeda])

Criamos a tabela \_Medidas que contém todas as medidas. Além das já mencionadas para quantidade e valor temos cinco utilizadas nos títulos dos gráficos, a MedidaDinamica e a Medida Selecionada, totalizando nove medidas.

A MedidaDinamica faz uso da Medida Selecionada que utiliza os campos da tabela \_Medida. Obs: a tabela \_Medida foi criada no Power BI para que a Medida Selecionada funcione.

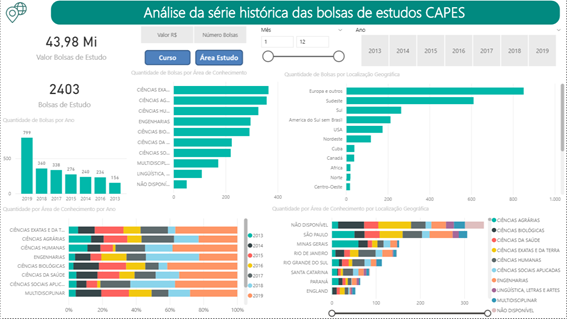
Foram agrupados os estados do Brasil nas regiões Norte, Sul, Nordeste, Sudeste e Centro-Oeste na tabela dwbolsascapes dim\_local, para estados e cidades de outros países o agrupamento foi feito por país.

## 8.2 Telas do Dashboard

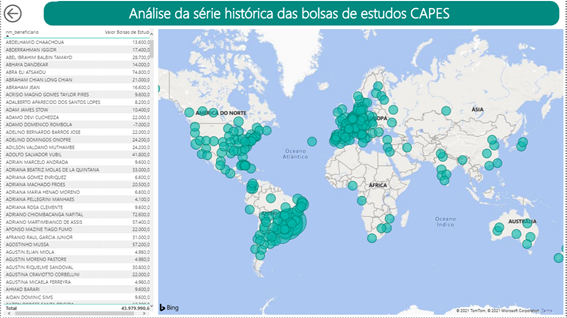
O dashboard é composto por duas telas, a primeira chamada de principal tem na parte superior esquerda dois cartões, um com o valor das bolsas e outro com a quantidade de bolsas concedidas.

Bem ao lado dos cartões, ainda na parte superior temos quatro botões que alternam os valores mostrados entre: valores em reais ou quantidade de bolsas e curso ou área de estudo. Esses botões junto com as segmentações de mês e ano fazem a seleção de filtros de forma conjunta de todos os elementos do dashboard.

Exemplos:



A segunda tela do dashboard é chamada de Mapa e pode ser acessada pelo ícone de globo no extremo superior esquerdo do dashboard Principal.



# Conclusão

A aplicação dos conceitos e ferramentas de business intelligence apresentados em aula possibilitou a construção do dashboard: Análise de série histórica das bolsas de estudos CAPES.

As 2404 linhas de dados obtidas do site “dados.gov.br”, foram transformadas nas informações apresentadas no dashboard. A escolha dos elementos da tela, sua distribuição e ordenação proporcionam uma visão global do histórico de concessão de bolsas de estudo.

Com os filtros disponíveis é possível interagir com a ferramenta e, baseado nos dados apresentados, obter entendimento de forma rápida para os questionamentos que esse trabalho se propõe a responder.

É possível diagnosticar e tirar conclusões sobre distribuição de bolsas de estudo por ano, por mês, por área de estudo, por local geográfico, por quantidade, por valor, beneficiário, etc.

Com isso podemos afirmar que o conhecimento adquirido a partir do dashboard possibilita a criação de ações para corrigir os desvios que existam entre o quadro real e o esperado de concessão de bolsas de estudos.

# Anexos

# Arquivos

* “br-capes-bolsas-proposta simplificado” - Arquivo em Excel base para o trabalho.
* “ETL\_pdi.KTR” - Processo de ETL feito no Pentaho.
* “multidimensional2.ktr” - Criação das tabelas dimensão e fato.
* “bolsas VF.architect” - Estrutura no Power Architect
* “Bolsascapes.sql” - visão do banco no Postgresql
* “Bolsa3\_capes” - Dashboard no Power Bi do projeto.