**DOCUMENTAÇÃO PROJETO BD CGE\_CE**

Data Analytics | Turma 18

Digital College Brasil | Fortaleza/CE

**Sobre o Projeto:**

A área de negócio do projeto foi uma base real de Execução Financeira e Despesa da Controladoria Geral do Estado do Ceará (CGE-CE).

O objetivo deste projeto é construir um ambiente de análise de dados robusto a partir de uma base relacional, utilizando as técnicas de ETL para criar um Data Warehouse eficaz.

O projeto visa desenvolver análises que servirão como base para tomadas de decisões estratégicas para o negócio.

**1. Time envolvido:**

Renato Albuquerque

Vinicius Costa

Alexsandro Cassiano

Antonio Marcos

Emmanuel Pinto

**2. Prazos:**

Início: 28/10/2024

Término: 23/11/2024

**3. Planejamento / Kick off:**

Leitura e discussão do material (Briefing do Projeto, Dicionário explicando as colunas e dados da tabela, Conhecimento do bd).

**3.1. Cronograma**

✅ 1ª Entendimento inicial sobre o PROJETO FINAL Unid. 01 (Leitura e discussão do material).

> bd de Execução Financeira e Despesa da Controladoria Geral do Estado do Ceará (CGE-CE).

> Briefing do projeto, dicionário explicando todas as colunas e dados da tabela.

✅ 2ª Criar um banco de dados do projeto (bd producao).

✅ 3ª Realizar um RESTORE na base de dados do projeto.

A base de dados possui apenas 01 tabela (execucao\_financeira\_despesa) com 30 colunas e várias linhas com dados.

✅ 4ª Criar um bd stage.

✅ 5ª Criar a tabela execucao\_financeira\_despesa.

✅ 6ª Inserir os dados na tabela.

✅ 7ª Realizar etapa de ETL (Extract, Transform, Load) / Tratamento dos dados (Verificar valores nulos, duplicados, inconsistentes [Negativos, zerados]).

✅ 8ª Criar um bd dw (data warehouse).

✅ 9ª Criar a modelagem lógica para o Data Warehouse, Ex.: Modelagem Star Schema (Tabelas Dimensão x Tabela Fato).

Obrigatório criar a dimensão ÓRGÃO (dim.orgao).

(BdSchema).

Elaborar métricas, indicadores, a partir da unidade de negócio: Item Elemento e Item Categoria.

✅ 10ª Criar as tabelas do dw (data warehouse).

(Modelagem física).

✅ 11ª Inserir os dados nas tabelas Dimensão e Fato.

✅ 12ª Conectar o dw (data warehouse) ao Microsoft Power BI.

✅ 13ª Criar um dashboard.

Elaborar gráficos a partir da unidade de negócio: Item Elemento e Item Categoria.

✅ 14ª Fazer apresentação do projeto.

**4. Execução:**

Etapa 01 referente ao entendimento inicial do projeto, desenvolvimento do cronograma de atividades.

**Etapa 02:**

-- Criar bd "projeto\_cgece".

create database projeto\_cgece;

**Etapa 03:**

-- Restore do bd do projeto.

-- Visualizar a tabela execucao\_financeira\_despesa (public), as colunas e os dados.

select \* from public.execucao\_financeira\_despesa

limit 5;

**Etapa 04:**

-- Criar o bd STAGE.

create schema stage;

**Etapa 05:**

-- Criar tabela execucao\_financeira\_despesa.

create table stage.execucao\_financeira\_despesa (

num\_ano character varying not null,

cod\_ne character varying not null,

codigo\_orgao character varying not null,

dsc\_orgao character varying,

cod\_fonte character varying,

dsc\_fonte character varying,

cod\_funcao character varying,

dsc\_funcao character varying,

cod\_item character varying,

dsc\_item character varying,

cod\_item\_elemento character varying,

dsc\_item\_elemento character varying,

cod\_item\_categoria character varying,

dsc\_item\_categoria character varying,

cod\_item\_grupo character varying,

dsc\_item\_grupo character varying,

dsc\_modalidade\_licitacao character varying,

cod\_item\_modalidade character varying,

dsc\_item\_modalidade character varying,

cod\_programa character varying,

dsc\_programa character varying,

cod\_subfuncao character varying,

dsc\_subfuncao character varying,

cod\_np character varying,

vlr\_empenho numeric(18,2),

vlr\_pagamento numeric(18,2),

dth\_empenho date,

dth\_pagamento date,

id serial not null primary key,

num\_ano\_np character varying

);

**Etapa 06:**

-- Carregar dados na tabela stage.execucao\_financeira\_despesa.

select \* from stage.execucao\_financeira\_despesa;

insert into stage.execucao\_financeira\_despesa (

num\_ano,

cod\_ne,

codigo\_orgao,

dsc\_orgao,

cod\_fonte,

dsc\_fonte,

cod\_funcao,

dsc\_funcao,

cod\_item,

dsc\_item,

cod\_item\_elemento,

dsc\_item\_elemento,

cod\_item\_categoria,

dsc\_item\_categoria,

cod\_item\_grupo,

dsc\_item\_grupo,

dsc\_modalidade\_licitacao,

cod\_item\_modalidade,

dsc\_item\_modalidade,

cod\_programa,

dsc\_programa,

cod\_subfuncao,

dsc\_subfuncao,

cod\_np,

vlr\_empenho,

vlr\_pagamento,

dth\_empenho,

dth\_pagamento,

id,

num\_ano\_np

)

select num\_ano,

cod\_ne,

codigo\_orgao,

dsc\_orgao,

cod\_fonte,

dsc\_fonte,

cod\_funcao,

dsc\_funcao,

cod\_item,

dsc\_item,

cod\_item\_elemento,

dsc\_item\_elemento,

cod\_item\_categoria,

dsc\_item\_categoria,

cod\_item\_grupo,

dsc\_item\_grupo,

dsc\_modalidade\_licitacao,

cod\_item\_modalidade,

dsc\_item\_modalidade,

cod\_programa,

dsc\_programa,

cod\_subfuncao,

dsc\_subfuncao,

cod\_np,

vlr\_empenho,

vlr\_pagamento,

dth\_empenho,

dth\_pagamento,

id,

num\_ano\_np

from public.execucao\_financeira\_despesa;

-- Visualizar a tabela execucao\_financeira\_despesa (stage), as colunas e os dados.

select \* from stage.execucao\_financeira\_despesa

limit 5;

**Etapa 07:**

-- Definir colunas a serem trabalhadas.

-- Obrigatórias para a equipe A: orgao, item\_elemento, item\_categoria.

-- Proposta para o dw: dim\_tempo, dim\_orgao, dim\_item\_elemento, dim\_item\_categoria, fato (5 tabelas).

**-- Tratamento dos dados - Valores nulos. Colunas codigo\_orgao, dsc\_orgao.**

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa;

select count(\*) as total, count(codigo\_orgao) as total\_cod\_orgao, count(dsc\_orgao) as total\_dsc\_orgao

from stage.execucao\_financeira\_despesa; -- Coluna dsc\_orgao com valores a menor, checar.

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa

where dsc\_orgao is null; -- 22 linhas null (439 linhas no total).

update stage.execucao\_financeira\_despesa

set dsc\_orgao = 'NÃO INFORMADO'

where dsc\_orgao is null; -- Valores atualizados.

select count(\*) as total, count(codigo\_orgao) as total\_cod\_orgao, count(dsc\_orgao) as total\_dsc\_orgao

from stage.execucao\_financeira\_despesa; -- Todas as colunas com valores iguais.

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa;

**-- Tratamento dos dados - Valores duplicados. Colunas codigo\_orgao, dsc\_orgao.**

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa

group by codigo\_orgao, dsc\_orgao; --1086 resultados.

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa

group by codigo\_orgao, dsc\_orgao

having count(codigo\_orgao) > 1; --1084 resultados.

update stage.execucao\_financeira\_despesa

set dsc\_orgao = 'FUNDO DE PREVIDENCIA PARLAMENTAR DA ASSEMB LEGISL DO CE'

where dsc\_orgao = 'FUNDO DE PREVIDENCIA PARLAMENTAR'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_orgao = 'SECRETARIA DA FAZENDA'

where dsc\_orgao = 'SECRETARIA DA FAZENDAsssssssssssssssssss'; -- Valores atualizados.

--

-- Dados com mesmo código e duas descrições (Checar). (Em desenvolvimento).

--

**-- Tratamento dos dados - Valores nulos. Colunas cod\_item\_elemento, dsc\_item\_elemento.**

select distinct cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa;

select count(\*) as total,

count(cod\_item\_elemento) as total\_cod\_item\_elemento,

count(dsc\_item\_elemento) as total\_dsc\_item\_elemento

from stage.execucao\_financeira\_despesa; -- Coluna count(\*) com valores a menor, checar.

select cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa

where dsc\_item\_elemento is null; -- 1683 linhas null.

update stage.execucao\_financeira\_despesa

set cod\_item\_elemento = '00'

where cod\_item\_elemento is null; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'NÃO INFORMADO'

where dsc\_item\_elemento is null; -- Valores atualizados.

select count(\*) as total,

count(cod\_item\_elemento) as total\_cod\_item\_elemento,

count(dsc\_item\_elemento) as total\_dsc\_item\_elemento

from stage.execucao\_financeira\_despesa; -- Todas as colunas com valores iguais.

select distinct cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa;

**-- Tratamento dos dados - Valores duplicados. Colunas cod\_item\_elemento, dsc\_item\_elemento.**

select distinct cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa

group by cod\_item\_elemento, dsc\_item\_elemento; -- 82 resultados.

select distinct cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa

group by cod\_item\_elemento, dsc\_item\_elemento

having count(dsc\_item\_elemento) > 1; -- 82 resultados.

**-- Corrigir dados duplicados, erro digitacao, texto minúsculo.**

update stage.execucao\_financeira\_despesa

set cod\_item\_elemento = '11'

where dsc\_item\_elemento = 'VENCIMENTOS E VANTAGENS FIXAS - PESSOAL'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTRAS DESPESAS VARIÁVEIS - PESSOAL CIVIL'

where dsc\_item\_elemento = 'OUTRAS DESPESAS VARIÁVEIS - PESSOAL CIVI'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = upper(dsc\_item\_elemento);

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'PREMIAÇÕES CULTURAIS, ARTÍSTICAS, CIENTÍFICAS, DESPORTIVAS E OUTRAS'

where dsc\_item\_elemento = 'PREMIAÇÕES CULTURAIS, ARTÍSTICAS, CIENTÍ'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTROS ENCARGOS SOBRE A DÍVIDA POR CONTRATO'

where cod\_item\_elemento = '22'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'ENCARGOS PELA HONRA DE AVAIS, GARANTIAS, SEGUROS E SIMILARES'

where cod\_item\_elemento = '27'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTRAS DESPESAS DE PESSOAL DECORRENTES DE CONTRATOS DE TERCEIRIZAÇÃO'

where cod\_item\_elemento = '34'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'SERVIÇOS DE CONSULTORIA'

where cod\_item\_elemento = '35'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTROS SERVIÇOS DE TERCEIROS - PESSOA FÍSICA'

where cod\_item\_elemento = '36'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTROS SERVIÇOS DE TERCEIROS - PESSOA JURÍDICA'

where cod\_item\_elemento = '39'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'OUTROS AUXÍLIOS FINANCEIROS A PESSOAS FÍSICAS'

where cod\_item\_elemento = '48'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'PENSÕES ESPECIAIS'

where cod\_item\_elemento = '59'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'CONSTITUIÇÃO OU AUMENTO DE CAPITAL DE EMPRESAS'

where cod\_item\_elemento = '65'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'RATEIO PELA PARTICIPAÇÃO EM CONSÓRCIO PÚBLICO'

where cod\_item\_elemento = '70'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_elemento = 'RESSARCIMENTO DE DESPESAS DE PESSOAL REQUISITADO'

where cod\_item\_elemento = '96'; -- Valores atualizados.

select cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa

group by cod\_item\_elemento, dsc\_item\_elemento;

**-- Tratamento dos dados - Valores nulos. Colunas cod\_item\_categoria, dsc\_item\_categoria.**

select distinct cod\_item\_categoria, dsc\_item\_categoria

from stage.execucao\_financeira\_despesa;

select count(\*) as total,

count(cod\_item\_categoria) as total\_cod\_item\_categoria,

count(dsc\_item\_categoria) as total\_dsc\_item\_categoria

from stage.execucao\_financeira\_despesa; -- Todas as colunas com a mesma qtde. de dados. 0 null.

**-- Tratamento dos dados - Valores duplicados. Colunas cod\_item\_categoria, dsc\_item\_categoria.**

select distinct cod\_item\_categoria, dsc\_item\_categoria

from stage.execucao\_financeira\_despesa;

update stage.execucao\_financeira\_despesa

set dsc\_item\_categoria = 'DESPESAS CORRENTES'

where dsc\_item\_categoria = 'DESPESA CORRENTE'; -- Valores atualizados.

update stage.execucao\_financeira\_despesa

set dsc\_item\_categoria = 'DESPESAS DE CAPITAL'

where dsc\_item\_categoria = 'DESPESA DE CAPITAL'; -- Valores atualizados.

select distinct cod\_item\_categoria, dsc\_item\_categoria

from stage.execucao\_financeira\_despesa;

**-- Tratamento dos dados - Valores nulos. Colunas vlr\_empenho, vlr\_pagamento.**

select \* from stage.execucao\_financeira\_despesa

limit 5;

select distinct vlr\_empenho, vlr\_pagamento

from stage.execucao\_financeira\_despesa;

select count(\*) as total,

count(vlr\_empenho) as total\_vlr\_empenho,

count(vlr\_pagamento) as total\_vlr\_pagamento

from stage.execucao\_financeira\_despesa; **-- Coluna vlr\_pagamento com valores a menor.**

select vlr\_pagamento

from stage.execucao\_financeira\_despesa

where vlr\_pagamento is null; -- 111276 linhas null.

update stage.execucao\_financeira\_despesa

set vlr\_pagamento = '0.00'

where vlr\_pagamento is null; -- Valores atualizados.

select count(\*) as total,

count(vlr\_empenho) as total\_vlr\_empenho,

count(vlr\_pagamento) as total\_vlr\_pagamento

from stage.execucao\_financeira\_despesa; -- Valores corrigidos.

**-- Tratamento dos dados - Valores negativos. Colunas vlr\_empenho, vlr\_pagamento.**

select vlr\_pagamento, vlr\_empenho

from stage.execucao\_financeira\_despesa

where vlr\_pagamento < 0 or vlr\_empenho < 0;

**-- Tratamento dos dados - Valores nulos. Colunas dth\_empenho, dth\_pagamento.**

select distinct dth\_empenho, dth\_pagamento

from stage.execucao\_financeira\_despesa;

select count(\*) as total,

count(dth\_empenho) as total\_dth\_empenho,

count(dth\_pagamento) as total\_dth\_pagamento

from stage.execucao\_financeira\_despesa; -- Coluna dth\_pagamento com valores a menor.

select dth\_pagamento

from stage.execucao\_financeira\_despesa

where dth\_pagamento is null; -- 111276 linhas null.

update stage.execucao\_financeira\_despesa

set dth\_pagamento = '2024-01-01'

where dth\_pagamento is null; -- Valores atualizados.

select count(\*) as total,

count(dth\_empenho) as total\_dth\_empenho,

count(dth\_pagamento) as total\_dth\_pagamento

from stage.execucao\_financeira\_despesa; -- Valores corrigidos.

-- num\_ano, num\_ano\_np, cod\_ne, cod\_np (Checar necessidade de trabalhar com essas colunas). (Em desenvolvimento).

**Etapa 08:**

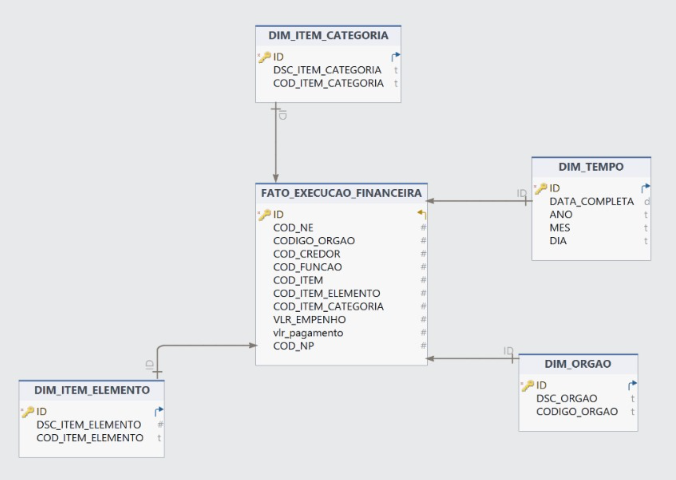
-- Criar o bd dw (data\_warehouse).

create schema dw;

**Etapa 09:**

Criar a modelagem lógica para o Data Warehouse (Modelagem Star Schema, Tabelas Dimensão x Tabela Fato).

(BdSchema).



**Etapa 10:**

-- Criar as tabelas do dw (data warehouse). 05 tabelas (dim\_tempo, dim\_orgao, dim\_item\_elemento

-- dim\_item\_categoria, fato\_execucao\_financeira).

-- (Modelagem física).

-- dim\_tempo

create table dw.dim\_tempo (

id serial not null primary key,

data\_inteira date,

ano integer,

mes integer,

dia integer

);

-- dim\_orgao

create table dw.dim\_orgao (

id serial not null primary key,

codigo\_orgao text,

dsc\_orgao text

);

-- dim\_item\_elemento

create table dw.dim\_item\_elemento (

id serial not null primary key,

cod\_item\_elemento text,

dsc\_item\_elemento text

);

-- dim\_item\_categoria

create table dw.dim\_item\_categoria (

id serial not null primary key,

cod\_item\_categoria text,

dsc\_item\_categoria text

);

-- fato\_execucao\_financeira

create table dw.fato\_execucao\_financeira (

id serial not null primary key,

id\_orgao INTEGER REFERENCES dw.dim\_orgao(id),

id\_item\_categoria INTEGER REFERENCES dw.dim\_item\_categoria(id),

id\_item\_elemento INTEGER REFERENCES dw.dim\_item\_elemento(id),

id\_dth\_empenho INTEGER REFERENCES dw.dim\_tempo(id),

id\_dth\_pagamento INTEGER REFERENCES dw.dim\_tempo(id),

vlr\_empenho NUMERIC(18,2),

vlr\_pagamento NUMERIC(18,2)

);

**Etapa 11:**

-- Inserir os dados nas tabelas Dimensão e Fato (dw).

-- dim\_tempo

INSERT INTO dw.dim\_tempo (data\_inteira, ano, mes, dia)

SELECT

dt as data\_inteira,

EXTRACT (YEAR FROM dt) AS ano,

EXTRACT (MONTH FROM dt) AS mes,

EXTRACT (DAY FROM dt) AS dia

FROM generate\_series (CURRENT\_DATE -INTERVAL '30 years', CURRENT\_DATE + INTERVAL '5 years', INTERVAL '1 day') AS dt;

-- dim\_orgao

insert into dw.dim\_orgao (codigo\_orgao, dsc\_orgao)

select distinct codigo\_orgao, dsc\_orgao

from stage.execucao\_financeira\_despesa;

-- dim\_item\_elemento

insert into dw.dim\_item\_elemento (cod\_item\_elemento, dsc\_item\_elemento)

select distinct cod\_item\_elemento, dsc\_item\_elemento

from stage.execucao\_financeira\_despesa;

-- dim\_item\_categoria

insert into dw.dim\_item\_categoria (cod\_item\_categoria, dsc\_item\_categoria)

select distinct cod\_item\_categoria, dsc\_item\_categoria

from stage.execucao\_financeira\_despesa

order by cod\_item\_categoria;

-- fato\_execucao\_financeira

insert into dw.fato\_execucao\_financeira (

id\_orgao, id\_item\_categoria, id\_item\_elemento, id\_dth\_empenho, id\_dth\_pagamento, vlr\_empenho, vlr\_pagamento)

select dor.id as id\_orgao,

dic.id as id\_item\_categoria,

die.id as id\_item\_elemento,

dt\_empenho.id as id\_data\_empenho,

dt\_pagamento.id as id\_data\_pagamento,

vlr\_empenho as valor\_empenho,

vlr\_pagamento as valor\_pagamento

from stage.execucao\_financeira\_despesa efd

inner join dw.dim\_orgao dor on efd.codigo\_orgao = dor.codigo\_orgao

inner join dw.dim\_item\_elemento die on efd.cod\_item\_elemento = die.cod\_item\_elemento

inner join dw.dim\_item\_categoria dic on efd.cod\_item\_categoria = dic.cod\_item\_categoria

inner join dw.dim\_tempo dt\_empenho on efd.dth\_empenho = dt\_empenho.data\_inteira

left join dw.dim\_tempo dt\_pagamento on efd.dth\_pagamento = dt\_pagamento.data\_inteira;

--

select \* from dw.dim\_tempo;

select \* from dw.dim\_orgao;

select \* from dw.dim\_item\_elemento;

select \* from dw.dim\_item\_categoria;

select \* from dw.fato\_execucao\_financeira;

--

**Etapa 12:**

Conectar o dw (data warehouse) com o Microsoft Power BI. (Ok).

**Etapa 13:**

-- Desenvolver o Dashboard.

****

**Etapa 14:**

Apresentação do projeto (Ok).