



Termo de Execução Descentralizada nº 19 – Realização de Estudos em Compras Públicas

Documento:

**Relatório de Extração de Dados
CNEP**

Data de Emissão:

07/02/2020

Elaborado por:

**Escola Nacional de Administração
Pública em parceria com Laboratório
de Tecnologias da Tomada de Decisão
– LATITUDE.UnB**

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

ENAP

Diogo Godinho Ramos Costa
Presidente

Diana Magalhães de Souza Coutinho
Diretor de Pesquisa e Pós-Graduação

Leonardo Monteiro Monasterio
Coordenador Geral de Ciência de Dados

Pedro Masson Sesconetto Souza
Coordenador de Ciência de Dados

EQUIPE TÉCNICA

Cristiano Alves Bezerra
Adalberto Bleme
Wanderson Maia Nascimento

UNIVERSIDADE DE BRASÍLIA

Márcia Abrahão Moura
Reitora

Marileusa Dosolina Chiarello
Diretora do Centro de Apoio ao
Desenvolvimento Tecnológico – CDT

Rafael Timóteo de Sousa Júnior
Coordenador do Laboratório de Tecnologias da
Tomada de Decisão – LATITUDE

EQUIPE TÉCNICA

Pesquisadores Sêniores
Ugo Silva Dias

EQUIPE TÉCNICA

Leticia Moreira Valle
Eduardo Calandrini Rocha da Costa
Anderson Alves de Oliveira
Andréia Campos Santana
Caio Matheus Campos de Oliveira
Danilo Santos Cardoso
Danilo Santos de Sales
Flávio Sousa da Vitória
Leonardo Pires Simões Vasconcelos
Samyra Lima Pereira

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

HISTÓRICO DE REVISÕES

Data	Versão	Autor	Descrição
15/10/2019	1.0	Eduardo Calandrini	Versão inicial do documento
06/10/2019	1.1	Leticia Valle	Inclusão dos dados da base CNEP
22/01/2020	1.2	Leticia Valle	Atualização do documento
07/02/2020	1.3	Leticia Valle	Atualização do documento



Universidade de Brasília – UnB
Campus Universitário Darcy Ribeiro - FT – ENE – Latitude
CEP 70.910-900 – Brasília-DF
Tel.: +55 61 3107-5598 – Fax: +55 61 3107-5590

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

SUMÁRIO

1. INTRODUÇÃO	6
2. ORIGEM DOS DADOS EXTRAÍDOS	6
3. QUANTITATIVO DE DADOS	6
4. MODELAGEM DO BANCO DE DADOS	6
5. FLUXOS DE ETL	8
<i>Tarefa 1 - Fazer download dos dados e salvar em arquivo</i>	8
<i>Tarefa 2 - Importar dados para o banco SQL</i>	8
<i>Tarefa 3 - Deletar arquivo</i>	9
6. DURAÇÃO DAS ROTINAS ETL	9
7. FLUXO DE TRATAMENTO DE ERROS	10
8. FLUXO DE AGENDAMENTO DE ROTINAS	10
9. ESTIMATIVA DE CRESCIMENTO	10
10. AUXÍLIO NOS ESTUDO DE COMPRAS PUBLICAS	10
11. EVIDÊNCIA DOS DADOS IMPORTADOS	11
12. BIBLIOGRAFIA	12
ANEXO I	13

1. INTRODUÇÃO

Este relatório tem como objetivo documentar o processo de extração, tratamento e carregamento de dados da base de Cadastro Nacional de Empresas Punidas (CNEP).

A base CNEP apresenta a relação das empresas que sofreram qualquer das punições previstas na Lei nº 12.846/2013 (Lei Anticorrupção) [1], além do registro dos acordos de leniência firmados pelas empresas com o poder público, inclusive aqueles que eventualmente sejam descumpridos.

Para a realização do trabalho, foram usadas as ferramentas de ETL Apache Airflow e o banco de dados SQL Server, rodando em um servidor Windows, requisito da equipe do Ministério da Economia.

2. ORIGEM DOS DADOS EXTRAÍDOS

Os dados são encontrados a partir da API (Application Programming Interface) disponível em <http://www.transparencia.gov.br/swagger-ui.html#/> e que retorna todos os dados das Empresas Punidas no formato JSON.

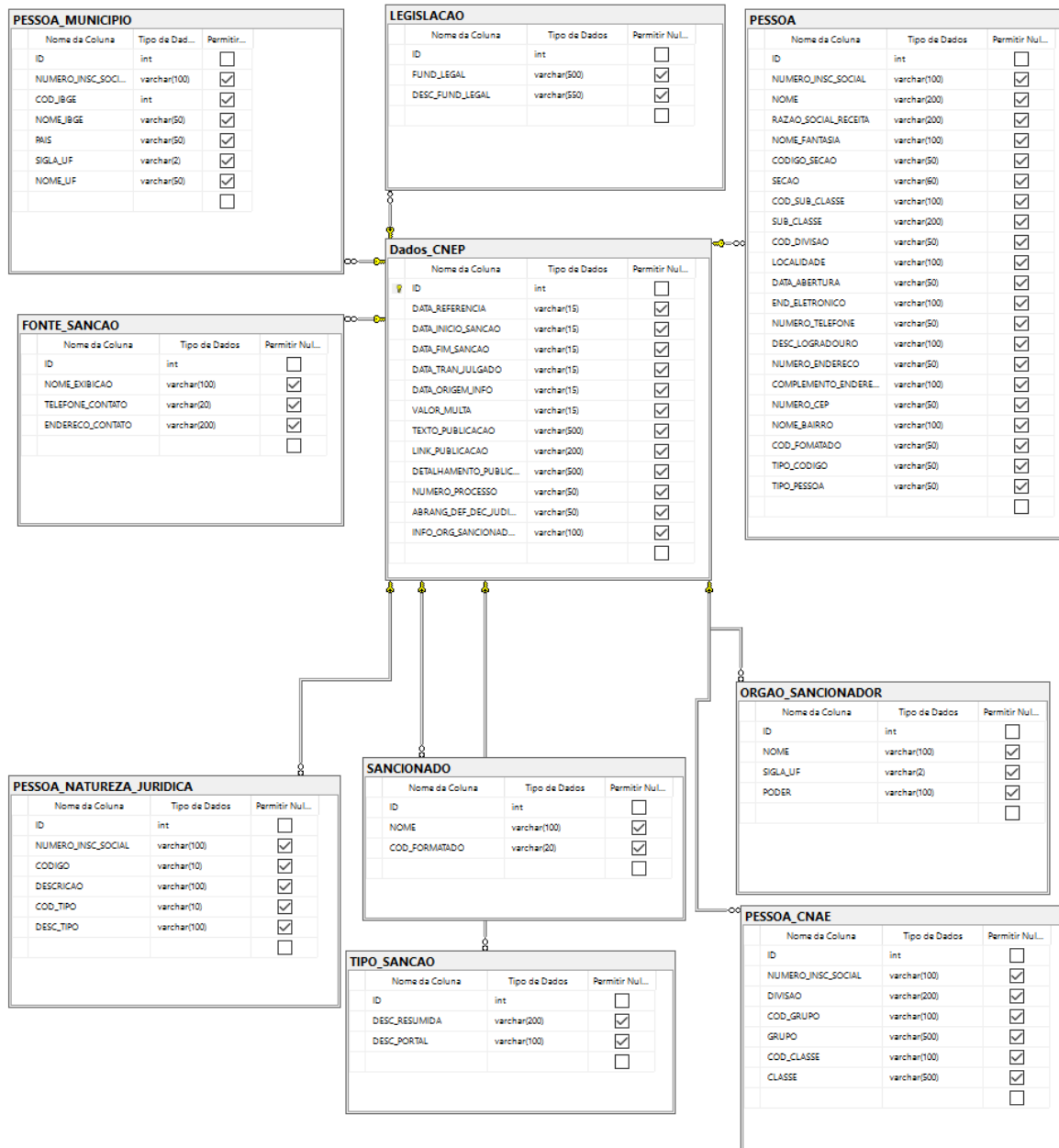
3. QUANTITATIVO DE DADOS

A base CNEP da forma como é disponibilizada, é composta por um arquivo JSON com 15 registros, na data de entrega desse documento.

4. MODELAGEM DO BANCO DE DADOS

Após análise da base CNEP, foi realizada a modelagem dos dados para posterior criação do banco e das tabelas. Para essa base específica, foram criadas 10 tabelas.

A seguir é apresentado o modelo lógico do banco. Por questões práticas de exibição, o script de criação de tabelas é apresentado no Anexo I.



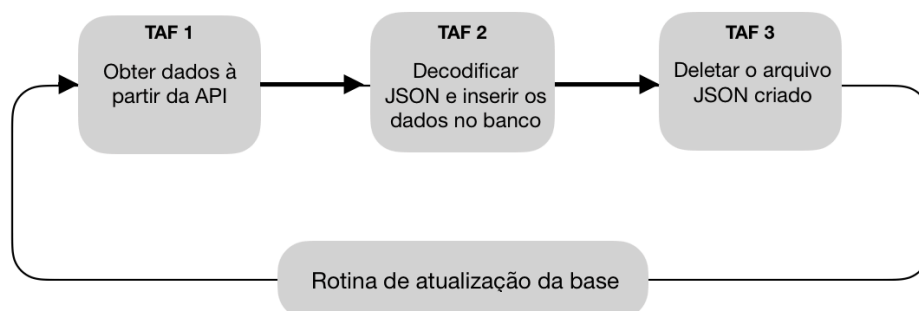
Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

5. FLUXOS DE ETL

A sessão a seguir apresenta um resumo do trabalho de extração, tratamento e carregamento da base CNEP.

O trabalho de ETL foi desenvolvido na ferramenta Apache Airflow e conta com 3 tarefas diretas, apresentadas no diagrama de bloco a seguir.



Tarefa 1 - Fazer download dos dados e salvar em arquivo

Rotina: **get_data**

Com o auxílio da biblioteca [requests](#) [2], é realizado o download da base de dados da base CNEP a partir da API. Os dados são salvos em um arquivo chamado cnep.json.

Tarefa 2 - Importar dados para o baco SQL

Rotina: **copy_to_sqlserver**

Com o arquivo cnep.json e com o auxílio das bibliotecas [json](#) [3] e [pyodbc](#) [4], é possível decodificar o arquivo json e conectar no banco sql para importar os dados das tabelas para o banco.

Foi utilizado o Driver ODBC [5] para SQL Sever para Linux.

```
conn_string = 'DRIVER={ODBC Driver 17 for SQL Server};SERVER='+
              host+';PORT=1433;DATABASE='+dbname+';UID='+
              username+';PWD='+ password +
              ';UseNTLMv2=yes;TDS_Version=8.0'
```

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

Tarefa 3 - Deletar arquivo

Rotina: **delete_file**

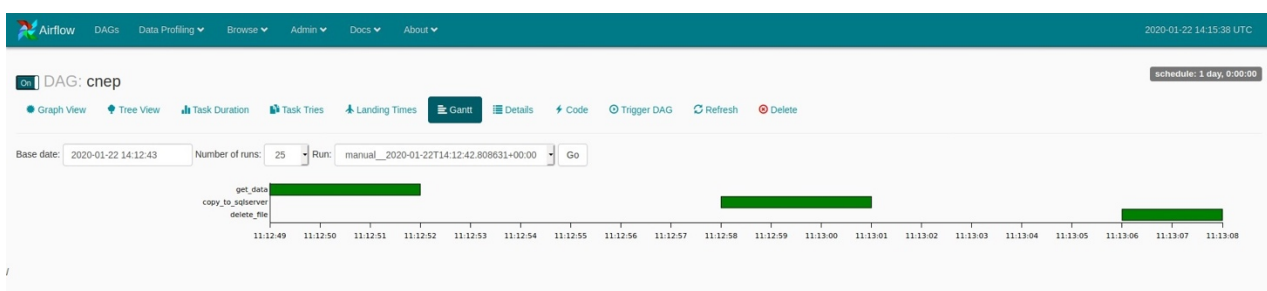
Com o auxílio da biblioteca [os](#) [6], o arquivo cnep.json criado é deletado

A imagem a seguir representa as tarefas ETL dentro da ferramenta Apache Airflow.



6. DURAÇÃO DAS ROTINAS ETL

A duração de importação dos dados depende do tempo de execução de todas as rotinas ETL. A imagem a seguir apresenta o tempo de execução de cada tarefa.



Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

Observa-se que cada tarefa possui um tempo de execução distinto dependendo da complexidade da tarefa. Para a base CNEP, o tempo de execução das rotinas foi de cerca de 3 minutos no total.

7. FLUXO DE TRATAMENTO DE ERROS

Os possíveis erros de execução das tarefas foram tratados à partir da criação de tarefas independentes. Cada tarefa procura pelos arquivos necessários para sua execução e gera como output arquivos que serão usados como entrada de outras tarefas. Dessa forma, caso alguma tarefa não seja cumprida, ou apresente erro, ao reiniciar o sistema, a rotina de tarefas será retomada e os arquivos salvos de tarefas anteriores continuam salvos.

Além disso, afim de evitar futuros erros relacionados à mudança da URL do API que contém o JSON para download, a tarefa inicial do ETL é fazer um filtro usando a URL do API disponibilizado na página, de forma a evitar erros gerados por mudanças estruturais na página de download do JSON.

8. FLUXO DE AGENDAMENTO DE ROTINAS

A base CNEP é atualizada diariamente. Dessa forma, no código de configuração do Apache Airflow, foi inserida uma rotina de atualização da base a cada 01 dias.

```
dag = DAG(dag_id='cnep', default_args=args, schedule_interval=timedelta(days=1))
```

9. ESTIMATIVA DE CRESCIMENTO

A estimativa de crescimento da base depende do volume atual de dados acrescido da estimativa de volume que vai ser inserido nas atualizações diárias.

Na entrega desse relatório, o volume total da base é de 54 KB, o que implica que mesmo após várias atualizações futuras, a base não devera passar de 1MB.

10. AUXÍLIO NOS ESTUDO DE COMPRAS PUBLICAS

A base de dados CNEP apresenta a relação das empresas que sofreram punições previstas na Lei Anticorrupção. Dentre as sanções previstas, estão a aplicação de multas, na esfera administrativa, até a perda de bens, a suspensão de atividades e a dissolução

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.

É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

compulsória, na esfera penal, além da proibição de receber incentivos, subvenções, doações ou empréstimos de órgãos e entidades públicas.

Com os dados da base CNEP, é possível realizar o cruzamento entre os dados das empresas punidas e os novos convênios, contratos de repasse ou termos de parceria com a administração pública federal, afim de evitar ou limitar a participação de tais empresas nesses pleitos.

11. EVIDÊNCIA DOS DADOS IMPORTADOS

Os dados importados corretamente no SQL Server são apresentados a seguir.

ID	DATA_REFERENCIA	DATA_INICIO_SANCAO	DATA_FIM_SANCAO	DATA_TRAN_JULGADO	DATA
75,900,488	21/01/2020	26/09/2019	Sem informação	Sem informação	27/09
75,900,489	21/01/2020	26/09/2019	Sem informação	Sem informação	27/09
76,100,274	21/01/2020	25/07/2017	Sem informação	Sem informação	16/05
76,205,940	21/01/2020	27/09/2019	27/09/2020	07/10/2019	07/10
76,205,941	21/01/2020	12/03/2018	Sem informação	Sem informação	16/03
76,205,942	21/01/2020	22/06/2018	22/06/2020	Sem informação	22/06
76,205,943	21/01/2020	07/11/2019	Sem informação	05/11/2019	19/11
76,205,944	21/01/2020	07/11/2019	Sem informação	05/11/2019	19/11
76,205,945	21/01/2020	15/08/2019	15/08/2020	24/08/2019	19/08
76,205,946	21/01/2020	27/09/2019	27/09/2020	07/10/2019	07/10
76,205,947	21/01/2020	15/08/2019	15/08/2020	Sem informação	11/09
76,205,950	21/01/2020	31/01/2018	Sem informação	Sem informação	07/02
76,205,952	21/01/2020	13/02/2019	Sem informação	07/02/2019	15/03
76,205,955	21/01/2020	27/09/2019	27/09/2020	07/10/2019	27/09
76,205,956	21/01/2020	27/09/2019	27/09/2020	07/10/2019	27/09

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

12. BIBLIOGRAFIA

1. Dados.gov. Página interna CNEP. Disponível em: < <http://dados.gov.br/dataset/cnep>>. Acesso em: 06 de novembro de 2019.
2. Request library documentation. Disponível em: <<https://docs.python.org/3/library/urllib.request.html>>. Acesso em: 21 de outubro de 2019.
3. JSON documentation. Disponível em: < <http://www.json.org/>>. Acesso em: 06 de novembro de 2019.
4. Driver SQL Python. Disponível em: < <https://docs.microsoft.com/pt-br/sql/connect/python/pyodbc/python-sql-driver-pyodbc?view=sql-server-ver15>>. Acesso em: 21 de outubro de 2019.
5. Microsoft ODBC Driver para SQL Server em Linux. <<https://docs.microsoft.com/en-us/sql/connect/odbc/linux-mac/installing-the-microsoft-odbc-driver-for-sql-server?view=sql-server-ver15#microsoft-odbc-driver-131-for-sql-server>> Acesso em: 21 de outubro de 2019.
6. Miscellaneous operating system interfaces. Disponível em: <<https://docs.python.org/3.4/library/os.html>>. Acesso em: 21 de outubro de 2019.

ANEXO I

```
CREATE TABLE dbo.Dados_CNEP(  
    ID int NOT NULL,  
    DATA_REFERENCIA varchar(15) NULL,  
    DATA_INICIO_SANCAO varchar(15) NULL,  
    DATA_FIM_SANCAO varchar(15) NULL,  
    DATA_TRAN_JULGADO varchar(15) NULL,  
    DATA_ORIGEM_INFO varchar(15) NULL,  
    VALOR_MULTA varchar(15) NULL,  
    TEXTO_PUBLICACAO varchar(500) NULL,  
    LINK_PUBLICACAO varchar(200) NULL,  
    DETALHAMENTO_PUBLICACAO varchar(500) NULL,  
    NUMERO_PROCESSO varchar(50) NULL,  
    ABRANG_DEF_DEC_JUDICIAL varchar(50) NULL,  
    INFO_ORG_SANCIONADOR varchar(100) NULL,  
    CONSTRAINT PK_Dados_CNEP PRIMARY KEY CLUSTERED  
    (ID ASC);
```

```
CREATE TABLE dbo.FONTE_SANCAO(  
    ID int NOT NULL,  
    NOME_EXIBICAO varchar(100) NULL,  
    TELEFONE_CONTATO varchar(20) NULL,  
    ENDERECO_CONTATO varchar(200) NULL);
```

```
CREATE TABLE dbo.LEGISLACAO(  
    ID int NOT NULL,  
    FUND_LEGAL varchar(500) NULL,  
    DESC_FUND_LEGAL varchar(550) NULL);
```

```
CREATE TABLE dbo.ORGAO_SANCIONADOR(  
    ID int NOT NULL,  
    NOME varchar(100) NULL,  
    SIGLA_UF varchar(2) NULL,  
    PODER varchar(100) NULL);
```

```
ALTER TABLE      dbo.ORGAO_SANCIONADOR      WITH      CHECK      ADD      CONSTRAINT  
FK_ORGAO_SANCIONADOR_Dados_CNEP FOREIGN KEY (ID)  
REFERENCES dbo.Dados_CNEP (ID);
```

```
ALTER      TABLE      dbo.ORGAO_SANCIONADOR      CHECK      CONSTRAINT  
FK_ORGAO_SANCIONADOR_Dados_CNEP
```

```
CREATE TABLE dbo.PESSOA(  
    ID int NOT NULL,  
    NUMERO_INSC_SOCIAL varchar(100) NULL,
```

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

```
NOME varchar(200) NULL,
RAZAO_SOCIAL_RECEITA varchar(200) NULL,
NOME_FANTASIA varchar(100) NULL,
CODIGO_SECAO varchar(50) NULL,
SECAO varchar(60) NULL,
COD_SUB_CLASSE varchar(100) NULL,
SUB_CLASSE varchar(200) NULL,
COD_DIVISAO varchar(50) NULL,
LOCALIDADE varchar(100) NULL,
DATA_ABERTURA varchar(50) NULL,
END_ELETRONICO varchar(100) NULL,
NUMERO_TELEFONE varchar(50) NULL,
DESC_LOGRADOURO varchar(100) NULL,
NUMERO_ENDERECO varchar(50) NULL,
COMPLEMENTO_ENDERECO varchar(100) NULL,
NUMERO_CEP varchar(50) NULL,
NOME_BAIRRO varchar(100) NULL,
COD_FOMATADO varchar(50) NULL,
TIPO_CODIGO varchar(50) NULL,
TIPO_PESSOA varchar(50) NULL);
```

```
CREATE TABLE dbo.PESSOA_CNAE(
    ID int NOT NULL,
    NUMERO_INSC_SOCIAL varchar(100) NULL,
    DIVISAO varchar(200) NULL,
    COD_GRUPO varchar(100) NULL,
    GRUPO varchar(500) NULL,
    COD_CLASSE varchar(100) NULL,
    CLASSE varchar(500) NULL);
```

```
ALTER TABLE dbo.PESSOA_CNAE WITH CHECK ADD CONSTRAINT
FK_PESSOA_CNAE_Dados_CNEP FOREIGN KEY (ID)
REFERENCES dbo.Dados_CNEP (ID);
```

```
ALTER TABLE dbo.PESSOA_CNAE CHECK CONSTRAINT FK_PESSOA_CNAE_Dados_CNEP;
```

```
CREATE TABLE dbo.PESSOA_MUNICIPIO(
    ID int NOT NULL,
    NUMERO_INSC_SOCIAL varchar(100) NULL,
    COD_IBGE int NULL,
    NOME_IBGE varchar(50) NULL,
    PAIS varchar(50) NULL,
    SIGLA_UF varchar(2) NULL,
    NOME_UF varchar(50) NULL);
```

```
ALTER TABLE dbo.PESSOA_MUNICIPIO WITH CHECK ADD CONSTRAINT
FK_PESSOA_MUNICIPIO_Dados_CNEP FOREIGN KEY (ID)
REFERENCES dbo.Dados_CNEP (ID);
```

```
ALTER TABLE dbo.PESSOA_MUNICIPIO CHECK CONSTRAINT
FK_PESSOA_MUNICIPIO_Dados_CNEP;
```

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.

```
CREATE TABLE dbo.PESSOA_NATUREZA_JURIDICA (
    ID int NOT NULL,
    NUMERO_INSC_SOCIAL varchar(100) NULL,
    CODIGO varchar(10) NULL,
    DESCRICAO varchar(100) NULL,
    COD_TIPO varchar(10) NULL,
    DESC_TIPO varchar(100) NULL);

ALTER TABLE dbo.PESSOA_NATUREZA_JURIDICA WITH CHECK ADD CONSTRAINT
FK_PESSOA_NATUREZA_JURIDICA_Dados_CNEP FOREIGN KEY (ID)
REFERENCES dbo.Dados_CNEP (ID);

ALTER TABLE dbo.PESSOA_NATUREZA_JURIDICA CHECK CONSTRAINT
FK_PESSOA_NATUREZA_JURIDICA_Dados_CNEP;

CREATE TABLE dbo.SANCIONADO (
    ID int NOT NULL,
    NOME varchar(100) NULL,
    COD_FORMATADO varchar(20) NULL);

ALTER TABLE dbo.SANCIONADO WITH CHECK ADD CONSTRAINT
FK_SANCIONADO_Dados_CNEP FOREIGN KEY (ID)
REFERENCES dbo.Dados_CNEP (ID);

ALTER TABLE dbo.SANCIONADO CHECK CONSTRAINT FK_SANCIONADO_Dados_CNEP;

CREATE TABLE dbo.TIPO_SANCAO (
    ID int NOT NULL,
    DESC_RESUMIDA varchar(200) NULL,
    DESC_PORTAL varchar(100) NULL);

ALTER TABLE dbo.TIPO_SANCAO WITH CHECK ADD CONSTRAINT
FK_TIPO_SANCAO_Dados_CNEP FOREIGN KEY (ID)
REFERENCES dbo.Dados_CNEP (ID);

ALTER TABLE dbo.TIPO_SANCAO CHECK CONSTRAINT FK_TIPO_SANCAO_Dados_CNEP;
```

Escola Nacional de Administração Pública

Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

www.ena.gov.br – www.rede.unb.br



Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a Enap.
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da Enap.