



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

Documento:

**Relatório de Extração de Dados
CNPJ**

Data de Emissão:

16/03/2020

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.

HISTÓRICO DE REVISÕES

Data	Versão	Autor	Descrição
22/11/2019	1.0	Flávio Vitória	Inclusão dos dados da base CNPJ
09/03/2020	1.1	Leticia Valle	Revisão
16/03/2020	1.2	Leticia Valle	Revisão



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.

SUMÁRIO

1.	INTRODUÇÃO	4
2.	ORIGEM DOS DADOS EXTRAÍDOS	4
3.	QUANTITATIVO DE DADOS	4
4.	MODELAGEM DO BANCO DE DADOS	5
5.	FLUXOS DE ETL	7
	<i>Tarefa 1 - Download do arquivo de dados</i>	<i>8</i>
	<i>Tarefa 2 – Extração dos dados</i>	<i>8</i>
	<i>Tarefa 3 – Tratamento dos dados</i>	<i>9</i>
	<i>Tarefa 4 – Importação da tabela no banco SQL Server</i>	<i>9</i>
	<i>Tarefa 5 – Apagar os arquivos intermediários</i>	<i>10</i>
6.	DURAÇÃO DAS ROTINAS ETL	10
7.	FLUXO DE TRATAMENTO DE ERROS	11
8.	FLUXO DE AGENDAMENTO DE ROTINAS	12
9.	ESTIMATIVA DE CRESCIMENTO	12
10.	AUXÍLIO NOS ESTUDO DE COMPRAS PUBLICAS	12
11.	EVIDÊNCIA DOS DADOS IMPORTADOS	12
12.	BIBLIOGRAFIA	14

1. INTRODUÇÃO

Este relatório tem como objetivo documentar o processo de extração, tratamento e carregamento de dados da base do Cadastro Nacional de Pessoas Jurídicas (CNPJ).

A base do CNPJ compreende as informações cadastrais das entidades de interesse das administrações tributárias da União, dos Estados, do Distrito Federal e dos Municípios. A administração do CNPJ compete à Secretaria da Receita Federal do Brasil (RFB). A mesma contém informações sobre a organização das empresas no que diz respeito à sua relação com a Receita Federal, assim como informações sobre o seu quadro societário. [1]

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 podem ser encontrados na página do portal da Receita Federal na categoria de dados do portal (dados abertos). A URL do local de origem dos dados é <https://receita.economia.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/dados-publicos-cnpj>

3. QUANTITATIVO DE DADOS

A base do CNPJ está dividida em 21 arquivos compactados, cada um contendo, em média, 250 MB. Descompactados, os mesmos podem chegar aos 100GB. Essa base é posicional, e o layout se encontra disponível em http://200.152.38.155/CNPJ/LAYOUT_DADOS_ABERTOS_CNPJ.pdf.

A base CNPJ contempla dois tipos de informações: as informações relativas às empresas e as informações relativas aos sócios das empresas. As informações estão misturadas nos registros, sendo identificáveis por um campo específico no início de cada registro. A base contém cerca de 67 milhões de registros.

4. MODELAGEM DO BANCO DE DADOS

Após análise da base do CNPJ, foi realizada a modelagem dos dados para posterior criação do banco e das tabelas. Como a base contém dois tipos diferentes de informação, essas informações foram quebradas em duas tabelas: a tabela DADOS_RECEITA_CNPJ e a tabela DADOS_RECEITA_SOCIOS.

A seguir são apresentados o modelo lógico do banco e o script para a criação da tabela.

DADOS_RECEITA_SOCIOS			
Nome da Coluna	Tipo de Dados	Permitir Nulos	
CNPJ	varchar(14)	<input checked="" type="checkbox"/>	
ID_SOCIO	int	<input checked="" type="checkbox"/>	
NOME_SOCIO	varchar(100)	<input checked="" type="checkbox"/>	
CNPJ_CPF_SOCIO	varchar(14)	<input checked="" type="checkbox"/>	
COD_QUALIFICACAO_S...	int	<input checked="" type="checkbox"/>	
PERCENTUAL_CAP_SOC...	int	<input checked="" type="checkbox"/>	
DATA_ENTRADA_SOCIE...	int	<input checked="" type="checkbox"/>	
COD_PAIS	int	<input checked="" type="checkbox"/>	
NOME_PAIS_SOCIO	varchar(100)	<input checked="" type="checkbox"/>	
CPF_REPRESENTANTE_L...	varchar(20)	<input checked="" type="checkbox"/>	
NOME_REPRESENTANT...	varchar(100)	<input checked="" type="checkbox"/>	
COD_QUALIFICACAO_R...	int	<input checked="" type="checkbox"/>	
		<input type="checkbox"/>	

```
CREATE TABLE DADOS_RECEITA_SOCIOS (
  CNPJ VARCHAR(14),
  ID_SOCIO INT,
  NOME_SOCIO VARCHAR(100),
  CNPJ_CPF_SOCIO VARCHAR(14),
  COD_QUALIFICACAO_SOCIO INT,
  PERCENTUAL_CAP_SOCIAL INT,
  DATA_ENTRADA_SOCIEDADE INT,
  COD_PAIS INT,
  NOME_PAIS_SOCIO VARCHAR(100),
  CPF_REPRESENTANTE_LEGAL VARCHAR(20),
  NOME_REPRESENTANTE_LEGAL VARCHAR(100),
  COD_QUALIFICACAO_REP_LEGAL INT
);
```

```
CREATE INDEX IDX_DADOS_RECEITA_SOCIOS ON DADOS_RECEITA_SOCIOS (CNPJ);
```

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.

DADOS_RECEITA_CNPJ			
Nome da Coluna	Tipo de Dados	Permitir Nulos	
CNPJ	varchar(14)	<input checked="" type="checkbox"/>	
IDENTIFICADOR_MATRIZ_FILIAL	int	<input checked="" type="checkbox"/>	
RAZAO_SOCIAL	varchar(200)	<input checked="" type="checkbox"/>	
NOME_FANTASIA	varchar(200)	<input checked="" type="checkbox"/>	
SITUACAO_CADASTRAL	int	<input checked="" type="checkbox"/>	
DATA_SITUACAO_CADASTRAL	varchar(10)	<input checked="" type="checkbox"/>	
MOTIVO_SITUACAO_CADASTRAL	int	<input checked="" type="checkbox"/>	
NOME_CIDADE_EXTERIOR	varchar(100)	<input checked="" type="checkbox"/>	
COD_PAIS	int	<input checked="" type="checkbox"/>	
NOME_PAIS	varchar(100)	<input checked="" type="checkbox"/>	
COD_NATUREZA_JURIDICA	int	<input checked="" type="checkbox"/>	
DATA_INICIO_ATIVIDADE	varchar(10)	<input checked="" type="checkbox"/>	
CNAE_FISCAL	int	<input checked="" type="checkbox"/>	
DESCRICAO_LOGRADOURO	varchar(250)	<input checked="" type="checkbox"/>	
LOGRADOURO	varchar(100)	<input checked="" type="checkbox"/>	
NUMERO	varchar(10)	<input checked="" type="checkbox"/>	
COMPLEMENTO	varchar(100)	<input checked="" type="checkbox"/>	
BAIRRO	varchar(100)	<input checked="" type="checkbox"/>	
CEP	varchar(10)	<input checked="" type="checkbox"/>	
UF	char(2)	<input checked="" type="checkbox"/>	
COD_MUNICIPIO	int	<input checked="" type="checkbox"/>	
MUNICIPIO	varchar(100)	<input checked="" type="checkbox"/>	
TELEFONE1	varchar(20)	<input checked="" type="checkbox"/>	
TELEFONE2	varchar(20)	<input checked="" type="checkbox"/>	
FAX	varchar(12)	<input checked="" type="checkbox"/>	
EMAIL	varchar(100)	<input checked="" type="checkbox"/>	
QUALIFICACAO_RESPONSAVEL	int	<input checked="" type="checkbox"/>	
CAPITAL_SOCIAL	varchar(14)	<input checked="" type="checkbox"/>	
PORTE_EMPRESA	varchar(2)	<input checked="" type="checkbox"/>	
OPCAO_SIMPLES	varchar(10)	<input checked="" type="checkbox"/>	
DATA_OPCAO_SIMPLES	varchar(10)	<input checked="" type="checkbox"/>	
DATA_EXCLUSAO_SIMPLES	varchar(10)	<input checked="" type="checkbox"/>	
OPCAO_MEI	char(1)	<input checked="" type="checkbox"/>	
SITUACAO_ESPECIAL	varchar(50)	<input checked="" type="checkbox"/>	
DATA_SITUACAO_ESPECIAL	varchar(10)	<input checked="" type="checkbox"/>	
		<input type="checkbox"/>	

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 DADOS_RECEITA_CNPJ (  
    CNPJ VARCHAR(14),  
    IDENTIFICADOR_MATRIZ_FILIAL INT,  
    RAZAO_SOCIAL VARCHAR(200),  
    NOME_FANTASIA VARCHAR(200),  
    SITUACAO_CADASTRAL INT,  
    DATA_SITUACAO_CADASTRAL VARCHAR(10),  
    MOTIVO_SITUACAO_CADASTRAL INT,  
    NOME_CIDADE_EXTERIOR VARCHAR(100),  
    COD_PAIS INT,  
    NOME_PAIS VARCHAR(100),  
    COD_NATUREZA_JURIDICA INT,  
    DATA_INICIO_ATIVIDADE VARCHAR(10),  
    CNAE_FISCAL INT,  
    DESCRICAO_LOGRADOURO VARCHAR(250),  
    LOGRADOURO VARCHAR(100),  
    NUMERO VARCHAR(10),  
    COMPLEMENTO VARCHAR(200),  
    BAIRRO VARCHAR(100),  
    CEP VARCHAR(10),  
    UF CHAR(2),  
    COD_MUNICIPIO INT,  
    MUNICIPIO VARCHAR(100),  
    TELEFONE1 VARCHAR(20),  
    TELEFONE2 VARCHAR(20),  
    FAX VARCHAR(12),  
    EMAIL VARCHAR(100),  
    QUALIFICACAO_RESPONSAVEL INT,  
    CAPITAL_SOCIAL VARCHAR(20),  
    PORTE_EMPRESA VARCHAR(2),  
    OPCAO_SIMPLES VARCHAR(10),  
    DATA_OPCAO_SIMPLES VARCHAR(10),  
    DATA_EXCLUSAO_SIMPLES VARCHAR(10),  
    OPCAO_MEI CHAR(1),  
    SITUACAO_ESPECIAL VARCHAR(50),  
    DATA_SITUACAO_ESPECIAL VARCHAR(10)  
);  
  
CREATE INDEX IDX_DADOS_RECEITA_CNPJ ON DADOS_RECEITA_CNPJ (CNPJ);
```

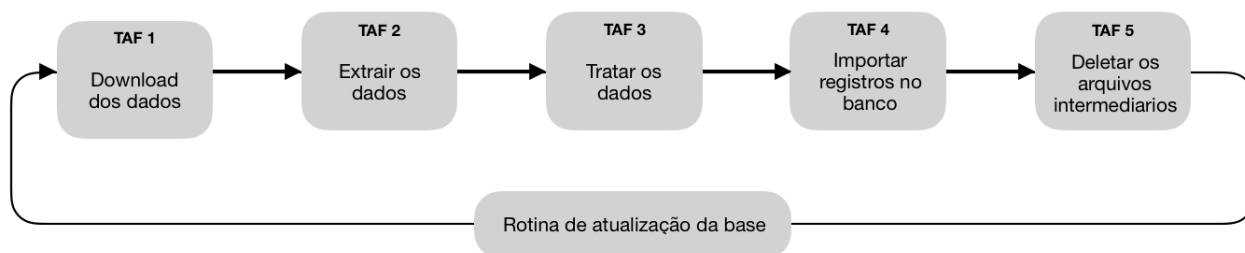
5. FLUXOS DE ETL

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

O trabalho de ETL foi desenvolvido na ferramenta Apache Airflow e conta com 5 tarefas diretas, apresentadas no diagrama de blocos a seguir.

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.



Considerando que a base está dividida em partes de tamanho considerável, e prevendo a possibilidade de novas partes serem acrescentadas no futuro, existe uma rotina de controle de fluxo para que todas as tarefas sejam realizadas seguindo um determinado roteiro e, em caso de queda de link, o processo consiga ser reiniciado do último ponto concluído com sucesso.

Tarefa 1 - Download do arquivo de dados

Rotina: **get_data_from_url**

Aproveitamos que existe uma regra de formação para a URL que disponibiliza os arquivos (http://200.152.38.155/CNPJ/DADOS_ABERTOS_CNPJ_{indice}.zip) aonde o índice vai de 01 até 20 para utilizar a biblioteca [requests](#) [2] e baixar todos os arquivos compactados (.zip) com o nome de “DADOS_ABERTOS_CNPJ_”(índice do arquivo)”.zip”.

Aproveitou-se para preparar para casos futuros em que a quantidade de arquivos aumente (para 21 por exemplo) testando de um em um até o link falhar.

Como a banda do site da receita é bem ruim, foi feito um tratamento extensivo de erros para garantir que ao final da etapa os dados estarão baixados de maneira idônea, apesar dos prováveis falhas de conexão com o servidor da receita federal.

Tarefa 2 – Extração dos dados

Rotina: **extract_data**

Com o arquivo compactado e com o auxílio da biblioteca [zipfile](#) [3], realizamos a extração do arquivo .csv guardado dentro do arquivo .zip. O arquivo resultante tem seu nome capturado e salvo no arquivo “file_temp_cod”(índice do arquivo)“.tmp”. Terminada a tarefa, o arquivo compactado é excluído, através da biblioteca [os](#) [4].

Tarefa 3 – Tratamento dos dados

Rotina: **analyze_data**

Após o arquivo original ser descompactado, o mesmo é lido e analisado para que os registros de sócios sejam separados dos registros relativos às informações das empresas. Nesse momento, também, o arquivo passa de posicional para separado por tabulações, em que o arquivo resultante para as informações das empresas tem 35 colunas e o arquivo resultante para as informações dos sócios tem 12 colunas. Terminada a leitura, o arquivo original, posicional, é excluído, via os [4].

Tarefa 4 – Importação da tabela no banco SQL Server

Rotina: **copy_to_sqlserver**

De posse do nome do arquivo final (via “file_temp_cod”(índice do arquivo)“.tmp”) e com o auxílio da biblioteca pyodbc [5], é possível conectar no banco SQL Server e importar os dados da tabela para o banco. Essa tarefa ocorrerá duas vezes por arquivo, uma para importar os dados do CNPJ e outra para importar os dados dos sócios. Terminada a importação, os arquivos são excluídos, assim como o .tmp, via os [4].

Foi realizado o download do pacote de instalação para Linux do driver de conexão (Driver ODBC [6] para SQL Server 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'
```

Devido ao tempo de demora nessa etapa, foram tomadas duas decisões de projeto importantes para aceleração da mesma: não se preocupar com perfeição na inserção (pelo menos em um primeiro momento) e processamento paralelo na inserção de dados.

Sobre a primeira questão, enquanto nas outras bases a inserção é um processo rápido que não leva mais que alguns minutos, nesta base é um processo muito longo. As tentativas de utilizar uma função análoga levavam dias para inserir alguns dos 20 arquivos. Além disso, o fato de serem 20 arquivos, cada um com milhões de registros torna quase inevitável o surgimento de excessões que interrompem o processo e necessitam uma alteração na tabela e a reinicialização do processo de inserção.

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.

Enquanto que nas outras bases entregues o tempo que o sistema leva para rodar rotinas de testes é de alguns minutos, nessa base leva pelo menos algumas horas com possibilidade de levar dias.

Frente à isso, foi decidido que a inserção tem que funcionar para o máximo de casos possível e as exceções que se apresentarem vão para um arquivo separado. Existem 8 exceções dentro de 42M de CNPJs e 339 exceções dentro de 26M de sócios. Como um todo, cerca de 5 exceções a cada milhão de registros. Esse nível foi considerado aceitável para as rotinas de tratamento de erro e um arquivo com os registros que apresentam erro ficará disponível.

Sobre o segundo problema, em testes preliminares, chegamos a esperar 7h para a inserção de 1 dos 20 arquivos de CNPJ no banco utilizando processo similar às outras bases. Embora isso seja suficiente para configurar sucesso, essa performance foi considerada inadequada e foram tomadas diversas medidas para melhorar essa performance.

Dessa forma, essa base não utiliza a biblioteca pandas para facilitar o acesso aos dados, para minimizar a utilização de RAM e CPU e permitir uma paralelização do processo. Mapeia-se cada arquivo a ser processado em um processo e os mesmos são postos em uma lista para serem executados à medida que o sistema operacional permitir. Na máquina de testes, por ter 4 cores, o S.O. faz quatro processos por vez. Isso significa 4 conexões com o banco e a inserção de 4 arquivos simultaneamente.

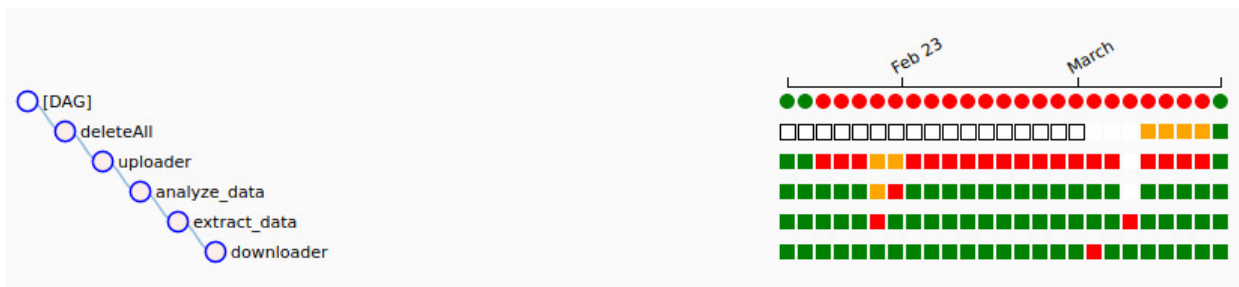
Tarefa 5 – Apagar os arquivos intermediários

Rotina: **delete_all**

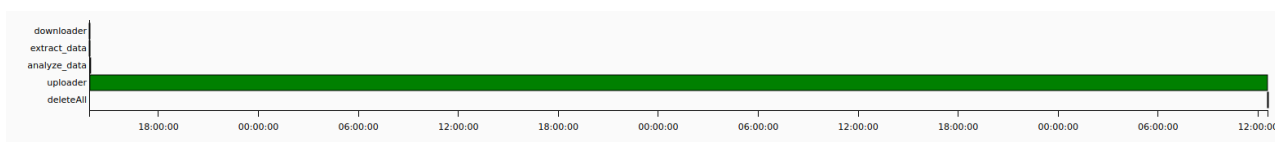
Depois de inserir os dados no banco de dados, uma tarefa apaga os arquivos intermediários baixados e libera o disco.

6. DURAÇÃO DAS ROTINAS ETL

O processo de ETL desta base dura dias. A estimativa é que todas as tarefas durem por volta de 7 dias no total para serem executadas.



A imagem a seguir apresenta a rotina de inserção dos dados no banco.



Uma estimativa mais realista do tempo é:

- 3 dias para baixar os dados (difícil melhorar, o problema é o servidor da Receita Federal)
- 1 dia para processar os dados (em uma máquina com SSD e bastante memória, pode ser bem mais rápido, levando poucas horas)
- 70h e 40 min dias para inserir os dados no banco (se o banco estiver na mesma rede do airflow, com uma conexão boa esse tempo pode ser bem mais rápido e com mais cores de CPU disponíveis pode executar mais arquivos por vez)

7. FLUXO DE TRATAMENTO DE ERROS

Os possíveis erros de execução das tarefas foram tratados a 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 que contém os dados para download, a tarefa inicial do ETL é fazer uma procura dinâmica pela existência e validade dos links para *download* das bases, de forma a evitar erros gerados por mudanças estruturais na página de download do arquivo.

Alguns registros se apresentam de forma não estruturada, de acordo com o padrão do resto dos registros da base. Esses casos, que representam 347 em 68 milhões de registros, não são importados no banco para evitar informações incorretas nas colunas das tabelas mas são salvos em arquivos .txt caso a equipe queira tratar esses dados no futuro.

8. FLUXO DE AGENDAMENTO DE ROTINAS

A base do CNPJ é atualizada trimestralmente. Dessa forma, no código de configuração do Apache Airflow, foi inserida uma rotina de atualização da base a cada 90 dias.

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 mensais. Como a atualização é feita trimestralmente e os arquivos de download são totalmente substituídos, não houve tempo hábil para estimar qual é o volume incremental da base à cada atualização.

10. AUXÍLIO NOS ESTUDO DE COMPRAS PÚBLICAS

Com os dados da base do CNPJ, é possível realizar o cruzamento entre os dados de porte, capital social, sócios, entre outras informações identificatórias das empresas relacionadas e os novos convênios, contratos de repasse ou termos de parceria com a administração pública federal, afim de conhecer mais sobre os fornecedores / parceiros do Governo Federal assim como validar informações dadas pelas mesmas, e também avaliar a capacidade de prestação de serviço das mesmas.

11. EVIDÊNCIA DOS DADOS IMPORTADOS

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

*DADOS_RECEITA_SOC dbo DADOS_RECEITA_CNPJ <SQL Server - maste

Properties Data ER Diagram

SQL Server - master CNPJ_RFB Schemas dbo Tables DADOS_RECEITA

DADOS_RECEITA_CNPJ Enter a SQL expression to filter results (use ▶) ▼

Grid	ABC CNPJ	123 IDENTIFICADOR_MATRIZ_FILIAL	ABC RAZAO_SOCIAL
1	05144075000160	1	AKIRA HASEGAWA
2	05144076000104	1	SERVICO AUTONOMO DE AGUA
3	05144077000159	1	A. C. ADRIANO SILVA
4	05144078000101	1	LUCIMAR T.M. DO NASCIMENTC
5	05144079000148	1	MARIA APARECIDA DE OLIVEIRA
6	05144080000172	1	AGROSERTAO - ASSESSORIA, CC
7	05144081000117	1	JORNAL NOVA CIDADE DE ORLA
8	05144082000161	1	J. L. A. MACHADO
9	05144083000106	1	DANIEL RIBEIRO DA SILVA - SUC
10	05144084000150	1	MADEIREIRA RIO DO OURO INDI
11	05144085000103	1	CIRCUITO CENTRO AUTOMOTIV
12	05144086000140	1	CONDOMINIO RESIDENCIAL ILH
13	05144087000194	1	TOP CONFECCOES LTDA
14	05144088000139	1	KART CLUBE RIO BRANCO
15	05144089000183	1	MASTEFARMA MERCANTIL ITT

Save Cancel Script 50 42.515.058

*DADOS_RECEITA_SOC dbo DADOS_RECEITA_CNPJ <SQL Server - maste

Properties Data ER Diagram

SQL Server - master CNPJ_RFB Schemas dbo Tables DADOS_RECEIT

DADOS_RECEITA_SOCIOS Enter a SQL expression to filter results (u ▶) ▼

Grid	ABC CNPJ	123 ID_SOCIO	ABC NOME_SOCIO	ABC CN
1	de 68M	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
2	26025213001133	2	PAULO AFONSO DE FARIA MAIA	000**
3	26025213001133	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
4	26025213001214	2	PAULO AFONSO DE FARIA MAIA	000**
5	26025213001214	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
6	26025213001303	2	PAULO AFONSO DE FARIA MAIA	000**
7	26025213001303	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
8	26025213001486	2	PAULO AFONSO DE FARIA MAIA	000**
9	26025213001486	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
10	26025213001567	2	PAULO AFONSO DE FARIA MAIA	000**
11	26025213001567	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
12	26025213001648	2	PAULO AFONSO DE FARIA MAIA	000**
13	26025213001648	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
14	26025213001729	2	MARIA DA CONCEICAO DE FARIA MAIA LEITE	000**
15	26025213001729	2	PAULO AFONSO DE FARIA MAIA	000**

Save Cancel Script 50 26.481.178

150 row(s) fetched - 8ms (+11ms)

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. Portal da Receita Federal. Página relativa ao CNPJ. Disponível em: <https://receita.economia.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/informacoes-gerais-sobre-o-cnpj>. Acesso em: 21 de outubro 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. Zipfile documentation. Disponível em: <https://docs.python.org/3/library/zipfile.html>. Acesso em: 21 de outubro de 2019.
4. Miscellaneous operating system interfaces. Disponível em: <https://docs.python.org/3.4/library/os.html>. Acesso em: 21 de outubro de 2019.
5. 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.
6. 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

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.