

# **Automatização do Processo de Transporte e Carga de Microdados Estatísticos (Censo Agropecuário 2017)**

**João Marcelo dos Santos Marques**

**Eduardo da Costa Romero**

**André Luiz da Fonseca Carvalho**

## Objetivo

*Relatar a **automatização** do processo de carga dos microdados estatísticos do Censo Agropecuário de 2017 do IBGE, através da utilização de programas que atuam diretamente no processo de ETL (extration, transformation, loader), visando a disseminação de informações com maior **rapidez** para a elaboração de **políticas públicas** para a sociedade civil.*

## Tópicos

- Motivação
- Conceito de microdados estatísticos
- Transporte dos microdados coletados
- Carga de microdados nas bases de dados
  - Extração
  - Transformação
  - Carga
- Conclusões
- Trabalhos Futuros

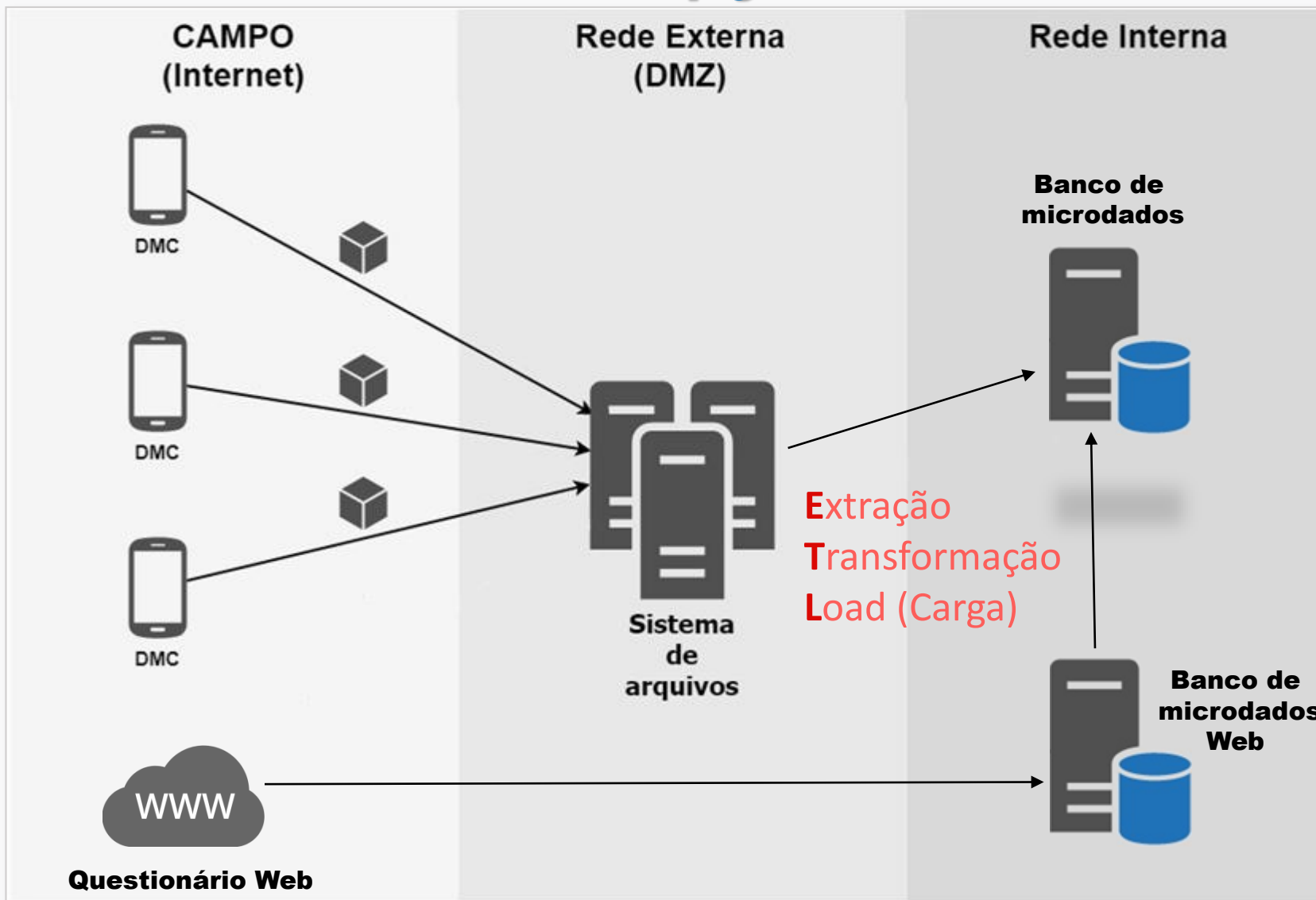
## Motivação

- Muitas organizações ainda enfrentam **desafios** na sua linha de produção no que tange ao processo de *ETL* em **base de dados** para disseminá-las.
- Um dos desafios para a execução de uma pesquisa de abrangência nacional é a **criação** de uma infraestrutura de comunicação de dados de forma **centralizada** nos postos de coleta para realizar a **transmissão** dos dados.

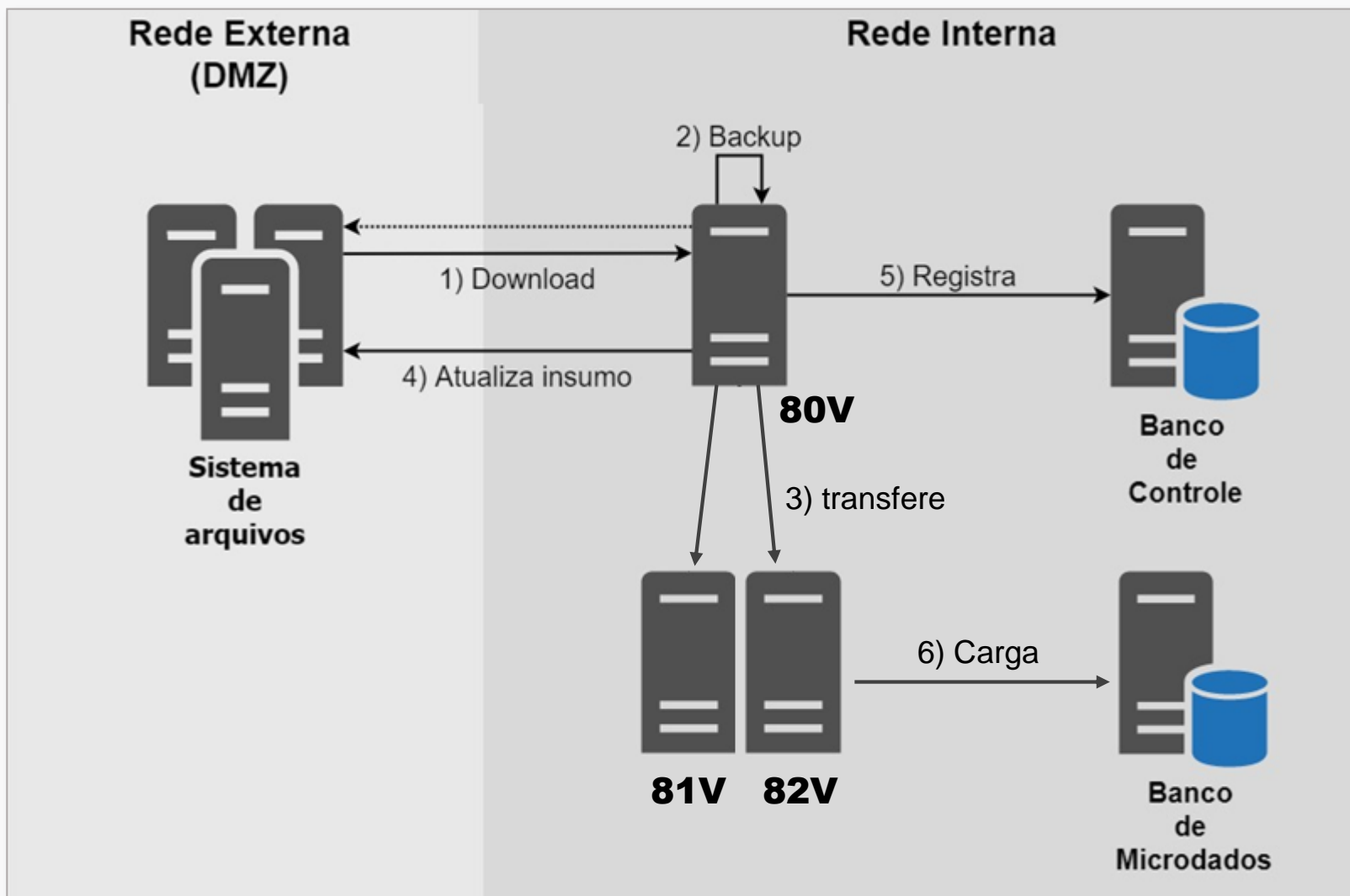
## Conceito de microdados

*Representa o menor nível de desagregação dos dados de uma pesquisa, retratando, sob a forma de códigos, o conteúdo dos questionários (portal do IBGE, 2017).*

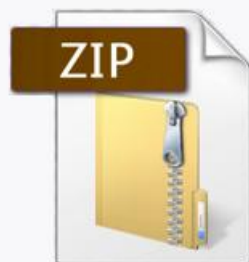
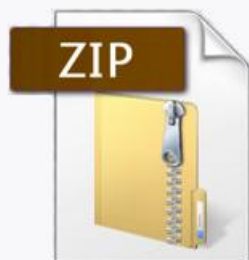
# Recepção dos microdados



## Recepção dos microdados coletados



## Extração



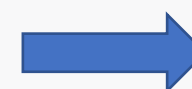
Descompactador dos Arqs de Entrada (v20140129)

Tipo arq	Total	Processados	Erros	Hora Inicial	Hora Final
01	0	0	0	11:42:09	11:42:09
02	0	0	0	11:42:09	11:42:09
03	52	52	0	11:42:09	11:42:29
04	3315	3315	0	11:42:30	11:55:29
05	0	0	0	11:55:29	11:55:29
06	0	0	0	11:55:29	11:55:29
07	3	3	0	11:55:29	11:55:31

E:\pesquisas\CENSO\_AGRD2017\rece\\*.zip ciclo: 00:00:05 modo: auto

Proxima Execução: 06/11/2017 11:55:36

Tempo restante para a próxima execução: 00:00:05





## Extração

↑ This PC > Local Disk (E:) > Pesquisas > CENSO\_AGRO2017 > recepcao > trabalho > 01 > 110001505000019

Name	Date modified	Type
110001505000000		
110001505000019		
110001505000020		
110001505000021		
110001505000022		
110001505000023		
110001505000024		
110001505000025		
110001505000026		
110001505000027		
110001505000028		
110001505000029		
110001505000030		
110001505000031		
110001505000032		
110001505000033		
110001505000034		
110001515000001		
110001515000002		
110001515000003		
110001520000000		
110001520000002		
110001520000003		
110001520000004		
110001525000000		
110001525000002		
110001525000003		
110001525000004		
110001505000019_0_12_12_1_01.csv	03/11/2017 14:04	CSV File
110001505000019_0_12_12_1_01_ASS.png	03/11/2017 14:04	PNG image
110001505000019_0_12_12_1_01_config.json	03/11/2017 14:04	JSON File
110001505000019_0_15_15_1_01.csv	03/11/2017 14:04	CSV File
110001505000019_0_15_15_1_01_config.json	03/11/2017 14:04	JSON File
110001505000019_0_17_17_1_01.csv	03/11/2017 14:04	CSV File
110001505000019_0_17_17_1_01_ASS.png	03/11/2017 14:04	PNG image
110001505000019_0_17_17_1_01_config.json	03/11/2017 14:04	JSON File
cep.csv	03/11/2017 14:04	CSV File
coordenada.csv	03/11/2017 14:04	CSV File
coordenadamutirao.csv	03/11/2017 14:04	CSV File
endereco.csv	03/11/2017 14:04	CSV File
enderecomutirao.csv	03/11/2017 14:04	CSV File
entorno.csv	03/11/2017 14:04	CSV File
especie.csv	03/11/2017 14:04	CSV File
face.csv	03/11/2017 14:04	CSV File
localidade.csv	03/11/2017 14:04	CSV File
logmutirao.csv	03/11/2017 14:04	CSV File
logradouro.csv	03/11/2017 14:04	CSV File
mapa.csv	03/11/2017 14:04	CSV File
mutirao.csv	03/11/2017 14:04	CSV File
partemutirao.csv	03/11/2017 14:04	CSV File
questcoordenada.csv	03/11/2017 14:04	CSV File
questionario.csv	03/11/2017 14:04	CSV File
resumosector.xml	03/11/2017 14:04	XML Document
rotaapp.csv	03/11/2017 14:04	CSV File
setor.csv	03/11/2017 14:04	CSV File

# Transformação

```
01;; [V011300]FAZENDA DO MEIO|4; [V011401]ALAOR BENTO RODRIGUES|4; [V01150000]2|4; [V01150102] (37) 99824-30  
02;; [V02010000]1|4; [V05010100]1|4; [V02020000]1|4; [V02190000]1|4; [V02200000]63|4; [V02240000]4|4; [V02210  
03;; [V01170100]34,5|4; [V01170200]24|4; [V03030000]34,5|4; [VW03030000]34,5|0;  
04;; [V01170100]34,5|4; [V01170200]24|4; [V04020001]2|4; [V04020000]2|4; [V04030000]2|4; [V04050001]2|4; [V04  
05;; [V05020100]2|4; [V05030100]2|4; [V05020300]1|4; [V05020400]1|4; [V05020500]2|4; [V05020600]1|4; [V050301  
06;; [V06010100]1|4;  
07;; [V07010100]2|4; [V07020100]1|4; [V07021201]1|4; [V07021100]2|4;  
08;; [V09000001]1|4; [V09000003]1|4; [V09010101]0,0|4; [V09010102]0,0|4; [V09010103]0,0|4; [V09010100]1|4; [V  
09;; [V08010102]2|4; [V10020101]1|4; [V10020100]2|4; [V10030100]1|4; [VW08010306]2|0; [VW10040001]1|0; [VW100  
10;; [V13011400]2|4; [V13011500]1|4; [V13011600]2|4; [V13011700]1|4; [V13011800]2|4; [V13011900]1|4; [V130120  
11;; [V14010101]15|4; [V14010000]2|4; [V14130000]1|4; [V14160000]1|4;  
13;; [V16010101]1|4; [V16010000]2|4; [V16010102]1|4;  
15;; [V18010101]2|4; [V18010000]2|4; [V18010102]1|4;  
19;; [V22010101]30|4; [V22010000]2|4; [V22050000]1|4; [V22080000]1|4;  
28;; [V33010100]2|4; [V33010200]2|4; [V33010300]1|4; [V33010400]1|4; [V33010500]1|4; [V33010600]2|3; [V330107  
29;; [V34010100]240|4; P240 [V34020300]2|4; P240 [V34020400]24|4; P240 [V34020500]120|4; P240 [V34020700]006|4;  
30;; [V35010100]101:103:108:132:139|4; P101 [V35020600]2|4; P101 [V35020300]8|4; P101 [V35020602]2|4; P101 [V35  
34;; [V39010100]505|4; P505 [V39020300]2000|4; P505 [VX04100011]3|4; [V01170200]24|4;  
36;; [V41010100]1|4;  
37;; [V46010001]1|4; [V46010600]27600,0|4; [V46010700]24000|4; [V46010000]2|4; [V46010601]2|4;  
38;; [V43010100]1|4;  
39;; [V45010300]420|4; [V45010500]500|4; [V45011400]1200|4; [V45011500]1200|4; [V45010900]200|4; [V45011900]  
41;; [V49010000]1|4; [V51010000]2|4; [V51010100]287,875,956-72|4; [V51020100]3|4;  
42;; [V51000001]preenchido|4;  
40;;
```

# Transformação

```
01;;[V011300]SITIO SAO MIGUEL|4;[V011401]MIGUEL DA SILVA|4;[V01150102](37)99999-1234|2;
```



```
SITIO SAO MIGUEL;MIGUEL DA SILVA;(37)99999-1234
```

```
30;;[V35010100]108:112|4;P108[V35020600]2|4;P108[V35020300]12000|4;P108[V35020602]1|4;P108[V35020700]4|4;P108[V35020800]29|4;P112[V35020600]2|4;P112[V35020300]1|4;P112[V35020602]1|4;P108[V35020801]24200|0;P108[VW35020307]1239,669|0;P108[VW35020701]9,68|0;
```



```
110001505000021;0;77;77;108;12000;2;;1;4;29;24200;,,,,,,,,,,,,,9,68;1239,669;,,,,;
110001505000021;0;77;77;112;1;2;;1;,,,,,,,,,,,,,
```

108=Banana - Kg;  
112=Café Arábica em Grão - Kg;

# Transformação

## Estabelecimentos

Dados Gerais do Estabelecimento 1  
Dados Gerais do Estabelecimento 2  
Dados Gerais do Estabelecimento 3

## Lavouras Permanentes

Dados do Produto 3 do Estabelecimento 1  
Dados do Produto 4 do Estabelecimento 2

## Lavouras Temporárias

Dados do Produto 1 do Estabelecimento 1  
Dados do Produto 1 do Estabelecimento 3

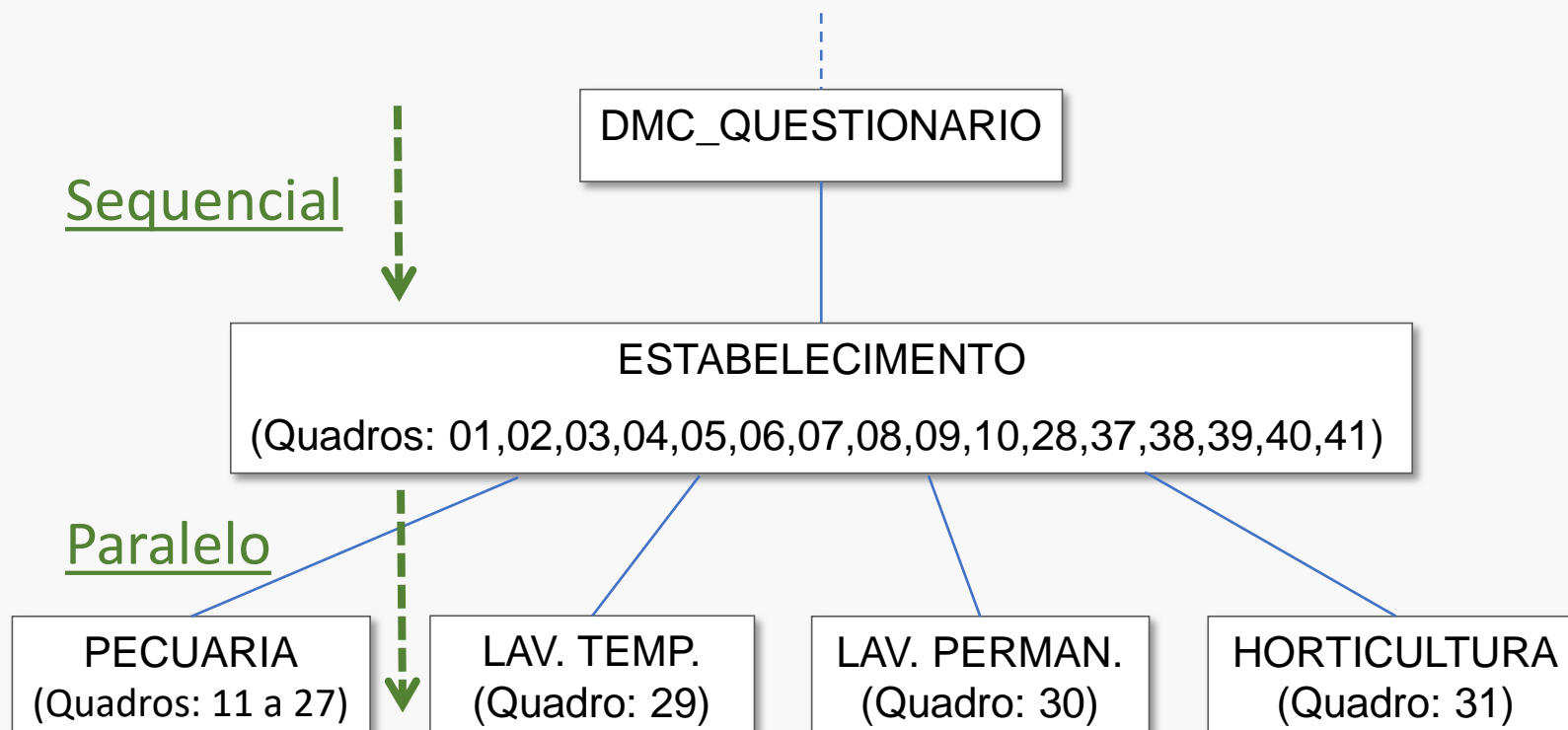
## Silviculturas

Dados do Produto 6 do Estabelecimento 3

## Horticulturas

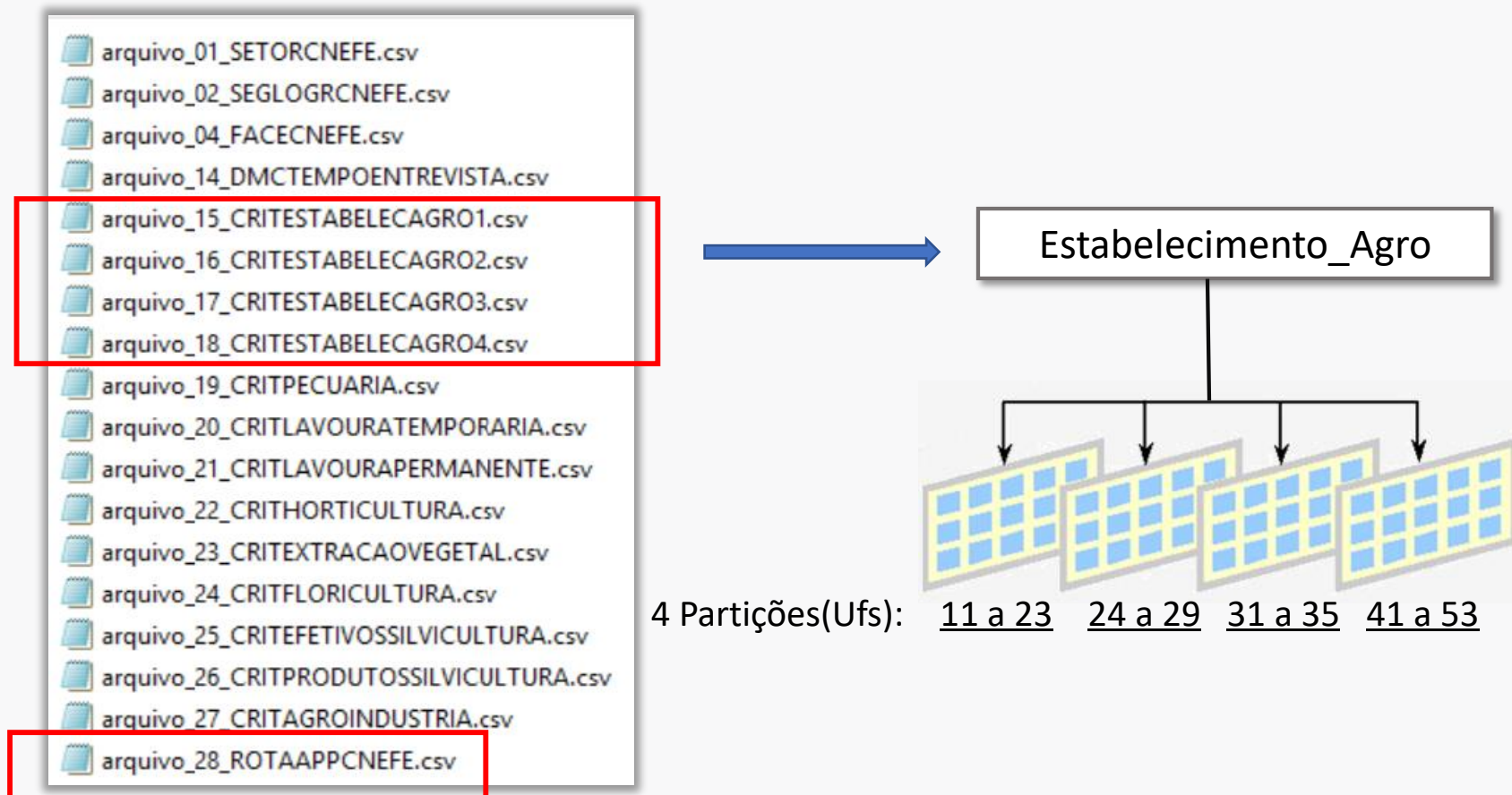
Dados Produto 5 do Estabelecimento 2

## Carga (SQLoader com paralelismo)



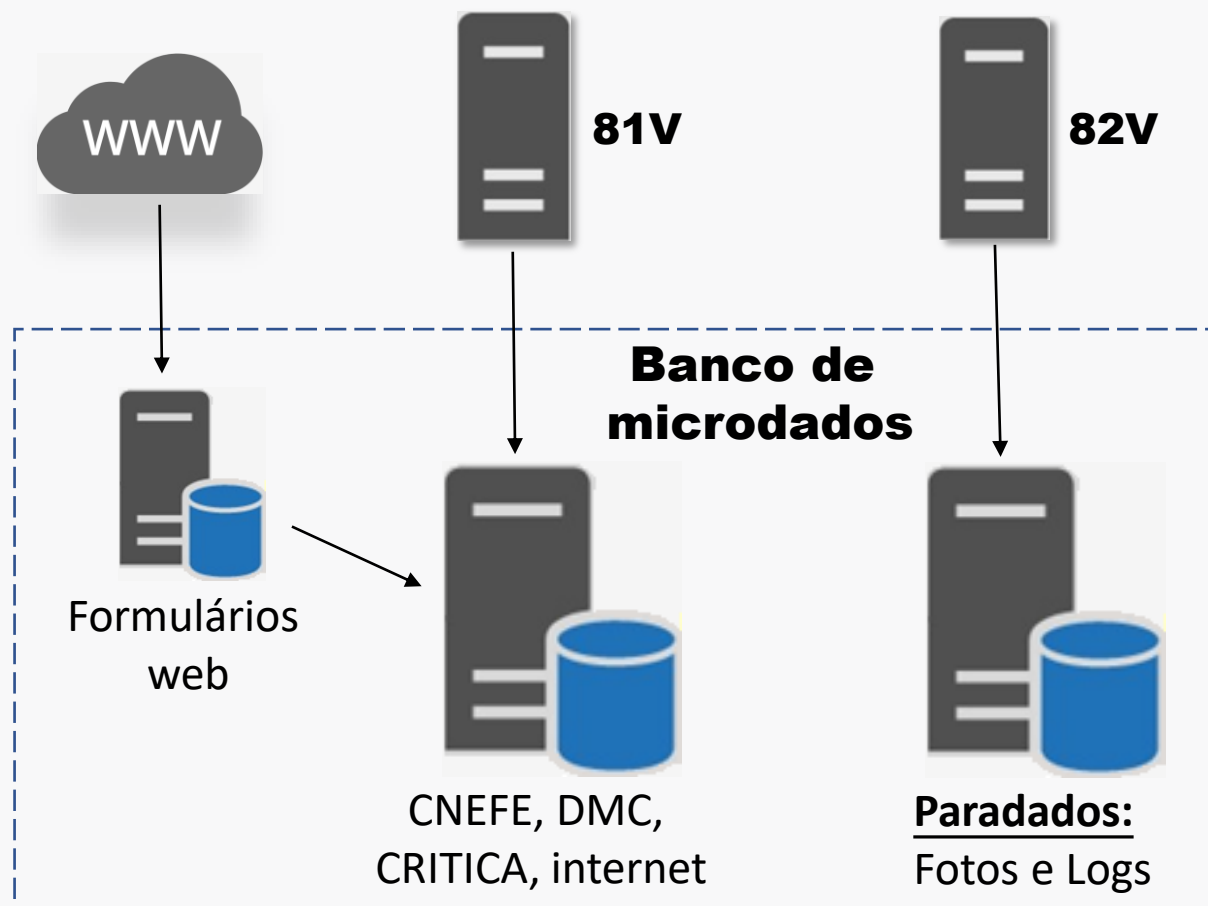
EXTR. VEGETAL (Quadro 32), FLORICULTURA (Quadro 33), EFET. SILVIC. (Quadro 34),  
PROD.SILVIC, (Quadro 35), AGROINDUSTRIA (Quadro 36)

## Carga (SQLoader com particionamento)



8 Partições(Ufs): 11 a 22, 23 a 28, 29, 31 e 32, 33 e 35, 41 e 42, 43, 50 a 53

## Carga (SQLoaders em 2 Bancos)



## Conclusões

1

A criação do programa **tradutor** no processo de *ETL* tem um papel fundamental na tradução das **tags** dos quadros dos questionários dos arquivos recebidos nas transmissões para a **carga** em banco de dados.

2

Este processo é de grande importância para o IBGE, uma vez que ele **produz** as informações necessárias para a **geração** de relatórios de acompanhamento da coleta e das bases de dados dos sistemas institucionais.



## Trabalhos Futuros



1

A **construção** de uma aplicação *web*, a fim de que todos os profissionais envolvidos possam acompanhar em tempo real todas as transmissões e monitorar o **processo de carga de microdados**.



2

Divulgar os **paradados** via *API* para a geração de relatórios gerenciais que sirvam de insumos para a melhoria do processo de transporte e carga de microdados e para o **planejamento** de censos futuros, identificando **erros** ou dificuldades ocorridas durante a coleta de dados.

## Contatos:

**[eduardo.romero@ibge.gov.br](mailto:eduardo.romero@ibge.gov.br)**

**+55 21 2142-0711**

**[andre.carvalho@ibge.gov.br](mailto:andre.carvalho@ibge.gov.br)**

**+55 21 2142-0555**

**[joao.marques@ibge.gov.br](mailto:joao.marques@ibge.gov.br)**

**+55 21 2142-4595**