

Dependências

MongoDB 3.6

<https://www.mongodb.com/download-center/community>

Python 3.x

<https://www.python.org/downloads/>

blast-dbf

<https://github.com/eaglebh/blast-dbf>

Script de conversão de DBC para DBF. Para instalá-lo é necessário fazer o download e descomprimir o arquivo em um diretório, em seguida o comando deve ser executado:

```
$make blast-dbf
```

Um executável com o nome **blast-dbf** será criado, em seguida é necessário movê-lo para o PATH do sistema. É necessário que o comando esteja disponível ao se digitar blast-dbf no terminal.

Uma versão obtida em 15/07/19 está disponível no arquivo **blast-dbf-master.zip**.

Dbf2csv

<https://github.com/akadan47/dbf2csv>

Script de conversão de DBF para CSV. Para instalá-lo é necessário fazer o download e descomprimir o arquivo em um diretório, em seguida o comando deve ser executado com o python 3 (em alguns sistema esse deve ser invocado com o comando python3 ao invés de simplesmente python):

```
$sudo python setup.py install
```

É necessário que o comando esteja disponível ao se digitar dbf2csv no terminal.

Uma versão obtida em 15/07/19 está disponível no arquivo **dbf2csv-master.zip**.

Arquivo de Entrada

Um arquivo de entrada é um arquivo no formato JSON que possui todas as especificações sobre o ETL a ser feito. Na pasta do projeto o arquivo *input.json* possui um exemplo de arquivo de entrada, já os arquivos *sih.json* e *sinasc.json* possuem o modelo das especificações para o ETL de SIH e SINASC.

Um arquivo de entrada pode conter os seguintes campo:

- base – Pode assumir SIM, SINASC ou SIH e especifica a base.
- Ano – Especifica o ano do qual a base deve ser obtida.
- Database – Nome da database no MongoDB em que a collection final do ETL será colocada.
- Collection – Nome da collection final do Etl no MongoDB.
- Campos – Especifica cada campo inserido na collection final, é um array onde cada campo contém o nome do campo e pode conter tabelas_auxiliares e campo_auxiliar_juncao caso o campo seja de junção.
- Campo – Nome do campo da base a ser inserida no MongoDB
- tabelas_auxiliares – Array com uma ou mais tabelas auxiliares para a execução da junção no ETL.
- campo_auxiliar_juncao – Campo da tabela auxiliar usado na junção.

Execução do ETL

Dado um arquivo de entrada *arquivo.json* com todas as especificações do ETL, a execução é invocada com o comando:

```
$/etl arquivo.json
```

O ETL também pode ser invocado explicitando o uso do python 3(em alguns sistema esse deve ser invocado com o comando `python3` ao invés de simplesmente `python`):

```
$python etl arquivo.json
```