Live de Python #49

Serializando objetos com pickle e shelve

Nome:
Eduardo Mendes
Instituição:
Unicamp / Diebold Nixdorf
Contatos:
{facebook, github, gist
instagram, linkedin,
telegram, twitter}/dunossauro

Roteiro

- O que é serialização?
- Por que serializar?
- Tipos de serialização nativas em Python
- Usando Pickle
- Usando Shelve
- Entendendo o Pickle

O que é serialização?

Tradução de estruturas de dados, ou objetos, falando em python, em um formato que pode ser armazenado. Armazenado em arquivos, buffers, bancos de dados, ...

Por que serializar?

- Tornar objetos/estruturas persistentes
- Comunicação web (socket, http)
- Distribuição de “estados”

Tipos de serializadores em Python

- Json
- Struct
- Pickle
- Shelve
- Marshal

Usando Pickle

O que é o pickle?

Pickles é uma biblioteca nativa para serialização em strings para objetos.

```
In [1]: import pickle

In [2]: lista = [1, 2, 3, 4]

In [3]: pickle.dumps(lista)
Out[3]: b'\x80\x03]q\x00(K\x01K\x02K\x03K\x04e.'

In [4]: lista = [1, 2, 3, 4, 'live de python']

In [5]: pickle.dumps(lista)
Out[5]: b'\x80\x03]q\x00(K\x01K\x02K\x03K\x04X\x0e\
x00\x00\x00live de pythonq\x01e.'
```

Funções de módulo

- dump
  - Faz o dump do objeto em um arquivo de texto (io.BytesIO)
- dumps
  - Faz o dump do objeto uma string
- load
  - Faz o load do objeto em um arquivo de texto (io.BytesIO)
- loads
  - Faz o load do objeto uma string

Não, eu não esqueci que dá pra ser mais rápido

No python2 existia uma interface em C para o pickles que segundo a documentação pode chegar a ser 1000 vezes mais rápido. Porém, não é possível estender a mesma, por ela é em C.

Fora isso, ainda temos grande dificuldades em debugar o código.

No Python3 o import é chamado com ‘_pickle’

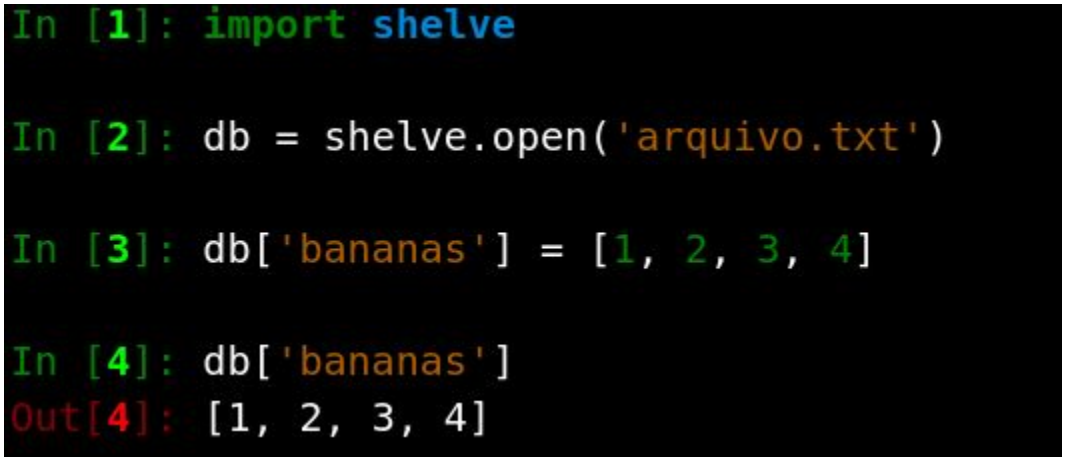
Usando Shelve

O que é o Shelve?

Shelve é uma biblioteca também nativa e que faz uso do Pickle. Ele cria uma estrutura baseada em MutableMaping, ou seja, se parece com um dicionário. Provem uma maneira “melhor estruturada” para para trabalhar com os dados serializados.

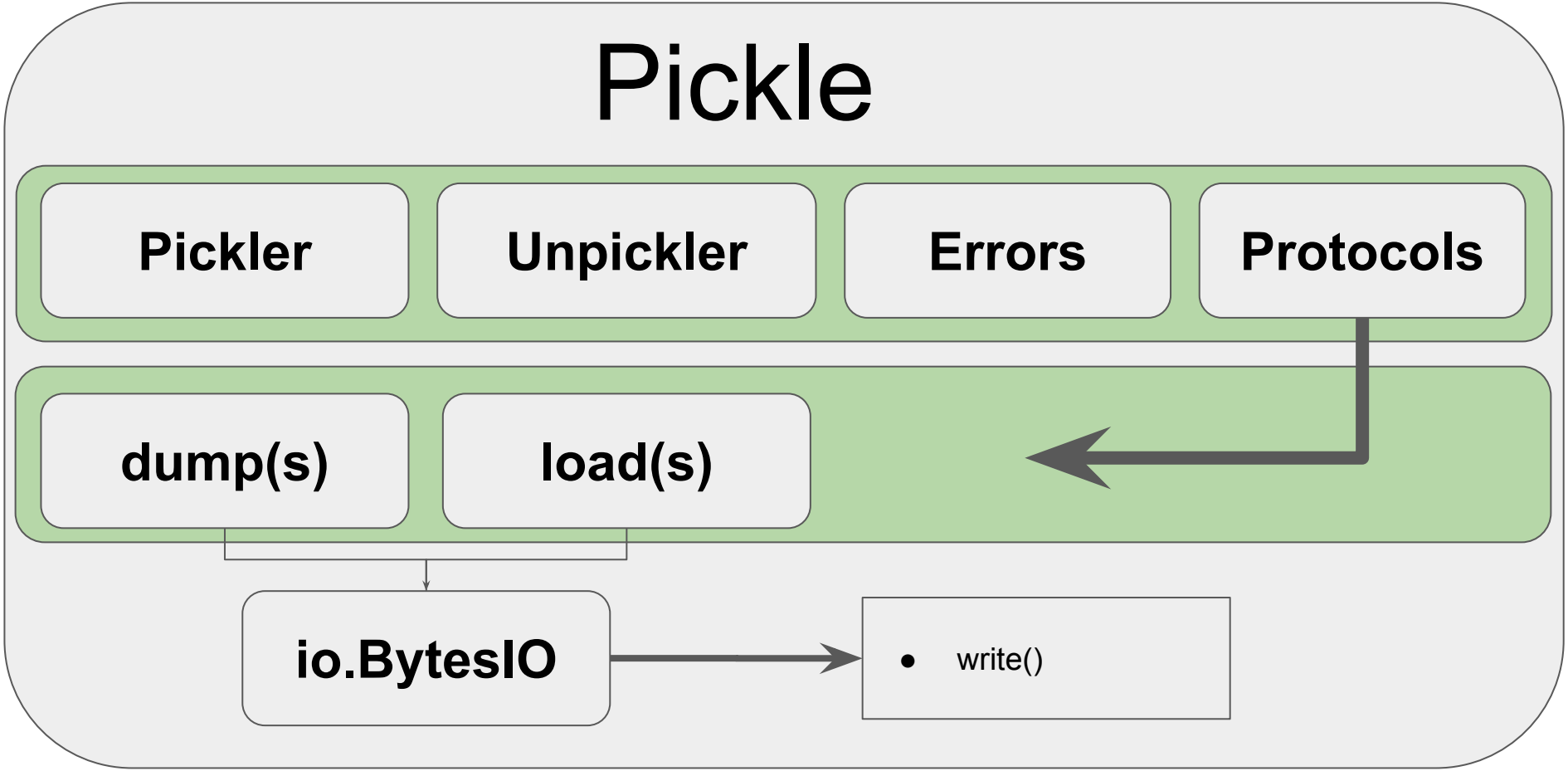
O que é o Shelve?

Ou seja, diferente de persistir usando unicamente o pickle, que força chamadas de load até que seja esgotada a leitura. Ele carrega todos os objetos e todos podem ser acessados usando a notação de dicionários



Entendendo o Pickle

Abstração


Pickle
Pickler
Unpickler
Errors
Protocols
dump(s)
load(s)
io.BytesIO
write()

O que pode ser serializado?

- None, True e False
- Numbers
- Strings, bytes e bytearrays
- tuplas, listas, conjuntos e dicionários contendo somente um objeto serializável
- Funções definidas anteriormente (usando def e não lambda)
- Funções built-in
- Classes definidas anteriormente

Protocolos

Existem diferentes protocolos para escrita do stream de texto, algumas vantagens e desvantagens existem em adotar alguma delas

- 0: Transforma exatamente o objeto transcrito para binário, compatível com a grande maioria de versões do python
- 1: Formato binário antigo, também compatível com versões anteriores
- 2 (PEP 307): Otimização da versão do binário gerado (Python 2.3)
- 3 (Python 3.0): Utilização dos Byte objs. Versão **default**, compatível com todas as versões 3 do Python, não aberto pelo python2
- 4 (PEP 3154): Otimização para arquivos grandes e otimização no formato da codificação, melhor opção, mas não compatível (Python 3.4)

Protocolos

```
pickle.dumps([1,2,3], protocol=0)
b'(lp0\nL1L\naL2L\naL3L\na.'

pickle.dumps([1,2,3], protocol=1)
b']q\x00(K\x01K\x02K\x03e.'

pickle.dumps([1,2,3], protocol=2)
b'\x80\x02]q\x00(K\x01K\x02K\x03e.'

pickle.dumps([1,2,3], protocol=3)
b'\x80\x03]q\x00(K\x01K\x02K\x03e.'

pickle.dumps([1,2,3], protocol=4)
b'\x80\x04\x95\x0b\x00\x00\x00\x00\x00\x00\x00]\x94(K\x01K\x02K\x03e.'
```

Erros

O pickle cria alguns modelos de erros:

- PickleError: Base para todos os erros, ou seja, todos geram PickleError
- PicklingError: Erro para serializar
- UnpicklingError: Erro para desserializar

```
try:
    dumped = pickle.dumps([1,2,3], protocol=4)
    print(dumped)
    loaded = pickle.loads(dumped)
    print(loaded)
except pickle.PickleError:
    print('deu ruim em algo')
except pickle.PicklingError:
    print('deu ruim para serealizar')
except pickle.UnpicklingError:
    print('deu ruim para desserealizar')
```

pickletools

Dúvidas?

Nome:
Eduardo Mendes
Instituição:
Unicamp / Diebold Nixdorf
Contatos:
{facebook, github, gist
instagram, linkedin,
telegram, twitter}/dunossauro