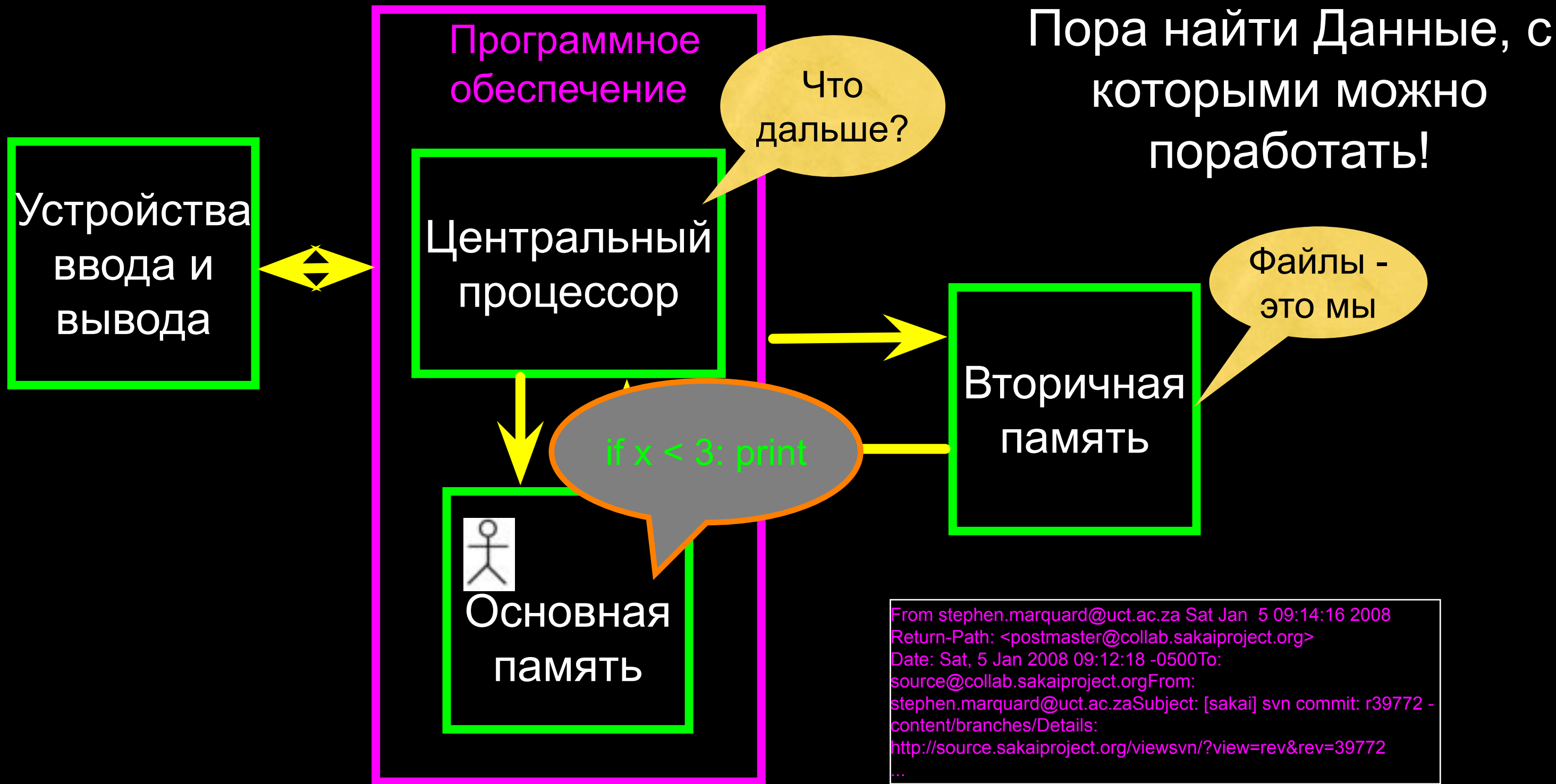


# Работа с файлами



# Работа с файлом

Текстовый файл можно представить как последовательность строк:

```
From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008
```

```
Return-Path: <postmaster@collab.sakaiproject.org>
```

```
Date: Sat, 5 Jan 2008 09:12:18 -0500
```

```
To: source@collab.sakaiproject.org
```

```
From: stephen.marquard@uct.ac.za
```

```
Subject: [sakai] svn commit: r39772 - content/branches/
```

```
Details: http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772
```

# Открытие файла

- Прежде чем мы сможем прочитать содержимое файла, нужно сообщить Пайтону, с каким файлом мы собираемся работать и что будем с ним делать
- Функция `open()` служит для открытия файлов
- Функция `open()` возвращает «**дескриптор файла**» — это переменная, используемая для операций с файлом
- Аналогично операции “Файл -> Открыть” в текстовом редакторе

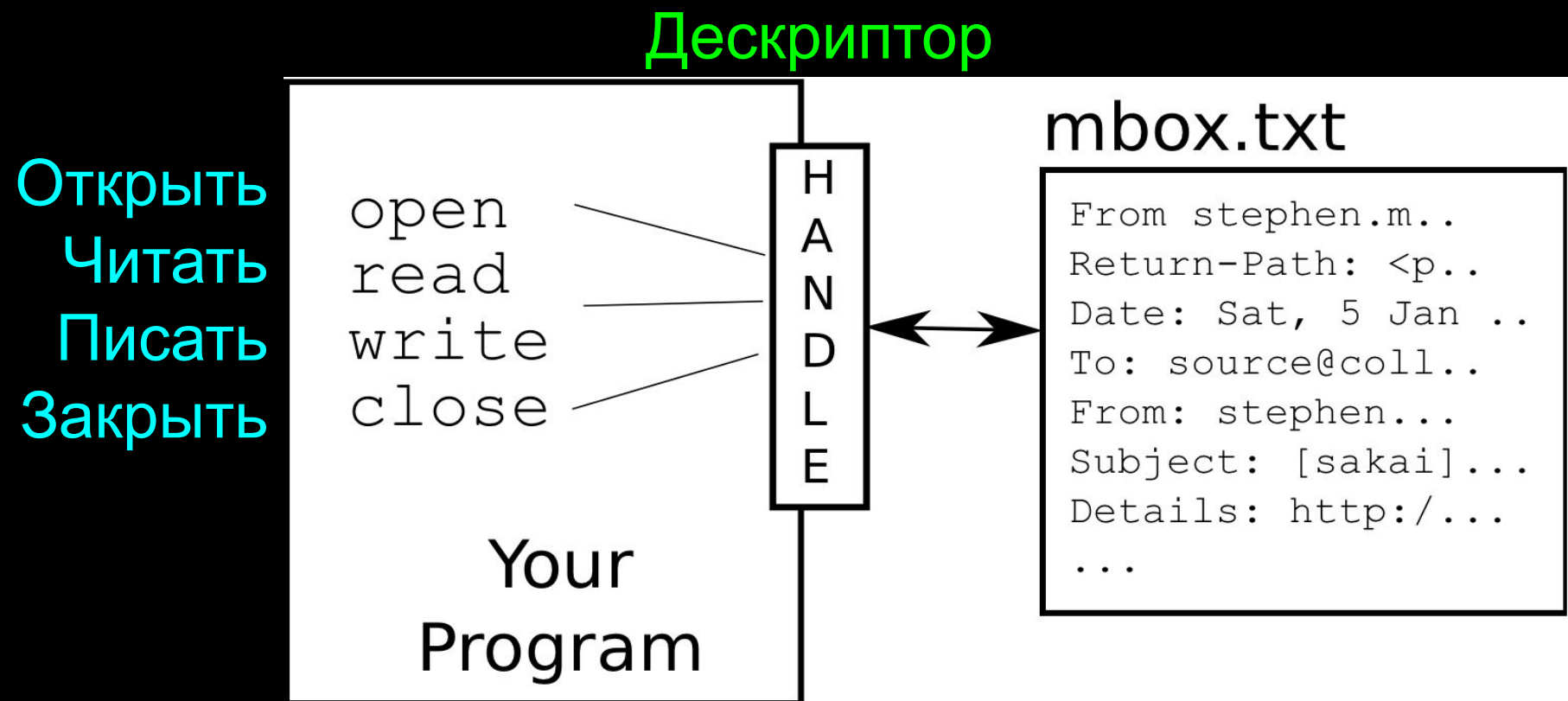
# Использование `open()`

```
fhand = open('mbox.txt', 'r')
```

- `handle = open(filename, mode)`
- возвращает дескриптор для управления файлом
- имя файла (`filename`) — это строка
- указание режима (`mode`) не является обязательным, но он должен быть `'r'`, если мы собираемся читать файл, и `'w'`, если собираемся записывать в файл

# Что такое дескриптор?

```
>>> fhand = open('mbox.txt')
>>> print(fhand)
<_io.TextIOWrapper name='mbox.txt' mode='r' encoding='UTF-8'>
```



Ваша программа

# Когда файл отсутствует

```
>>> fhand = open('stuff.txt')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
FileNotFoundError: [Errno 2] No such file or
directory: 'stuff.txt'
```

Данный файл или каталог отсутствует



# Символ перевода строки

- Чтобы обозначить конец строки, используется специальный символ — «перевод / разрыв строки»
- В строках пишется `\n`
- Перевод строки (`\n`) — это один символ, а не два

```
>>> string = 'Привет, \nМир!'
>>> string
'Привет, \nМир!'
>>> print(string)
Привет,
Мир!
>>> string = 'X\nY'
>>> print(string)
X
Y
>>> len(string)
3
```



# Работа с файлом

Текстовый файл можно представить как последовательность строк:

```
From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008
```

```
Return-Path: <postmaster@collab.sakaiproject.org>
```

```
Date: Sat, 5 Jan 2008 09:12:18 -0500
```

```
To: source@collab.sakaiproject.org
```

```
From: stephen.marquard@uct.ac.za
```

```
Subject: [sakai] svn commit: r39772 - content/branches/
```

```
Details: http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772
```

# Работа с файлом

Текстовый файл содержит символ перевода строки в конце каждой строки:

```
From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008\nReturn-Path: <postmaster@collab.sakaiproject.org>\nDate: Sat, 5 Jan 2008 09:12:18 -0500\nTo: source@collab.sakaiproject.org\nFrom: stephen.marquard@uct.ac.za\nSubject: [sakai] svn commit: r39772 - content/branches/\n\n\nDetails: http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772\n
```

# Чтение файлов в Пайтон

# Дескриптор файла как последовательность

- **Дескриптор файла**, открытый для чтения, можно рассматривать как **последовательность** строк, где каждая строка файла — строка в последовательности
- Мы можем использовать оператор **for**, чтобы **последовательно** перебирать строки
- Помните: **последовательность** — это упорядоченный набор

```
xfile = open('inbox.txt')  
for s in xfile:  
    print(s)
```

# Подсчет строк в файле

- Открываем **файл** в режиме «только для чтения»
- Используем цикл с **for**, чтобы прочесть каждую строку
- **Подсчитываем** строки и выводим количество строк

```
fhand = open('mbox.txt')
count = 0
for line in fhand:
    count = count + 1
print('Всего строк:', count)
```

```
Всего строк: 132045
```

# Чтение файла \*целиком\*

Мы можем **прочитать** весь файл (всё содержимое и переводы строк) в **одну строку**

```
>>> fhand = open('mbox-short.txt')
>>> inp = fhand.read()
>>> print(len(inp))
94626
>>> print(inp[:20])
From stephen.marquar
```

# Поиск по файлу

С помощью оператора **if** внутри цикла **for** можно обрабатывать только строки, соответствующие определённому критерию

```
fhand = open('mbox-short.txt')
for line in fhand:
    if line.startswith('From:') :
        print(line)
```

# Ой!

Что здесь делают все  
эти пустые строки?

From: `stephen.marquard@uct.ac.za`

From: `louis@media.berkeley.edu`

From: `zqian@umich.edu`

From: `rjlowe@iupui.edu`

...



# Ой!

Что здесь делают все эти пустые строки?

- В конце каждой строки файла находится **СИМВОЛ перевода строки**
- Оператор **print** добавляет **СИМВОЛ перевода строки** к каждой строке

```
From:
stephen.marquard@uct.ac.za\n
\n
From:
louis@media.berkeley.edu\n
\n
From: zqian@umich.edu\n
\n
From: rjlowe@iupui.edu\n
\n
...
```

# Поиск по файлу (исправленный)

- Мы можем удалить пробелы с правой стороны строки, используя `rstrip()` из библиотеки работы со строками
- Строка, содержащая символ перевода строки, считается «пустым пространством» и **удаляется**

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if line.startswith('From:') :
        print(line)
```

```
From: stephen.marquard@uct.ac.za
From: louis@media.berkeley.edu
From: zqian@umich.edu
From: rjlowe@iupui.edu
....
```

# Пропуск с использованием `continue`

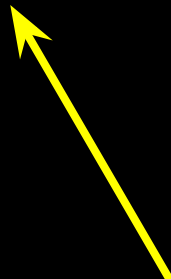
Если необходимо, то можно пропустить строку, используя оператор `continue`

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if not line.startswith('From:') :
        continue
    print(line)
```

# Используем `in`, чтобы выбрать строки

Мы можем поискать подстроку `в (in)` любой строке (`line`), используя соответствующий критерий поиска

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if not '@uct.ac.za' in line :
        continue
    print(line)
```



```
From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008
X-Authentication-Warning: set sender to stephen.marquard@uct.ac.za using -f
From: stephen.marquard@uct.ac.za
Author: stephen.marquard@uct.ac.za
From david.horwitz@uct.ac.za Fri Jan 4 07:02:32 2008
X-Authentication-Warning: set sender to david.horwitz@uct.ac.za using -f...
```

```
fname = input('Введите имя файла: ')
fhand = open(fname)
count = 0
for line in fhand:
    if line.startswith('Subject:') :
        count = count + 1
print('Найдено', count, 'строк, содержащих subject в',
fname)
```

## Запрос имени файла

Введите имя файла: **mbox.txt**  
Найдено 1797 строк, содержащих subject в  
mbox.txt

Введите имя файла: **mbox-short.txt**  
Найдено 27 строк, содержащих subject в  
mbox-short.txt

# Неверное имя файла

```
fname = input('Введите имя файла: ')
try:
    fhand = open(fname)
except:
    print('Неверное имя файла:', fname)
    quit()

count = 0
for line in fhand:
    if line.startswith('Subject:') :
        count = count + 1
print('Найдено', count, 'строк, содержащих subject  
в', fname)
```

Введите имя файла: mbox.txt

Найдено 1797 строк, содержащих subject в mbox.txt

Введите имя файла: na na boo boo

Неверное имя файла: na na boo boo

# Задача. Сколько писем получено за каждый час

```
From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008
From louis@media.berkeley.edu Fri Jan 4 18:10:48 2008
From zqian@umich.edu Fri Jan 4 16:10:39 2008
From rjlowe@iupui.edu Fri Jan 4 15:46:24 2008
From zqian@umich.edu Fri Jan 4 15:03:18 2008
From rjlowe@iupui.edu Fri Jan 4 14:50:18 2008
From cwen@iupui.edu Fri Jan 4 11:37:30 2008
From cwen@iupui.edu Fri Jan 4 11:35:08 2008
From gsilver@umich.edu Fri Jan 4 11:12:37 2008
...
...
```

## Задача. Сколько писем получено за каждый час

```
fname = input('Введите имя файла: ')
```

```
try:
```

```
    fhand = open(fname)
```

```
except:
```

```
    print('Неверное имя файла:', fname)
```

```
    quit()
```

```
# From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008 -
```

```
так выглядит строка файла
```

```
counter = dict()
```

```
for line in fhand:
```

```
    if line.startswith('From'):
```

```
        parts = line.split()
```

```
        hour = parts[5].split(':')[0]
```

```
        counter[hour] = counter.get(hour, 0) + 1
```

```
for key, val in sorted(counter.items()):
```

```
    print("hour: ", key, "emails: ", val)
```



# Резюме

- Открытие файла, дескриптор файла
- Структура файла, символ перевода строки
- Построчное чтение файла с помощью цикла `for`
- Поиск строк в файле
- Чтение имен файлов
- Неверное имя файла