

# How to Use In-Memory Streams

Hayao Suzuki

PyCon JP 2020

August 28, 2020

# Who am I ?

お前誰よ

**Name** Hayao Suzuki (鈴木 駿)

**Twitter** @CardinalXaro

**Work** Python Programmer at iRidge, Inc.

# In-Memory Streams

# Stream?

そもそもストリームって何？

ストリームはファイルオブジェクトである。

# File Object?

## ファイルオブジェクトって何？

- `read()` や `write()` などのメソッドを持つオブジェクト
- ディスク上のファイルや別の場所にあるストレージ、入出力機器とやりとりができる

## ファイルオブジェクトたち

- 生バイナリファイル
- バッファ付きバイナリファイル
- テキストファイル

## テキストファイル

```
f = open("myfile.txt", "r")
```

## バッファ付きバイナリ

```
f = open("myfile.jpg", "rb")
```

## open 関数の裏側

open は何をしているのか？

OS のシステムコール API を呼ぶ



## 例：CSV に加工する

```
with open("events.csv", "w") as csv_file:
    fieldnames = ["title", "started_at", "ended_at"]
    writer = csv.DictWriter(csv_file, fieldnames)
    writer.writeheader()
    writer.writerows(events)
```

## 例 : Windows

- CreateFile (ファイルのアクセス権取得)
- QueryAllInformationFile (ファイル情報の取得)
- WriteFile (ファイルへ書き込む)
- CloseFile (ファイルを閉じる)

Process Monitor 経由で確認した。

## 例 : Ubuntu on WSL

- openat (ファイルのオープン)
- fstat (ファイル情報の取得)
- ioctl (デバイス制御)
- lseek (ファイルのシーク)
- brk (ブレークポイントの調整)
- write (ファイルへ書き込む)
- close (ファイルを閉じる)

strace 経由で確認した。

# 最後に笑うのは誰だ

## 最終的な成果物はどこに置く？

- ファイルをローカルに保存するのがゴールではない
- ファイルを AWS S3 などの外部に置きたい

# In-Memory Streams

# インメモリーストリーム

## インメモリーストリームとは

- `str` や `bytes` をファイルオブジェクトのように扱える
- 読み書き可能、ランダムアクセス可能

# StringIO

## StringIO

テキストファイルのためのインメモリストリーム

## 例：CSV を StringIO で取り扱う

```
import io
with io.StringIO() as csv_file:
    fieldnames = ["title", "started_at", "ended_at"]
    writer = csv.DictWriter(csv_file, fieldnames)
    writer.writeheader()
    writer.writerows(events)
```