

俺のWSL生活

たいせつなデータとバックアップの話

令和元年版

2019/07/21

Hajime MIZUNO@Ubuntu Japanese Team

みんなバックアップ取ってる？

- お手軽にUSB-HDDに…
- クラウドにまるなげ…
- 男のノーガード戦法

そもそもバックアップとは？

- 何を守るのか
- どこまで守るのか
- 失うことを許せるのか

何を守るのか？

- 全部に決まってるでしょ
→無限にカネがかかる！

どこまで守るのか？

- 全データの全履歴を死ぬまで
- →無限にカネがかかる!!

失うことを許せるのか？

- それを失うなんてとんでもない
→無限に力ネがかかる!!!

どこで折り合いをつけるのか？

- 守りたいデータの範囲
- 絶対に失いたくないもの
- クラウドに依存するリスク
- ディザスタリカバリ

これらをコストと天秤にかけ、最適な戦略を考える必要がある

筆者の実例

写真のバックアップを取りたい！

- 大容量
→ 1ショット70MB、1日1500～3000ショット
- 継続して増え続ける
→ だいたい毎週撮りに行くよね…
- 二度と取り戻せないデータ

データの特徴的に、どう考えてもカネがかかる案件である

大容量データをどう扱うか

- 無料で使えるクラウドサービスもあるけども
- 二度と取り戻せないデータを第三者に預けてよいのか？
- 結局HDDが最強

大容量ディスクとクラウドの折衷案

- 増え続ける大容量データ
- 第三者に丸投げできない
 - HDDを並べるしかない
 - 複製は何個作る？

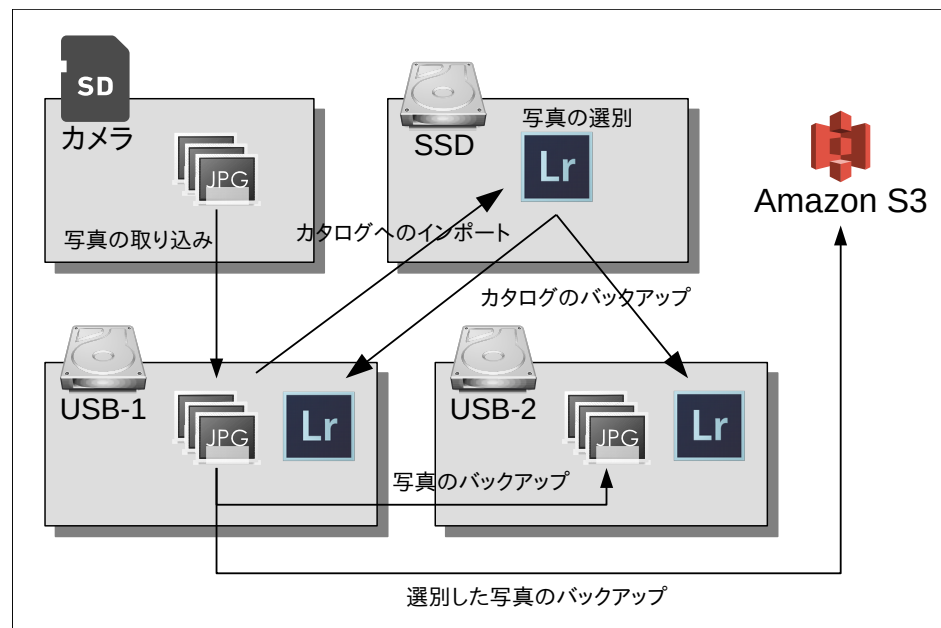
大容量ディスクとクラウドの折衷案

- 自宅に隕石が落ちたら？
→ やはりクラウドも不可欠
- ありもののサービスは不安
→ 自分でやるしかあるまい

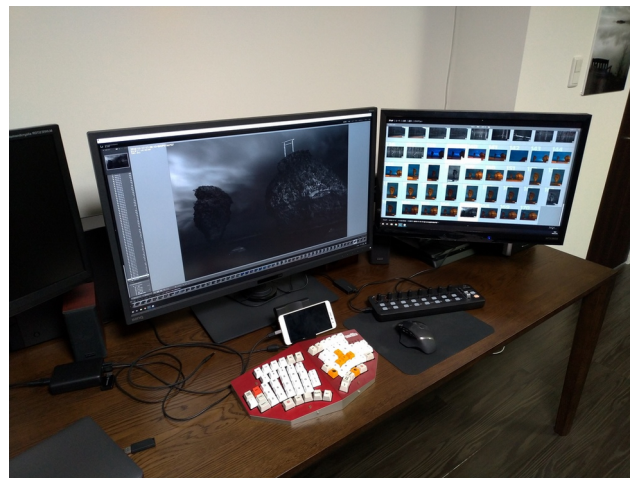
コストを考えた現実解は

- 全データは常に複数のHDDに複製を持つ
- 可能であれば遠隔地に保管
- 重要なデータのみを厳選してクラウドへ保管

筆者宅の構成



筆者宅の構成



実際のワークフロー

1. 写真をPCへ取り込み → `rsync`
2. 撮影日時ごとに整理 → `bash`
3. 全体をバックアップ → `rsync`
4. 写真を選別 → `Lightroom`
5. 選別した写真をバックアップ → `aws-cli`

Adobeから離れられない問題

- 写真を管理するカタログは必須
- OSSの現像アプリは品質面で純正 or Adobeに勝てない
→ Windowsが必要
- とはいえワークフローには強力なUnixツールが使いたい
→ WSLで解決

WSL is 何？

- LinuxバイナリをWindows上で実行する互換レイヤー
- MicrosoftストアからインストールできるLinux実行環境
- VMのような「隔離環境」ではない点が便利
- ストレージくそ遅い
- WSL2に超期待

撮影日時ごとに整理

- 一定のシャッター回数ごとにファイル名はループする
 - 重複を避けたい
 - 同じカメラが複数台あるともうカオス
- メーカー特有のフォルダ構造
 - 撮影年月日、時刻をわかるようにしたい

撮影日時ごとに整理

- exiftoolで写真の情報を取得して移動

```
for file in $(find $@ -type f)
do
    createdate=$(exiftool -n -b -DateTimeOriginal $file)

    date=$(awk '{print $1}' <<< $createdate | tr ':' '-')
    time=$(awk '{print $2}' <<< $createdate | tr ':' '-')

    if [ ! -d $date ]; then
        mkdir $date
    fi
    mv $file $date/$date-$time-$(basename $file)
done
```

カタログと写真をバックアップ

- rsyncでまるごとコピー

```
$ rsync -avz --delete /mnt/c/Users/mizuno/Documents/Catalog /mnt/y  
$ rsync -avz --delete /mnt/x/Photo /mnt/y
```

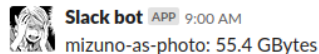
選別した写真をバックアップ

- LightroomのカatalogはSQLite3のデータベース
- SQLを使ってレーティングされた写真を抽出する

```
select absolutePath || pathFromRoot || originalFilename
from Adobe_images inner join (AgLibraryFile inner join
(AgLibraryFolder inner join AgLibraryRootFolder on
AgLibraryRootFolder.id_local = AgLibraryFolder.rootFolder)
on AgLibraryFile.folder = AgLibraryFolder.id_local)
on Adobe_images.rootFile = AgLibraryFile.id_local
where rating is not null;
```

選別した写真をバックアップ

- データベースから抽出した写真をワークエリアに一旦コピー
- ワークエリアからs3 sync
- S3の使用量は毎日Slackへ通知 →



```
for file in $(sqlite3 "${LRCAT}" "${SQL}"); do
  if [ -f "${file}" ]; then
    cp -v -u "${file}" "${DEST}${DIR}/"
  fi
done

${AWSCLI} --profile "${KEY}" s3 sync "${DEST}" s3://"${BUCKET}"/
```

```
$ aws cloudwatch get-metric-statistics --namespace AWS/S3 \
  --metric-name BucketSizeBytes \
  --dimensions Name=BucketName,Value=バケット名 \
    Name=StorageType,Value=StandardStorage \
  --statistics Maximum \
  --start-time $(date '+%Y-%m-%dT00:00:00' -d '1 days ago') \
  --end-time $(date '+%Y-%m-%dT00:00:00') \
  --period 86400 --unit Bytes
```

まとめ

- 取り戻せないデータという意味を真面目に考えよう
- 家がなくなる可能性は思っているより高いかもしれない
- 既存サービスを使うならリスクを判断しよう
- まだまだHDD最強
- クラウドは適材適所で