

百万件くらいの データの扱い方

Masafumi Yokoyama

Rabbitについて

- ✓ プレゼンテーションツール
- ✓ 実装: Ruby/GTK+
- ✓ 動作: UNIX/Win/Mac
- ✓ 文章とデザインの分離
 - ✓ バージョン管理しやすい

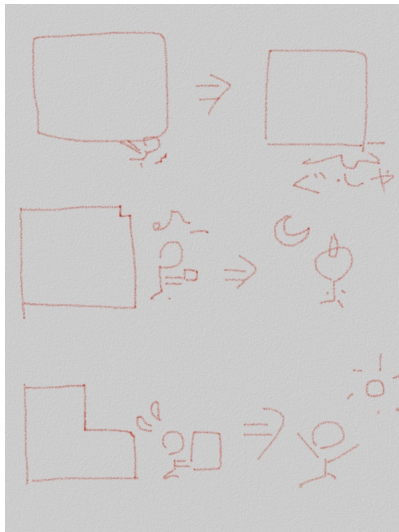


本日のテーマ



- ✓ 件数が多いデータの扱い方
 - ✓ データ移行
 - ✓ ログフォーマット変換

キーワードは『荷物運び』



一度に運ぶ量

✓ 大量

✓ 持てない

✓ 少量

✓ 夜になっても終わらない

✓ 適量

✓ 明るいうちに終わる



まとめ



- ✓ 件数が多いデータは、適量に分けて処理する
- ✓ 適量を見極めたら勝ち



第一部 完

目次



- ✓ 第一部 考え方
- ✓ 第二部 具体例
 - ✓ UNIXシェル
 - ✓ DBMSのトランザクション処理
- ✓ 第三部 スレッド

第二部



目次



- ✓ 第一部 考え方
- ✓ 第二部 具体例
 - ✓ UNIXシェル
 - ✓ DBMSのトランザクション処理
- ✓ 第三部 スレッド

UNIXシェル



- ✓ UNIXのコマンドプロンプト
- ✓ コマンドプロンプトよりも高機能

問題



- ✓ 大量ファイルの削除
 - ✓ ディレクトリ内の.gomiファイルを全て削除したい
 - ✓ 数千～数万ファイル

事前準備



- ✓ 名前が長いファイルを大量生成
 - ✓ ファイル名: 210バイト
 - ✓ ファイル数: 10,000

```
% cd /tmp ; mkdir gomi ; cd gomi  
% ruby -e '0.upto((10 ** 4) - 1) {|i| `touch #{"gomi" * 50}_#{"%04d" % i}.gomi` }'  
% (cd .. ; tar xzf gomi.tar.gz gomi)
```

A. 普通に削除



```
% ls | wc -l
```


```
10000
```

```
% rm *.gomi
```

```
zsh: 引数リストが長すぎます: rm
```

```
% ls | wc -l
```

```
10000
```



消せない！