

データアクセス速度 を測ってみた

わしだ

答えられますか？

- ハッシュやDBのアクセスは何要素でどれだけ時間がかかるのか？
- メモリアクセスは1秒で何回くらい可能か？
- ディスクアクセスは？

こんな事、

- 「XXは速い/遅い」という話に流されていませんか？

基本操作の速度すら
あまり知られていない

桁くらしいは
知っておきたい

調べてみました

1. メモリアクセス速度

2. メモリ/ディスクのコピー速度

3. 応用操作

1. メモリアクセス

計測1

- 指定サイズのメモリを百万回アクセス
- アクセスの順番はランダム
- 領域のサイズは512B～1GBの範囲

スペック

- MacBook Pro (Core 2 Duo 15", 2007)
- 2GB + 1GB RAM

結果

ランダムアクセス

- ある値を境に段階的に遅くなる

⇒ メモリキャッシュ

CPU キャッシュ(1)

高速なRAMをCPUに内蔵、キャッシュ

- L1キャッシュ

サイズ ≒ 数十kB、速度 ≒ 数ナノ秒、四則演算レベル

- L2キャッシュ

サイズ ≒ 数MB、速度 ≒ L1の1/10、メインメモリの10倍

- L3キャッシュ (高スペックマシンに搭載)

CPU キャッシュ (2)

- 命令キャッシュ

データとは別に命令用にL1キャッシュを用意

- Translation Lookaside Buffer、TLB

仮想アドレスのマッピング

数段階ある

計測2

- 指定サイズのメモリを百万回アクセス
- アクセス方法を変えてみる
 - シーケンシャルアクセス
 - ランダム読み込み後、書き込み

結果

結論

- キャッシュに乗るかが大きな分かれ道
- シーケンシャルアクセスは意外と速い
- データの局所性は常に重要

2. コピー

行った計測

- デバイス

1. メインメモリ

2. ディスク

- 計測方法

- A. 小さいデータ、定数回コピー

- B. 大きいデータ、合計サイズ固定でコピー

計測1A

- メモリから1000回コピー
- コピー元の領域は1kB～1GB
- 1回のコピーサイズを8B～1MB

結果

小さいサイズのコピー

- 1kB程度以下ならサイズによる速度差はない
- コピー対象の容量もあまり関係ない

? 通常のアクセスより速い事がある

計測1B

- メモリから合計1GBコピー
- コピー元の領域は1kB～1GB
- 1回のコピーサイズは1kB～64MB

結果

大きいサイズのコピー

- 回数ではなく合計のコピー容量が重要

? 一回のコピー量が増えると段階的に遅くなる

? バイトをずらすと速い

計測2A

- ディスクから16回コピー
- コピー元は16GBのファイル
- 1回のコピーサイズを256B~1MB
- readとmmapで比較

補足: ディスクアクセス速度

- 平均シーク時間 + 平均回転待ち時間 + 転送時間
- 検証マシンでは
 - シーク: 12ms
 - 回転: 5400rpm
- 待ち時間: $12 + 1000 / (5400 / 60) / 2 \doteq 17.6(\text{ms})$

結果

小さいサイズのコピー

- 数十kB程度以下ならサイズによる速度差はない
- 遅い
 - 1回で数十ミリ秒、メインメモリの10万倍程度

計測 2B

- ディスクから合計の64MBコピー
- コピー元は16GBのファイル
- 1回のコピーサイズは8MB～64MB

補足: readとmmap

UNIX系OSでの代表的なファイルアクセス方法

1. read ... 指定範囲のデータをメモリにコピー

2. mmap ... ファイルをメモリのように扱う

- 大抵メモリにキャッシュされる
 - キャッシュの仕方が異なる

結果

大きいサイズのコピー

- 回数ではなく合計のコピー容量が重要
- 遅い
 - 64MBで2.8秒

結論

- 小さいデータのコピー時間は一定
- 大きいデータのコピーは容量に比例
- キャッシュの効果がある
- ディスクアクセスは非常に遅い

3. 応用

代表的な例を検証

- ハッシュ
- grep
- ソート
- join

その1: HashMap にアクセス