

Windows、Linuxにおけるソート処理性能 ベンチマークレポート

～ SORTKITのご紹介～

日本電気株式会社(NEC)

2014 年05 月

昨今ビッグデータに注目され、ストレージシステムの大容量化しているなか、多くのシステムでは扱うデータ量が以前に比べ大幅に増えています。このような状況の中、バッチ処理などのデータ処理において重要な位置を占めるソート処理への負荷は、データの増大にともない、より一層大きくなっています。ソート処理の遅れが、データ処理全体の遅延の原因になり、ひいては企業戦略の意思決定の遅れをまねくかもしれません。多くのシステムにおいてソート処理をいかに速くするかが、そのシステムを構築した企業のビジネスを成功させる鍵の一つといえるでしょう。

本稿では、この問題に対し解決策をご提供するSORTKITをご紹介します。

1. はじめに

Windows®、Linux®に標準で付属しているsortコマンドでソート処理を行うことが可能です。OS標準sortコマンドとSORTKIT®のソート処理の性能ベンチマークを行います。

2. SORTKITの概要

SORTKITは、マルチCPU対応や効率のよいソートアルゴリズム、その他高速化手法により、大量データのソート・マージ処理を行うユーティリティです。コマンドやC言語インターフェースなど複数のインターフェースを持ち、バッチ処理やC言語で作成されたアプリケーションから利用することができます（本稿では、バッチ処理からの利用を想定して、コマンドインターフェースを利用）。

その他に、OS標準sortコマンドでは扱えない固定長レコード形式ファイルや索引順編成ファイルなどのファイル編成や、パック10進数、シフトJISなど様々なソートキーなどを扱えます。また、レコード選択削除機能や、レコードのフィールド構成の変更、集計処理、複数ファイルへの出力などをソート・マージ処理と同時にを行うことができます。

3. 目的

SORTKIT付属のソートコマンドと、OS標準sortコマンドとの実行時間の比較を行います。データサイズを変化させた場合の各コマンドの実行時間を確認します。

4. システム構成

OSは、次のバージョンを使用します。

- Windows

サーバー向けOS : Windows Server® 2012
クライアント向けOS : Windows® 8

- Linux

Red Hat® Enterprise Linux® 6.3

測定するマシンのスペックを図1に示します。

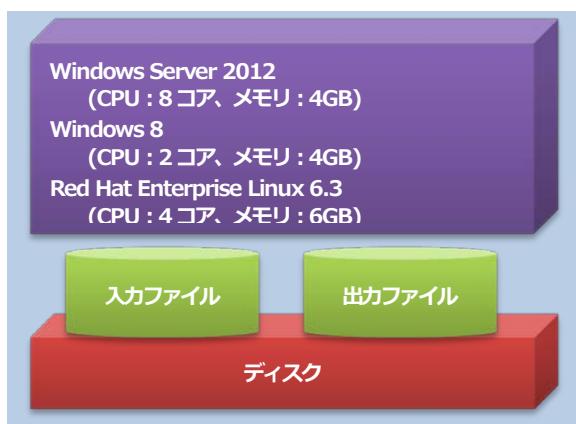


図1 マシンスペック

5. 測定方法

測定に使用するSORTKITは、Windows Server 2012ではSORTKIT/Enterprise Ver6.2、Windows 8では

SORTKIT for PC Ver8.1、Red Hat Enterprise LinuxではSORTKIT(x64版) Ver2.1とします。

ソートコマンドを起動してから、ソートコマンドが終了するまでの時間を実行時間として測定します。測定条件を揃えるため、WindowsではOSの再起動、Linuxではメモリのクリアを各測定の前に行います。

ファイル形式は、OS標準sortコマンドで処理できる可変長レコード形式とし、データサイズは、次の通りとします。

- Windows Server 2012
- Red Hat Enterprise Linux 6.3
250MB～7GB
- Windows 8
200MB～600MB

その他のデータ条件は、下記の通りです。

平均レコードサイズ : 765バイト
キーサイズ : 22バイト

6. 結果

6.1. Windowsにおける性能比較結果

Windows Server 2012の結果を図2、3に、Windows 8の結果を図4に示します(縦軸が実行時間、横軸がデータサイズです)。

Windows Server 2012は、ソート処理方法に応じて、2つのグラフに分けています。図2は、すべてのデータをメモリ上でソート処理を行える条件(メモリソート)の結果です。図3は、データサイズが大きく、1度にメモリ上でソート処理を行えないため、中間ファイルを使用する条件(ファイルソート)の結果です。

図4のWindows 8の場合、すべてのデータをメモリ内で処理できるソート処理であり、図2と同様の傾向になっています。

各測定結果とも、データサイズが小さい場合は、OS標準sortとSORTKITで大きな差は出ていませんが、データサイズが大きくなるほど、OS標準sortとSORTKITの差が大きくなっています。

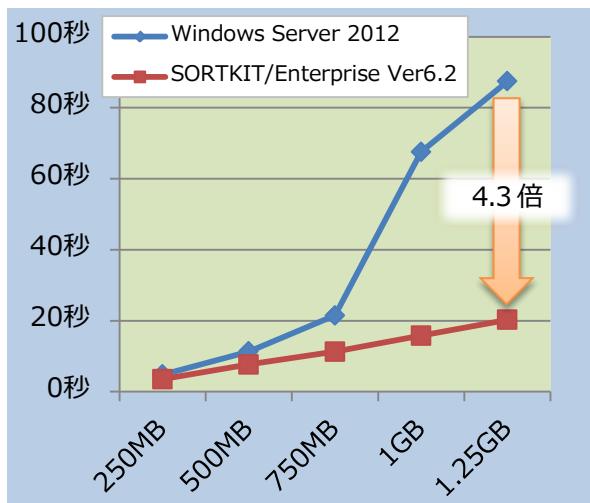


図2 Windows Server 2012(メモリソート)

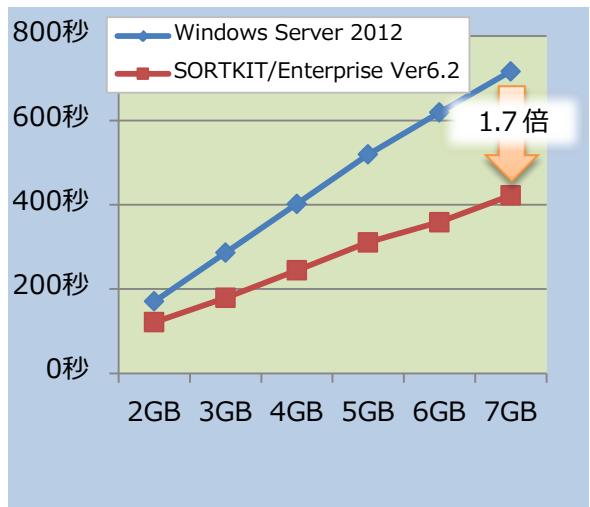


図3 Windows Server 2012(ファイルソート)

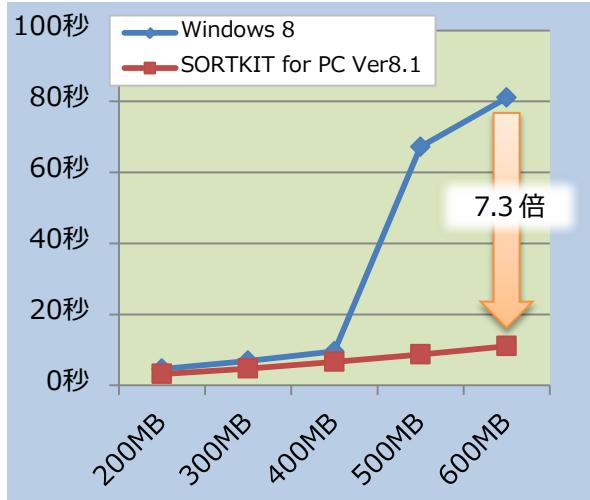


図4 Windows 8 上での比較

6.2. Linuxにおける性能比較結果

Red Hat Enterprise Linux 6.3の結果を図5、6に示します。Windows Server 2012の場合と同様に、ソート処理方法に応じて、2つのグラフに分けています。測定したすべてのデータサイズにおいて、SORTKITはOS標準sortに比べ、メモリソートの条件では約4倍、ファイルソートの条件で約2.5倍の性能があります。Windowsと同様、Linuxにおいても、大きいデータサイズほど実行時間に差があることが確認できます。

7. おわりに

Windows、LinuxともにOS標準sortをSORTKITに置き換えることで、実行時間の短縮が見込め、データサイズが大きいほど、大きな効果が期待できます。

【お問い合わせ先】
N E C プラットフォームビジネス本部

〒108-8424
東京都港区芝五丁目33番8号（第一田町ビル）
TEL : 03(3798)7177
受付時間：9:00～12:00 13:00～17:00
月曜日～金曜日（祝日・NEC所定の休日を除く）
<http://www.nec.co.jp/sortkit/>

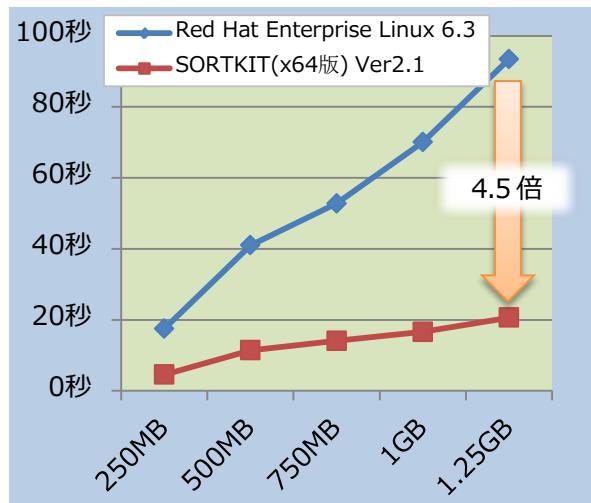


図 5 Red Hat Enterprise Linux 6.3(メモリソート)

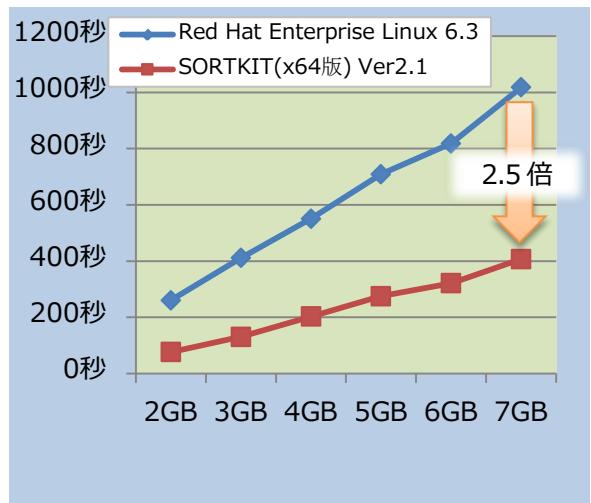


図 6 Red Hat Enterprise Linux 6.3(ファイルソート)

免責事項

※ 当社は、本資料で提供される内容に関し、その正確性、有用性、確実性その他いかなる保証もするもではありません。本ホワイトペーパーで提供される内容のご利用により万一何らかの損害が発生したとしても、当社は一切責任を負いません。

- ・ Windows、および、Windows Server は、米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です。
- ・ Linux は、Linus Torvalds 氏の米国およびその他の国における登録商標または商標です。
- ・ Red Hat は、Red Hat, Inc.の米国およびその他の国における登録商標または商標です。
- ・ SORTKIT は、日本電気株式会社の登録商標です。
- ・ その他、本書に登場する会社名、製品名は一般に各社の登録商標または商標です。