Japan Geoscience Union Meeting 2014
(28 April - 02 May 2014 at Pacifico YOKOHAMA, Kanagawa, Japan)

MGI37-03 会場:413 時間:4 月 29 日 09:30-09:45

分散ファイルシステムによる並列データ I/O 測定
An Examination of Data I/O Speed on a Parallel Data Storage System

村田 健史 [1*] ; 渡邉 英伸 [1] ; 鵜川 健太郎 [2] ; 村永 和哉 [2] ; 鈴木 豊 [2] ; 建部 修見 [3] ; 田中 昌宏 [3] ; 木村 映善 [4]
MURATA, Ken T.[1*] ; WATANABE, Hidenobu[1] ; UKAWA, Kentaro[2] ; MURANAGA, Kazuya[2] ; YUTAKA, Suzuki[2] ; TATEBE, Osamu[3] ; TANAKA, Masahiro[3] ; KIMURA, Eizen[4]

[1] 情報通信研究機構, [2] 株式会社 セック, [3] 筑波大学, [4] 愛媛大学
[1]National Institute of Information and Communications Technology, [2]Systems Engineering Consultants Co., LTD., [3]University of Tsukuba, [4]Ehime University

現在、多くの科学研究分野ではデータのほとんどがデジタル化され、その量および種類は大規模化の一途をたどっている。これからますます大規模化・複雑化するデータ指向型科学時代を踏まえて、ビッグデータ処理がより容易に、また一元的行うことができるクラウドシステムが求められている。

NICT サイエンスクラウドは、地球惑星科学を含む様々な科学研究データおよびソーシャルデータのためのクラウドシステムである。NICT サイエンスクラウドでは（1）データ伝送・データ取集機能、（2）データ保存・データ管理機能、（3）データ処理・データ可視化機能の 3 つの柱（機能）から構成されている。それぞれの機能についての基盤技術を開発するだけではなく、複数の基盤技術を組み合わせることでシステム化を行うことができる。システムを実際に科学研究に応用・適用することで、様々な分野でのビッグデータ科学・データインテンシブ科学が可能となる。

本研究では、NICT サイエンスクラウド上で科学研究のビッグデータ処理を行うための基盤技術について議論する。データサイズが大きい場合にクラウドデータ処理で解決すべき問題点の一つはデータ I/O である。例えば、100MB/sec で 100TB のデータを読み出すとすると、1,000,000 秒（約 11.5 日）かかる。すなわち、大規模科学データを処理するためには、高速 I/O 技術が不可欠である。本発表では、並列ファイルシステム（GPFS）と分散ファイルシステム（Gfarm）の 2 つのシステムでのデータ読み出し速度の比較を行い、それらのスケーラビリティーを比較する。