

## NICTサイエンスクラウドによるビッグデータ処理技術 Techniques of Big-Data Processing on the NICT Science Cloud

村田 健史<sup>1\*</sup>; 渡邊 英伸<sup>1</sup>; 鶴川 健太郎<sup>2</sup>; 村永 和哉<sup>2</sup>; 鈴木 豊<sup>2</sup>; 建部 修見<sup>3</sup>; 田中 昌宏<sup>3</sup>; 木村 映善<sup>4</sup>  
MURATA, Ken T.<sup>1\*</sup>; WATANABE, Hidenobu<sup>1</sup>; UKAWA, Kentaro<sup>2</sup>; MURANAGA, Kazuya<sup>2</sup>; YUTAKA, Suzuki<sup>2</sup>; TATEBE,  
Osamu<sup>3</sup>; TANAKA, Masahiro<sup>3</sup>; KIMURA, Eizen<sup>4</sup>

<sup>1</sup> 情報通信研究機構, <sup>2</sup> 株式会社 セック, <sup>3</sup> 筑波大学, <sup>4</sup> 愛媛大学

<sup>1</sup>NICT, <sup>2</sup>Systems Engineering Consultants Co., LTD., <sup>3</sup>University of Tsukuba, <sup>4</sup>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、また一元的に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・データ収集機能、(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。本発表では、分散ファイルシステム(Gfarm)でのデータ読み出し速度の評価を行い、将来の大規模データ処理のためのスケーラビリティについても検討する。

さらに、この高速I/Oを元にしたビッグデータ処理技術について議論する。NICTサイエンスクラウドでは、衛星観測データ処理や3次元シミュレーションデータ処理、3次元気象レーダデータ処理(リアルタイム)等で60倍から100倍を超えるビッグデータ処理の高速化を達成している。この技法は汎用性が高いため様々な研究分野において応用が可能である。