

Gfarm/Pwrakeによるリアルタイムデータ処理技法とその応用事例紹介

Real-time Data Processing via Gfarm/Pwrake and Applications on the NICT Science Cloud

*村永 和哉¹、村田 健史²、山本 和憲²、長屋 嘉明²、鷗川 健太郎¹、村山 純一¹、鈴木 豊¹、建部 修見³、田中 昌宏³、木村 映善⁴

*Kazuya Muranaga¹, Ken T. Murata², Kazunori Yamamoto², Yoshiaki Nagaya², Kentaro Ukawa¹, Junichi Murayama¹, Yutaka Suzuki¹, Osamu Tatebe³, Masahiro Tanaka³, Eizen Kimura⁴

1.株式会社セック、2.情報通信研究機構、3.筑波大学、4.愛媛大学医学部

1.Systems Engineering Consultants Co., LTD., 2.National Institute of Information and Communications Technology, 3.University of Tsukuba, 4.Department of Medical Informatics Ehime Univ.

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（3）大規模データの並列処理技術について議論する。本発表では分散ファイルシステム（Gfarm）とアフィニティスケジューラであるPwrakeを組み合わせた並列分散データ処理技術について紹介する。さらに、Gfarm/Pwrakeをリアルタイムデータ処理に応用した事例として、3次元気象レーダデータ処理、気象衛星データ処理などの事例を紹介する。