

平成27年度北海道大学情報基盤センター共同研究成果報告書

1. 研究領域番号 A2

2. 研究課題名 クラウド環境における高性能統計計算

3. 研究期間 平成 27 年 4 月 1 日 ~ 平成 28 年 3 月 31 日

4. 研究代表者

氏 名	所属機関・部局名	職 名	備 考
中野純司	統計数理研究所・モデリング研究系	教授	

5. 研究分担者

氏 名	所属機関・部局名	職 名	備 考
清水信夫	統計数理研究所・データ科学研究系	助教	
本多啓介	情報・システム研究機構・総合企画本部 URAステーション	リサーチ・アドミニ ストレーター	
棟朝雅晴	北海道大学・情報基盤センター	教授	
水田正弘	北海道大学・情報基盤センター	教授	

6. 共同研究の成果

下欄には、当該研究期間内に実施した共同研究の成果について、その具体的内容、意義、重要性等を、共同研究申請書に記載した「研究目的」と「研究計画・方法」に照らし、800字～1,000字で、できるだけ分かりやすく記載願います。文章の他に、研究成果を端的に表す図表を貼り付けても構いません。なお、研究成果の論文・学会発表等を行った実績（発表等の予定を含む。）があれば、あわせて記載して下さい。

まず、北海道大学情報基盤センターと統計数理研究所の情報交換の場として、2015年5月に統計数理研究所の中野と本多が北大情報基盤センターを訪問し、互いのシステムの説明を行うとともに、それらのシステムの改良点や将来のインタークラウドの方向について議論した。その結果、大学と共同利用機関におけるスパコンやクラウドの目的や役割における違いを明確に理解することができた。そして、北大の課金システムの思想や内容が大学共同利用機関である統計数理研究所の今後のシステム構築において参考になることがわかった。

さらに、多量データの統計解析において必要とされている重要ソフトウェアとして統計解析環境 R と多量分散処理インフラストラクチャ Apache Hadoop が注目されていることを鑑み、R から Hadoop を利用できる RHIPE (<https://github.com/tesseractata/RHIPE>) を導入し、その実用性を検討した。その結果、Hadoop は成熟したソフトウェアであるが、その導入は容易とは言えず、自分でシステムを構築することはかなり難しいことが確認できた。この点、北大のようにすぐに動くクラウド環境を提供することが非常に重要であることを強く認識した。また、RHIPE はまだ実験的なソフトウェアであり、OS や Hadoop のバージョンとの相性が非常に微妙であることも確認できた。本研究の範囲外で行われたことではあるが、RHIPE がその一部となっているプロジェクト Tessera (<http://tessera.io>) の主要メンバーと議論する機会があったことは本研究にとっても有益であった。

(研究成果のつづき)

目的の一つとして、多量のシミュレーションが現在の統計解析では多く利用され、そのために伝統的なスパコンを利用した高性能計算が重要であるという認識のもとで、それを比較的容易に実現するためのクラウド上に、統計数理研究所で開発した R 上の高性能並列計算用パッケージ Rhpc のチューニングを行うということをあげていた。これに関しては、残念ながら具体的な成果を得ることができなかった。少なくとも実験したクラウドにおいてはネットワークのスピードがボトルネックになっているようで、それを解決するには至らなかった。

その他、2015 年 12 月に他の共同研究と合同で研究集会を行った。そこでもクラウド上の統計計算に関して議論することができた。また、統計数理研究所の清水が 2016 年 3 月に北大を訪問し、北大メンバーと多量データの統計計算について議論した。