学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　Phyo Thandar Thant

学　位　論　文　題　名

Multi-Objective Resource Optimization of Big Data Applications in Cloud Computing Infrastructure
（クラウドコンピューティング基盤におけるビッグデータアプリケーションの多目的資源割当最適化）

Experimental applications in the scientific area have become very complex with ever increasing amounts of Big Data that highlight the criticality of high-performance computing environments such as clusters, grids, and the cloud. Grid computing systems were widely used in scientific applications in the past as they provide parallelism through high-performance computational infrastructure. However, the adoption of cloud computing to process these scientific data has increased recently because it enables researchers to execute workflows at low costs without setting up their own infrastructure. Moreover, the elastic nature of cloud facilitates the changing of resource quantities and characteristics to vary at runtime, facilitating dynamic scaling of infrastructure resources according to application requirements and budget. Further, nowadays, large-scale genome data presents a significant challenge for data processing and analytics in Big Data environment. In this scenario, Galaxy, a widely used scientific workflow management system, provides an open, web-based platform for data-intensive computational analyses and data integration. However, the resources needed by scientific workflows vary significantly during execution on Galaxy and may become inefficient in terms of resource usage and cost. Therefore, researchers are focused on deploying Galaxy in the cloud for better resource utilization and data analytics. Nevertheless, several research issues associated with the deploying of scientific workflow applications in the cloud still exists. Among them, resource selection to make the right choice of instances for a certain application of interest is a challenging problem for researchers. In addition, providing services with optimal performance at the lowest financial resource deployment cost based on users' resource selection is quite challenging for cloud service providers.

Scientific workflows are a series of computation-dependent tasks associated with scientific experiments. They have emerged as an alternative to script programming for performing in-silico experiments and are concerned with the processing of high-volume scientific Big Data. Several workflow optimization studies have focused on optimizing makespan through scheduling mechanisms, whereas others have conducted multi-objective optimization during workflow execution. However, conventional workflow optimization approaches only consider makespan and cost optimization and do not cover crucial factors related with virtual machine (VM) instances, such as VM failure probability and availability probability, which can significantly affect the service level agreements (SLAs) of the cloud during workflow execution.

This thesis begins with preliminary studies of Big Data applications in Hadoop infrastructure by using VM instances from cloud computing infrastructure. Subsequently, it focuses on investigation of the challenges of resource optimization of scientific workflow executions in the cloud and develops a new constrained multi-objective optimization model for general scientific workflow applications such as bioinformatics, environmental issues, and astronomy using multi-objective evolutionary algorithms that benefit both stakeholders (users and cloud service providers). The following contributions are made during resource optimization of the proposed system: (i) the modeling of the resource optimization problem as a constrained multi-objective optimization problem, (ii) randomized search-based resource optimization for rightsized instance resource selection, (iii) level-wise workflow execution optimization for task dependencies and resource elasticity, and (iv) level-wise balanced task clustering and assignment for better load balancing and resource sharing.

Thus, the system provides several advantages, such as elimination of the resource selection problem confronting novice cloud users, optimal resource assignment, optimal budget management, and improved service satisfaction for the cloud users from the service providers. This system can support rapid adoption of scientific applications to more efficient cloud applications and can also handle the output validity problem of large critical scientific applications in certain situations.