

Large-Scale Data Management and Analysis for Astronomical Research

The improvement of information technology enables precise scientific observation that demands larger storage and faster data processing techniques than ever before. From the perspective of astronomical research, one of the most important challenges is to extract useful astronomical information efficiently from a huge collection of observed data. Even though the existing distributed computing techniques, such as grid computing and cloud computing, have provided the scientists a better way to access powerful computing resources, the development of big-data management and analysis software is still lagging far behind. The awkward predicament obstructs the connected computing resources from being utilized efficiently. Therefore, it is beneficial to provide an integrated, efficient information management and analysis system for astronomical research. This research, conducted by the PAN-STARRS research team at Taiwan, focuses on the issues of integrating commercial data warehouse and large-scale grid computing techniques, and develops a system for efficient data management and fast analysis in astronomy-related fields. Our system can be viewed as a data grid system that supports analysis of large data collections such as Large Synoptic Survey Telescope (LSST). The system consists of two analytical sub-systems and one data presentation and management sub-system. The first one is called the PARAllel Hierarchical Agglomerative Clustering System (PARHACS), which uses a distributed message-passing algorithm to efficiently calculate a hierarchical cluster, given a set of astronomical data. The second sub-system is called the SIMilarity Classification System (SIMCS), which uses a decentralized Multiple Classifier System (MCS) framework to support a complex classification procedure using multiple classifiers. The last sub-system is called the ASTROnomical Information Management System (ASTROIMS), which utilizes a multidimensional data-warehouse design to construct a more concise, integrated, and scalable platform for fast data retrieval and management. It is able to perform data maintenance procedures automatically and to reduce maintenance and operation costs easily. In addition, the sub-system provides a user-friendly interface to facilitate a variety of data analytical tasks on line.

Primary authors : Mr. TANG, Cheng-hsien (Department of Computer Science and Information Engineering, National Central University, Taiwan)

Co-authors : WANG, Min-feng (Department of Computer Science and Information Engineering, National Central University, Taiwan) ; WANG, Wei-jen (Department of Computer Science and Information Engineering, National Central University, Taiwan) ; TSAI, Meng-feng (Department of Computer Science and Information Engineering, National Central University, Taiwan) ; URATA, Yuji (Institute of Astronomy, National Central University, Taiwan) ; NGEOW, Chow-choong (Institute of Astronomy, National Central University, Taiwan) ; LEE, Induk (Institute of Astronomy, National Central University, Taiwan) ; HUANG, Kuiyun (Academia Sinica Institute of Astronomy and Astrophysics, Taiwan)

