

High Performance Distributed Data Processing Pipeline for Chinese Spectral Radioheliograph

Feng WANG^(1,2), Hui DENG⁽¹⁾, and Wei WANG⁽³⁾

(1) Kunming University of Science and Technology, Kunming, 650500

(2) Yunnan Observatories, Chinese Academy of Sciences, Kunming, China

(3) Chinese National Observatory, Chinese Academy of Sciences, Beijing, China, 100011

The Chinese Spectral RadioHeliograph (CSRH) is a solar-dedicated synthetic aperture radio interferometer built in Inner Mongolia, China. CSRH is capable of producing high quality radio images at frequency range from 400 MHz to 15 GHz with high temporal, spatial, and spectral resolution. To implement high cadence imaging at wide-band and obtain more than 2 order higher multiple frequencies, the implementation of the data processing system for CSRH is a great challenge.

Referring to the stream computing technology, we first develop a novel distributed computing infrastructure named OpenCluster, which is a wholly own-designed software for quickly designing scientific data processing pipeline. Fig.1 shows a conceptual diagram of the OpenCluster. We then develop a high performance distributed data processing pipeline (DDPP) built on the OpenCluster for processing CSRH observational data including data storage, archiving, preprocessing, image reconstruction, deconvolution, and real-time monitoring. We comprehensively elaborate the system architecture of the pipeline and the implementation of each subsystem (see Fig.2).

The real application of the DDPP proves that the DDPP has several distinguished advantages. 1. The DDPP is a loose-coupled system and easy to expand. All processing components are encapsulated into the standalone services and deployed upon the network. 2. The DDPP is a robust system that can be reliable operation without any maintenance. 3) The DDPP is a hybrid computing system which has integrated distributed computing technology and GPU technology for high performance data processing. 4) The OpenCluster provides some level of support for workflow technology.

The study presents an valuable reference for other radio telescopes especially aperture synthesis telescopes, and also gives an valuable contribution to the current and/or future data intensive astronomical observations.

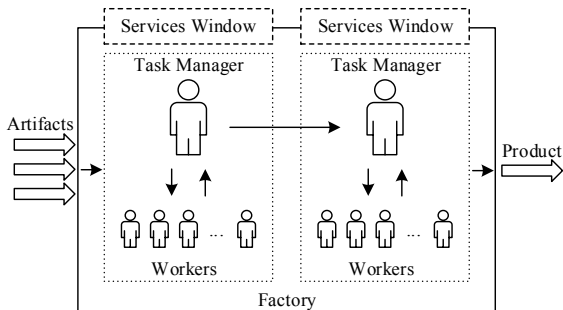


Fig.1 The conceptual diagram of the OpenCluster.

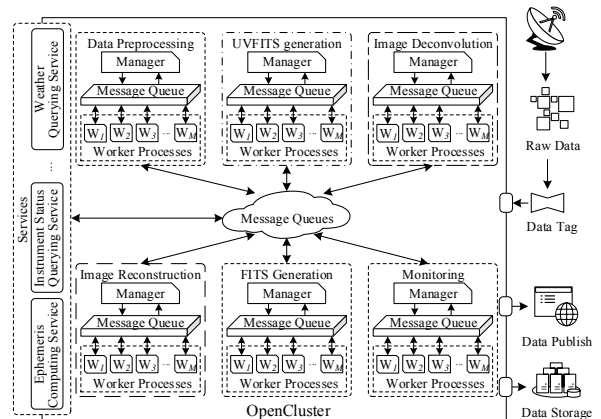


Fig.1 The system infrastructure diagram.