

A lightweight task submission and management infrastructure

Monday, October 10, 2016 2:15 PM (15 minutes)

In the near future, many new experiments (JUNO, LHAASO, CEPC, etc) with challenging data volume are coming into operations or are planned in IHEP, China. The Jiangmen Underground Neutrino Observatory (JUNO) is a multipurpose neutrino experiment to be operational in 2019. The Large High Altitude Air Shower Observatory (LHAASO) is oriented to the study and observation of cosmic rays, which is going to collect data in 2019. The Circular Electron Positron Collider (CEPC) is planned to be a Higgs factory and upgraded to a proton-proton collider in second phase. The DIRAC-based distributed computing system has been enabled to support multi experiments. Development of task submission and management system is the first step for new experiments to have a try or use distributed computing resources in their early stages. In the paper we will present the design and development of a common framework to ease the process of building experiment-specific task submission and management system. Advanced object-oriented programming technology has been used to make infrastructure easy to extend for new experiments. The framework covers the functions including user interface, task creation and submission, run-time workflow control, task monitor and management, dataset management. YAML description language has been used to define tasks, which can be easily interpreted to get configurations from users. The run-time workflow control adopts the concept of DIRAC workflow and allows applications easily to define their several steps in one job and report status separately. Common modules including splitter to split tasks, backend to heterogeneous resources, job factory to generate the related parameters and files for submission have been provided. The monitoring service with web portal has been provided to monitor status for tasks and the related jobs. The dataset management module has been designed to communicate with Dirac File Catalog to implement query and register of dataset. At last the paper will show two experiments JUNO and CEPC how to use this infrastructure to build up their own task submission and management system and complete their first scale try on distributed computing resources.

Secondary Keyword (Optional)

Data processing workflows and frameworks/pipelines

Primary Keyword (Mandatory)

Distributed workload management

Tertiary Keyword (Optional)

Primary authors: Mr SUO, Bing (Shandong University); YAN, Tian (Institution of High Energy Physics, Chinese Academy of Science); Mr ZHAO, Xianghu (Nanjing University); ZHANG, Xiaomei (Chinese Academy of Sciences (CN))

Co-authors: ZHANG, Yao (Institute of High Energy Physics and); DENG, ziyang (Institute of High Energy Physics, Beijing)

Presenters: Mr SUO, Bing (Shandong University); ZHANG, Xiaomei (Chinese Academy of Sciences (CN))

Session Classification: Track 3: Distributed Computing

Track Classification: Track 3: Distributed Computing