The JSUB task submission and management software

Yifan Yang, Xiaomei Zhang, Xianghu Zhao

Institute of High Energy Physics (IHEP), CAS

Motivation

Future high energy physics experiments like JUNO [1] and CEPC [2] would give rise to needs of processing large amount of data. To encourage experiment users to make good use of the distributed computing resources of the collaborations, it's essential to provide them with handy tools to save them from troubles of mastering various relevant technical details.
For this need, JSUB is developed to:

- Ease the procedure of using DIRAC [3], and potentially other computing platforms.
- Automatically manage massive jobs.
- Be highly extensible to different HEP experiments.

Introduction

A lightweighted front-end task submission and management tool which covers functionalities including the submission, job management, and relevant data management of tasks in distributed computing scenario.

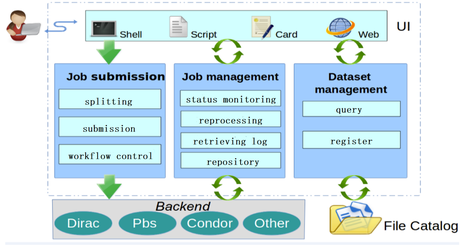

Figure 1:Functionalities of JSUB

```
Commands:
  create      Create a task from a task description file.
  getlog      Retrieve log files of selected subjobs.
  jobvar      View the values of jobvar lists
  ls          List all tasks.
  package     Show active packages.
  remove      Delete a task.
  rename      Rename a task.
  reschedule  Reschedule selected subjobs.
  resubmit    Equivalent to 'jsub submit -r' command
  run         Create from a task profile, and submit.
  show        Show detailed description of a task.
  status      Show the backend status of a task.
  submit      Submit a task to backend.
  version     Show the version of the software.
```

Figure 2:Supported commands for JSUB

Task management with JSUB

Life cycle of a task includes the steps of creation, submission, running, monitoring, rescheduling, and retrieval of output.

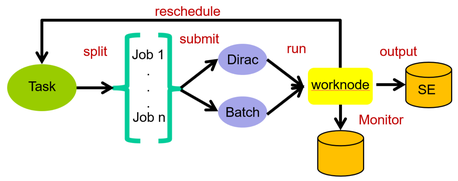

Figure 3:Life cycle of a task in distributed computing

To create a task, JSUB parses YAML task description file (TDF) to get the settings, split a large task into jobs that can each run on a single machine, and generate necessary files for submission.
Next, JSUB can communicate with backend softwares such as DIRAC and HTCondor to send these jobs to working nodes in computing sites.
On working nodes, the main executable of these jobs is a driver program that runs applications according to the user-defined workflow, keeps track of logging information, and may send monitoring messages to backend platform.
JSUB backend modules communicate with backend platforms to monitor running status of jobs, and to reschedule failed jobs.
Data downloading and uploading are automatically added as the first and last steps of the workflow.

Parametric subjob submission

Instead of generating task files for each subjob and submit jobs one by one, JSUB accelerate the process of task creation by only generating a common task file template for all subjobs and submit them with subjob-specific parameters. With the feature of parametric submission on DIRAC, the submission process is also greatly accelerated (10,000 jobs in 5 minutes).

Flexible task splitting

To support complex needs from experiment users, an advanced task splitter has been developed to allow for customable task setting. Users can define their subjob-specific variables lists (jobvars), cite their values in task settings, and combine them into variable sets to determine the splitting of tasks. This splitter module gives users large freedom and good control over the details of their tasks.

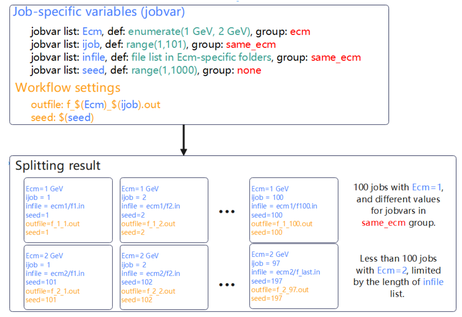

Figure 4:An example task with jobvar splitter, where subjob parameters are combined by groups to determine task splitting and their values are cited by workflow settings.

High extensibility

JSUB is designed to be highly extensible to different backend platforms and physics experiments. The nonspecific outline of task management processes are handled by common modules, while the varying details are handled by extension modules that are dynamically loaded at runtime.

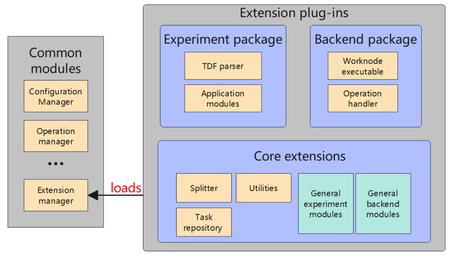

Figure 5:Extensibility design of JSUB

Usage of JSUB

JSUB aims to serve JUNO, CEPC, and potentially other physics experiments.
Since 2020, JSUB has been used in JUNO experiment. It has been tested that actual tasks for simulation and reconstruction of Monte Carlo samples can be successfully handled by the software with good performances, while supports for user analyses have been developed and tested with mock-up tasks and need further improvement in robustness and user-friendliness with more user feedbacks after JUNO starts taking data.

Conclusion

A lightweighted task submission and management software has been developed for the needs of users in JUNO and CEPC experiments.
Major features of the software include parametric job submission, flexible task splitting, and high extensibility.

References

[1] Zelimir Djurcic et al.
JUNO Conceptual Design Report.
8 2015.

[2] Mingyi Dong et al.
CEPC Conceptual Design Report: Volume 2 - Physics & Detector.
11 2018.

[3] Brook N. et al.
DIRAC - Distributed Infrastructure with Remote Agent Control.
arXiv e-prints, page cs/0306060, June 2003.

Contact Information

- Email: yangyf@mail.ihep.ac.cn