Procedia Computer Science

Volume 80, 2016, Pages 690–699

ICCS 2016. The International Conference on Computational Science

Running simultaneous Kepler sessions for the parallelization of parametric scans and optimization studies applied to complex workflows

Michał Owsiak[1], Marcin Płóciennik[1], Bartek Palak[1], Tomasz Zok[1], Cedric Reux[2], Luc Di Gallo[2], Denis Kalupin[3], Thomas Johnson[4], and Mireille Schneider[2]

[1] Poznań Supercomputing and Networking Center IBCh PAS, Poznan, Poland
michalo@man.poznan.pl, marcinp@man.poznan.pl, bartek@man.poznan.pl, tzok@man.poznan.pl
[2] CEA, IRFM, F-13108 Saint-Paul-lez-Durance, France
luc.digallo@cea.fr, cedric.reux@cea.fr, mireille.schneider@cea.fr
[3] EUROfusion Programme Management Unit, Boltzmannstr. 2, 85748 Garching, Germany
denis.kalupin@euro-fusion.org
[4] Fusion Plasma Physics, EES, KTH, SE-10044 Stockholm, Sweden
johnso@kth.se

Abstract

In this paper we present an approach taken to run multiple Kepler sessions at the same time. This kind of execution is one of the requirements for Integrated Tokamak Modelling platform developed by the Nuclear Fusion community within the context of EUROFusion project[2]. The platform is unique and original: it entails the development of a comprehensive and completely generic tokamak simulator including both the physics and the machine, which can be applied for any fusion device. All components are linked inside workflows. This approach allows complex coupling of various algorithms while at the same time provides consistency. Workflows are composed of Kepler and Ptolemy II elements as well as set of the native libraries written in various languages (Fortran, C, C++). In addition to that, there are Python based components that are used for visualization of results as well as for pre/post processing. At the bottom of all these components there is a database layer that may vary between software releases, and require different version of access libraries. The community is using shared virtual research environment to prepare and execute workflows. All these constraints make running multiple Kepler sessions really challenging. However, ability to run numerous sessions in parallel is a must - to reduce computation time and to make it possible to run released codes while working with new software at the same time. In this paper we present our approach to solve this issue and examples that show its correctness.

Keywords: Kepler, parallel execution, parametric scan

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016

doi:10.1016/j.procs.2016.05.362

Acknowledgement

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

1 Introduction

Integrated modelling efforts for ITER[1] experiment focuse currently on the assessment on the validation of comprehensive models against present facilities results. In view of supporting ITER operation, modelling tools are necessary, both for pulse validation and plasma control. The Nuclear Fusion community has developed for this purpose the software infrastructure framework for integrated modelling activities as well as a validated suite of simulation codes, platform that is currently operated under EUROFusion[1] project.

EUROfusion strategy includes the integration of the most advanced EU fusion codes into a centrally maintained suite of Integrated Modelling tools. Kepler plays major role in this development as it serves as a basis for integration of all the components and codes developed as part of the project.[2] The work follow up the Integrated Tokamak Modelling Task Force that operated under EFDA from 2004 until 2013.

Within EUROFusion WPCD, Kepler is used as a basis for linking various components that are used in numerical computations. Each of these components can be developed in programming language that is supported by ITM (Integrated Tokamak Modelling) platform: C/C++, Fortran, Java, Matlab, Python. All these components (developed separately) are linked together to form workflows performing numerical computations. Different workflows focus on different aspects of plasma simulation. Thanks to using Kepler, each component (numerical code) can be wrapped by Java code and exposed as Kepler actor. This way, regardless of the workflow type, each actor (numerical code) can be easily reused without too much effort.

After workflows are released, there are two main ways of using them. First one, is to use released workflow for actual simulations, second one is to optimize it. These two actions require two, different, installations of Kepler. Both should be able to run at the same time. Apart from that, there is another requirement related to running multiple Kepler sessions at the same time, that is batch execution[10]. In order to reduce computation time one typically runs numerous Kepler sessions running the same workflow with different parameters (see Figure 1).

This way, it is possible to run multiple simulations at the same time. However, this is not an easy task to achieve when we talk about Kepler being run at batch nodes in multi user environment.

Current developments are influenced by numerous factors that affect execution of Kepler in parallel. These factors are either result of project's specifics or are based on internal Kepler's limitations. We will discuss these factors in next section.

2 Limitations of Kepler's mechanisms

Kepler itself provides solutions for running multiple Kepler instances at the same time. However, these solutions are not fully applicable in case of EUROfusion based developments. We are

[1] https://www.euro-fusion.org
[2] http://portal.efda-itm.eu/itm/portal/