

---

# HERCULE data management library : boosting RAMSES I/O for exascale era

Loic Strafella\*<sup>1</sup>

<sup>1</sup>Institut de Recherches sur les lois Fondamentales de l'Univers – Commissariat à l'énergie atomique et aux énergies alternatives : DRF/IRFU/DEDIP/LILAS – France

## Abstract

### Résumé:

Les sorties du code RAMSES sont aujourd'hui uniquement de type protection/reprise et sont principalement basées sur un format binaire dédié. Chaque snapshot d'une simulation peut occuper plusieurs téraoctets de données répartis sur plusieurs dizaines de milliers de fichiers, pénalisant la charge du réseau et le système de fichiers se traduisant par des temps d'écriture particulièrement important et une fragilité la scalabilité du code dans une perspective de calcul exaflopique. De plus, bien que le format de protection/reprise permette de relancer une simulation en cas d'arrêt imprévu, il nécessite une conversion vers un format de "dépouillement", plus léger, pour la visualisation et l'analyse des données à l'aide d'outils spécifiques tels que PyMSES. Les objectifs de ces travaux sont d'intégrer dans RAMSES une librairie d'entrées/sorties parallèles développée au CEA DAM appelée HERCULE, qui permettra de générer une sortie du type protection/reprise tout en améliorant les performances d'entrées/sorties du code ainsi que sa scalabilité. De plus, l'implémentation d'un format plus compacte spécifique au dépouillement directement dans RAMSES apportera un gain de productivité lors des phases de transfert sur le réseau et accélérera la visualisation et l'analyse des résultats par PyMSES, auquel sera ajouté un module de lecture des sorties au format HERCULE. La librairie de gestion de données et un convertisseur de format RAMSES-legacy/HERCULE seront présentés aux utilisateurs ainsi que les résultats de benchmarks de performances des entrées/sorties. Une introduction au nouveau format de dépouillement HERCULE projettera l'utilisateur sur les futures fonctionnalités qu'offre l'intégration de la librairie.

Abstract:

---

\*Speaker

As of today, RAMSES code outputs are checkpoint/restart outputs only, written as binary files in a specific format. Every simulation snapshot can be very large up to terabytes, split on dozens of thousands of files, thus hindering the network and filesystem load resulting in large I/O times and in a poor scalability of RAMSES code in the exascale context. Moreover, even if the checkpoint/restart file format contains all informations required by the code in case the simulation needs to be restarted upon failure, it needs to be post-processed to a lightweight format before any visualisation or data analysis by dedicated tools such as PyMSES.

The two main objectives of this work are to integrate in RAMSES a parallel I/O library called HERCULE developed at CEA DAM (Military Applications Division), thus improving RAMSES I/O performance and scalability. Furthermore, the implementation of a lightweight output format dedicated to post-processing within RAMSES will result in faster data transfer and improve significantly the overall performance of visualisation and analysis dedicated tools such as PyMSES, to which a specific reader for HERCULE format will be added. The HERCULE library and a RAMSES-legacy/HERCULE data converter, as well as some I/O performance benchmark results (checkpoint/restart output format comparison) will be presented to the RAMSES user community.

An introduction to the HERCULE "post-processing" output format will enlighten the RAMSES users with the future performance and functionalities provided by the integration of this new library.