Contribution ID : **5**   Type : **not specified**

Arkadiusz Hypki (UAM) Distribute data analysis of Big Data and machine learning applied on a large number of detailed MOCCA numerical simulations

Wednesday, 23 October 2024 14:15 (15)

MOCCA code is able to perform detailed numerical simulations of globular clusters of any size within a few days (http://moccacode.net). Because of its speed and a close agreement with N-body codes MOCCA code is perfect to perform a grid of simulations for various initial conditions. It is currently beyond the capabilities of any N-body codes. At this moment our MOCCA database consists of over 2500 models for broad range of initial parameters (200 TBs). To handle such amount of data from numerical simulations I created BEANS - a general purpose web-based software for interactive distributed data analysis with a clear interface for querying, filtering, aggregating, and plotting data (http://beanscode.net). BEANS software relies on software which already proved its value in the industry like Apache Hadoop (distributed processing), Apache Pig (high level language for Apache Hadoop), Elastic (petabyte scale search engine) and more. Recently, a plugin to BEANS was added, which provides simple interface to train machine learning algorithms and then test their predictions. During the talk I will demon- strate how one can query huge amount of MOCCA numerical simulations with Apache Pig to build a training set with some parameters describing the numerical models of GCs (e.g. core half-mass radii, relaxation time, pre/post core collapse phase). Then, using a few different machine learning algorithms, I will show how one can train them and finally test the predictions. More specifically, the BEANS software will be used to show how one can automatically determine the dynamical age of GCs (whether a GC is in pre- or post-collapse phase).

Session Classification : Status of in-kind program, astroinformatics and synergies