H L R S

University of Stuttgart
Germany

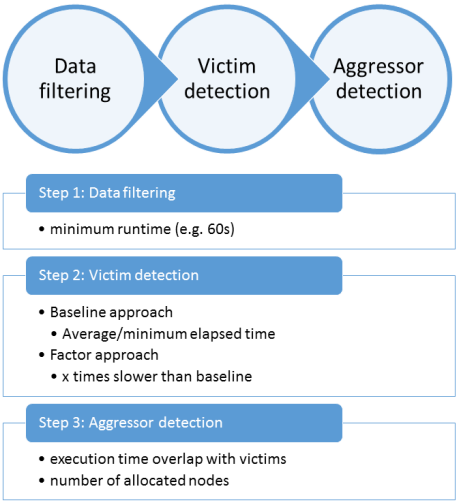
CATALYST

Combining HPC and High Performance Data Analytics for Academia and Industry

At the High Performance Computing Center Stuttgart (HLRS), customers tend to execute more and more data-intensive applications. One Petabyte of result data for large-scale simulations is not uncommon anymore, and since systems grow, data will as well. Since it no longer becomes feasible that data is processed and analysed manually by domain experts, HLRS and Cray Inc. have launched the CATALYST project to advance the field of data-intensive computing by converging HPC and Big Data to allow a seamless workflow between compute-intensive simulations and data-intensive analytics. For that purpose, Cray Inc. designed the Urika-GX data analytics hardware, which supports Big Data technologies and furthermore, enhances the analysis of semantic data. This system has been installed as an extension of Hazel Hen, the current HPC-flagship system of HLRS.

Main objective of CATALYST is to evaluate the hardware as well as the software stack of the Urika-GX and its usefulness with a particular focus on applications from the engineering domain. As the majority of today's data analytics algorithms are oriented towards text processing (e.g. business analytics) and graph analysis (e.g., social network studies), we are further in need to evaluate existing algorithms with respect to their applicability for engineering. Thus, CATALYST will examine future concepts for both hardware and software.

The first case study conducted in collaboration with Cray Inc. addresses the performance variations of our Cray XC40 system. Performance variability on HPC platforms is a critical issue with serious implications for the users: irregular runtimes prevent users from correctly assessing performance and from efficiently planning allocated machine time. Thus, monitoring today's IT infrastructures has actually become a big data challenge on its own. The analysis workflow used to identify the causes of runtime variations consists of three steps including different configuration parameters:



1. *Data filtering*
2. *Detection of applications that show high variability (Victim)*
3. *Detection of applications that potentially causing the variability (Aggressor)*

Outlook:

- Big Data application evaluation
- Close cooperation with partners from both, industry and academia
- Seamless integration of the Big Data system into our existing HPC infrastructure
- Develop and evaluate practical case studies to advertise the solution

Project Information:

Runtime: October 2016 – September 2019

Funding Organization: Ministry of Science, Research and the Arts Baden-Württemberg

Partners: HLRS, Cray Inc. & Daimler AG (associated)

Contact: Michael Gienger
High Performance Computing Center Stuttgart
Nobelstraße 19, 70569 Stuttgart, Germany
Phone: +49 (0) 711/ 685-65891
E-Mail: gienger@hlrs.de