

## Distributed Radio Interferometric Calibration using SAGECal

The SKA will deliver tremendously high data rates when it becomes operational. In order to handle that amount of data, processing pipelines should be very efficient. SAGECal is one of the popular interferometric calibration tools capable of handling such data rates. SAGECal uses GPU acceleration and distributed computing using MPI in order to this. In order to meet the challenges posed by the SKA, SAGECal needs big data analytics to improve its robustness and scalability. Big data frameworks such as Apache Spark is a good option for these tasks.

In this work, we have integrated SAGECal into a big data ecosystem, namely Apache Spark. We set up a cluster for benchmarking and deployed it to SURFsara HPC Cloud platform. In order to deploy the services, such as Apache Spark and Apache Hadoop, we have developed a tool which makes the deployment easier (see <https://github.com/NLeSC/lokum> and <https://github.com/NLeSC/baklava>). All other components used in this work are publicly available at <https://github.com/nlesc/Dirac>. The setup consists of Apache Spark components, a master node and slave nodes, and storage units. Measurement sets (MS) from radio astronomical observations are stored in a Hadoop Distributed File System (HDFS) and processed by Apache Spark. As SAGECal is written in C/C++/CUDA and Apache Spark does not have a native support for this language, we have used Java Native Interface (JNI) to generate a compatible version of SAGECal.

Previously, we presented the use of these tools in a virtual cluster which was created by Docker swarm. This time, we will present our results using a real cluster. We will also present the adaptation of SAGECal code base for the Apache Spark platform. Moreover, the performance comparison of MPI and Apache Spark versions of SAGECal will be shown. The technical details of the setup and the software architecture will be also presented.

### Suggested duration

**Primary authors:** DIBLEN, Faruk (Netherlands eScience Center); YATAWATTA, Sarod (Kapteyn Institute)

**Presenter:** DIBLEN, Faruk (Netherlands eScience Center)

**Session Classification:** Science Operations I

**Track Classification:** Science Operations