

sPHENIX data processing

The production system needs to be able to manage jobs on the grid to cover simulation needs during data taking, user analysis jobs, our coordinated analysis (analysis train), the near online and second pass processing. The number of simultaneous active jobs is in the multi 100k range.

The sPHENIX collaboration will use PanDA as workflow management system. Our requirements are well within the capabilities of PanDA and we have local expertise in the NPPS group. Furthermore the plan of ATLAS to establish a PanDA server in SDCC provides a good opportunity to join forces and set up a system in parallel for use by sPHENIX. All of this makes PanDA the logical choice for the sPHENIX workflow management system.

The sPHENIX collaboration will use Rucio for its data management. It is well integrated into PanDA and its capabilities make it suitable to manage all temporary files. SDCC already runs a Rucio instance and via the NPPS group we have access to local expertise.

Our goal is to have a first implementation of our data processing scheme for the next planned Mock Data Challenge (MDC2) scheduled to start in December 2021.