

# Evaluation of a new data staging framework for the ARC middleware

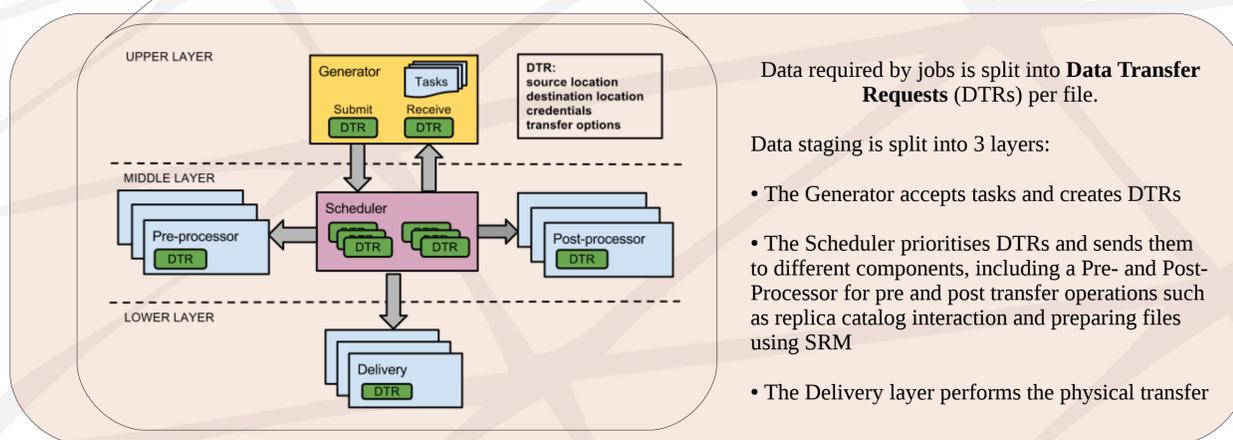
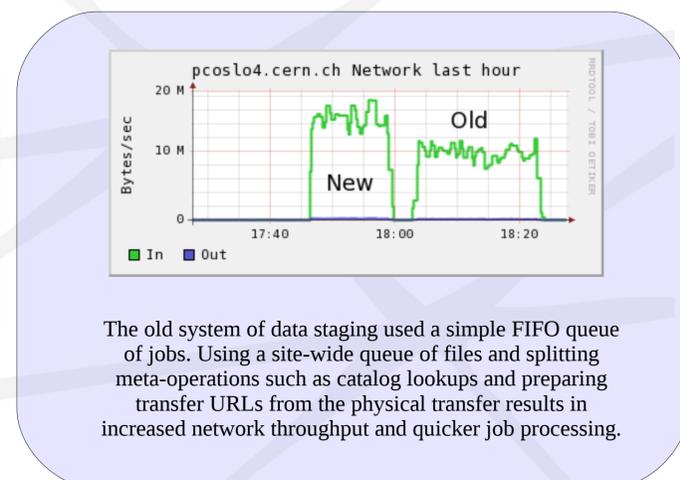
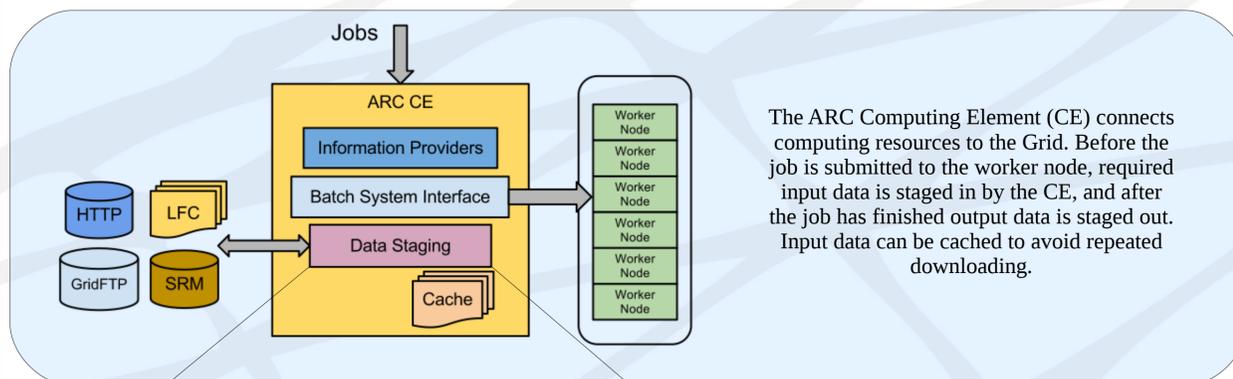
D. Cameron<sup>1</sup>, A. Filipcic<sup>2</sup>, D. Karpenko<sup>1</sup>, A. Konstantinov<sup>1,3</sup>

<sup>1</sup>University of Oslo, Norway

<sup>2</sup>Jozef Stefan Institute, Slovenia

<sup>3</sup>Vilnius University, Lithuania

Staging data to and from remote storage services on the Grid for users' jobs is a vital component of the ARC computing element. A new data staging framework for the computing element has recently been developed to address issues with the present framework, which has essentially remained unchanged since its original implementation 10 years ago. This new framework consists of an intelligent data transfer scheduler which handles priorities and fair-share, a rapid caching system, and the ability to delegate data transfer over multiple nodes to increase network throughput. We use data from real user jobs running on production ARC sites to present an evaluation of the new framework. It is shown to make more efficient use of the available resources, reduce the overall time to run jobs, and avoid the problems seen with the previous simplistic scheduling system. In addition, its simple design coupled with intelligent logic provides greatly increased flexibility for site administrators, end users and future development.



The time was measured to run jobs requiring several different datasets using the old and new system.

Dataset	No files	Ave file size	Total size
Dataset A	600	500 KB	320 MB
Dataset B	39	15 MB	622 MB
Dataset C	200	50 MB	10 GB
Dataset D	20	1.5 GB	29 GB

