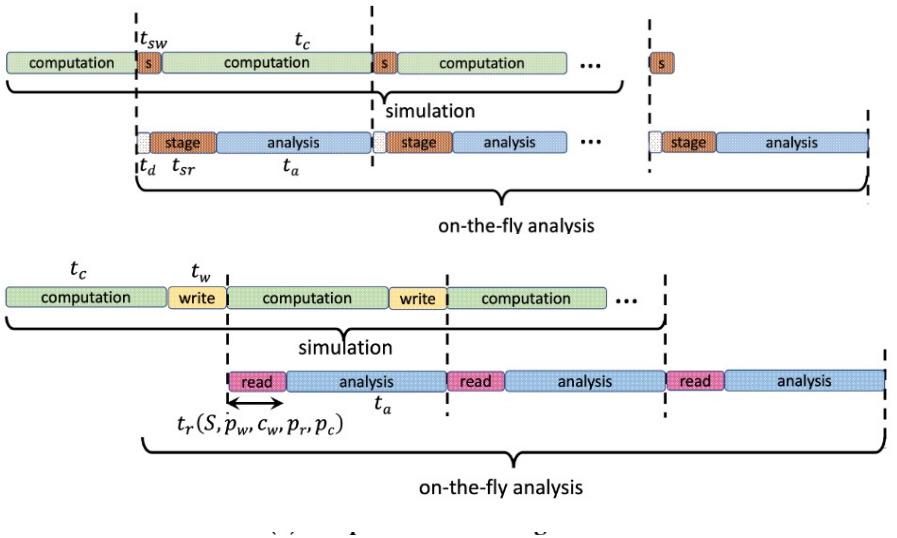
Data staging performance for coupled simulation workflows

Scientific Achievement

A comprehensive study that assesses the robustness of state-of-the-art data staging solutions to optimize the data movement performance of coupled workflows on large scale systems based on their behavior.

Significance and Impact

The study extracts the relation between the application behavior and the effectiveness of staging techniques. It highlights configurations that might lead to lower performance and/or higher cost and gives recommendations for emerging applications.

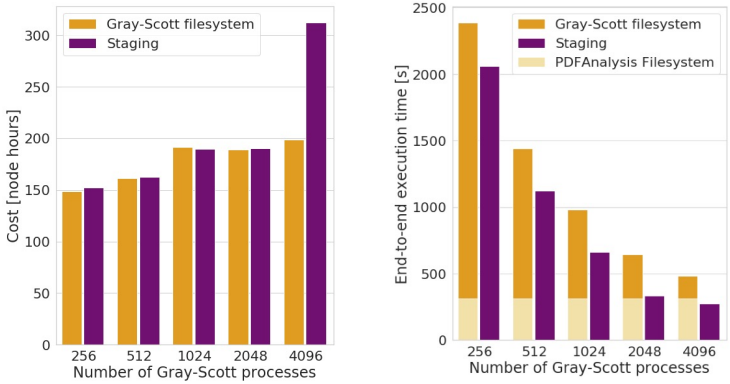


Streaming vs filesystem data movement

Research Details

The study looks at three main computational techniques: a stencil code, particle-in-cell and neural networks. The main findings show:

- Streaming algorithms deliver faster time to solution compared to filesystems but with the price of higher node hours
- Inline staging algorithms are zero-copy methods but achieve best performance only for embarrassingly parallel analysis codes that do not require interaction between processes
- Storing the data to storage can be hidden during staging as long as the write is done in the correct process



Performance of staging compared to filesystems for the Gray-Scott application

U.S. DEPARTMENT OF ENERGY | Office of Science

Work was performed at Oak Ridge National Laboratory

OAK RIDGE National Laboratory