After more than 10 years of operation, the beamlines at the Australian Synchrotron are well established and the demand for automation of research tasks is growing. In order to meet these demands we have developed a generic distributed workflow system.

Lightflow models a workflow as a set of individual tasks arranged as a directed acyclic graph (DAG). This specification encodes the direction that data flows as well as dependencies between tasks. Each workflow consists of one or more DAGs. While the arrangement of tasks within a DAG cannot be changed at runtime, other DAGs can be triggered from within a task, therefore enabling a workflow to be adapted to varying inputs or changing conditions during runtime.

The phaseID pipeline at the SAXS/WAXS beamline identifies diffraction peak positions within SAXS profiles and infers the most likely space group. This pipeline enables researchers to rapidly determine phase diagrams for self-assembled lyotropic liquid crystal systems.

The data compression and management pipeline at the two crystallography beamlines (MX1, MX2) has recently been upgraded to use Lightflow in order to take advantage of a distributed system to compress multiple experiments at the same time.

Lightflow employs a worker-based queuing system, in which workers consume individual tasks. This allows the processing of workflows to be distributed. In order to avoid single points of failure, such as a central daemon as is often found in other workflow tools, the queuing system is also used to manage and monitor workflows and DAGs.

Tasks can receive data from upstream tasks and send data to downstream tasks. Any data that can be serialized can be shared between tasks. Typical examples for data flowing from task to task are file paths, pandas DataFrames or numpy arrays.

Lightflow provides a fully featured command line interface for starting, stopping and monitoring workflows and workers.

The source code has been published as open source on GitHub and on PyPI.

https://github.com/AustralianSynchrotron/Lightflow
https://australiansynchrotron.github.io/Lightflow

Architecture

Implementation

Examples

Workflow definition

Download

The phaseID pipeline at the SAXS/WAXS beamline identifies diffraction peak positions within SAXS profiles and infers the most likely space group. This pipeline enables researchers to rapidly determine phase diagrams for self-assembled lyotropic liquid crystal systems.

The data compression and management pipeline at the two crystallography beamlines (MX1, MX2) has recently been upgraded to use Lightflow in order to take advantage of a distributed system to compress multiple experiments at the same time.