Decaf: Decoupled Dataflows for In Situ Workflows

Motivation
The need to distill enormous amounts of data into useful knowledge is pushing the limits of computational science. Tightly coupled data analysis and data generation—making the analysis interdependent and closely coordinated with the computation—limits the flexibility provided by individual modules. The Decaf project explores a hybrid approach that combines both types of coupling—tight and loose—in effect decoupling tightly coupled applications.

An example workflow in cosmology transforms raw particle positions in an N-body simulation into a Voronoi mesh, which is then used to deposit particle density onto a regular grid. Subsequent density statistics are computed in postprocessing.

---

Dataflows for Workflows

A workflow is a directed graph of tasks and communication between them. The graph can have cycles. Graph nodes are the tasks and links are the communication. Both nodes and links are actually parallel processes. A dataflow is the communication over the links in a workflow. Workflows consist of high-level tasks; dataflows consist of communication between MPI process ranks. For every producer-consumer pair in the workflow, 5 MPI communicators are used in the dataflow.

---

Programming Model

Below: The workflow software stack consists of a workflow definition layer, a dataflow definition layer, and a transport layer.

- **Dataflow Definition and Runtime**
  - Dataflow Definition: Specifies the tasks and dataflows.
  - Dataflow Runtime: Coordinates and manages the execution of the tasks.

- **Transport Layer**
  - Communicators: Used in the dataflow to handle communication between tasks.
  - Executors: Execute the tasks in parallel.

The core of Decaf is the dataflow layer. Other workflow definition tools can be used, and transport layers are provided by systems software such as MPI.

---

Data Redistribution

Below: Four redistribution patterns are available in Decaf: left-to-right round-robin, contiguous, bounding box, and block. Others can be customized by the user.

Redistributing data models requires splitting and merging while preserving semantics. Specifying a particle data model (right) allows Decaf to split and merge the particles correctly (below).

---

Scalable Performance

Below: Good scaling of cosmology workflow and executing tasks in space division mode results in total time equal to the slowest task, not the sum of the tasks.

---

Fault Tolerance

Right: Silent data corruption in a time-stepping solver detected with our inexpensive auxiliary method that outperforms the best known detectors (AID and BBS14) for numerical integration.

Bottom: Modeling the dataflow and optimally adding replication and roll back to recover from hard and soft errors detected above. Bottom right: Comparing group and process duplication and triplication.

---

Applications

An in situ cosmology workflow that converts dark matter tracer particles from an N-body simulation to a density image is used to estimate the gravitational field for lensing.

An in situ biology workflow for computational steering of an iron complex through a protein channel of a bacteria cell is used to design new drugs.

---

The Decaf Project   https://bitbucket.org/tpeterka1/decaf

---

Tom Peterka, Jay Lofstead, Franck Cappello, Patrick Widener, Matthieu Dreher, Clement Mommesin, Pierre-Louis Guhur, Aurelien Cavelan, Alexis Champsaur