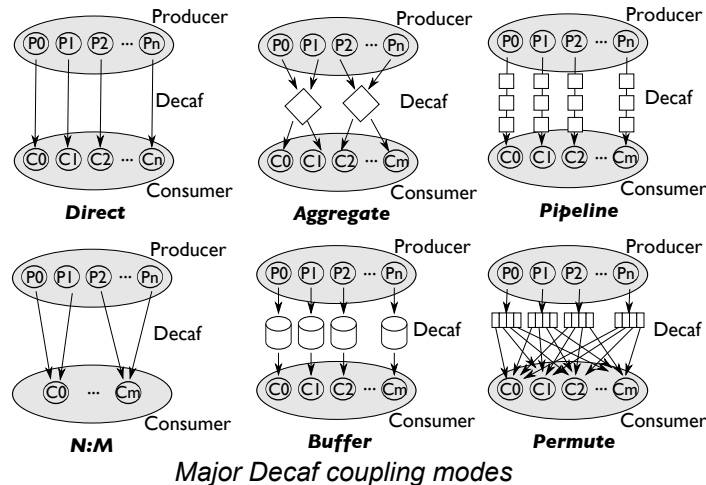


Decaf: Decoupling Tightly-Coupled Data Flows



Novel Ideas

Decaf is a dataflow middleware providing scalable and generic connections of producers and consumers in a data analysis workflow.

- Decouples tightly-coupled workflow links into general dataflows
- Dataflows are composable from small set of primitives
- Automatic buffering for persistence, flow control, and resilience
- Fault tolerance to hard and soft errors in dataflows

Impact and Champions

Decaf is vital to the success of ASCR's mission to develop extreme-scale scientific data analysis:

- In situ analysis coupled to simulations
- Analysis tasks coupled together in workflows
- Addresses HPC architecture challenges: resilience, concurrency, heterogeneity
- Scalability, productivity, and usability
- Integrates with Argo and Hobbes OS/R research
- Application drivers: cosmology, reactor design, superconductivity

Principal Investigator(s): Tom Peterka (PI), ANL; Jay Lofstead (coPI), SNL; Franck Cappello (coPI), ANL

Milestones/Dates/Status

Year	1				2				3			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4
Data Description	Develop schema and API				Integrate into transport and dataflow				Integrate resilience and test			
Transport	Develop API for primitives and develop pipeliner and buffer				Develop aggregator and selector				Integrate resilience in primitives			
Dataflow	Develop basic dataflow deployment computation				Integrate dataflow computation in Decaf				Incorporate resilience in dataflow			
Resilience	Strong, weak, none process failure resilience				Weak resilience to silent data corruption				Evaluate in proxy and full apps			

Project timetable with interconnections between tasks