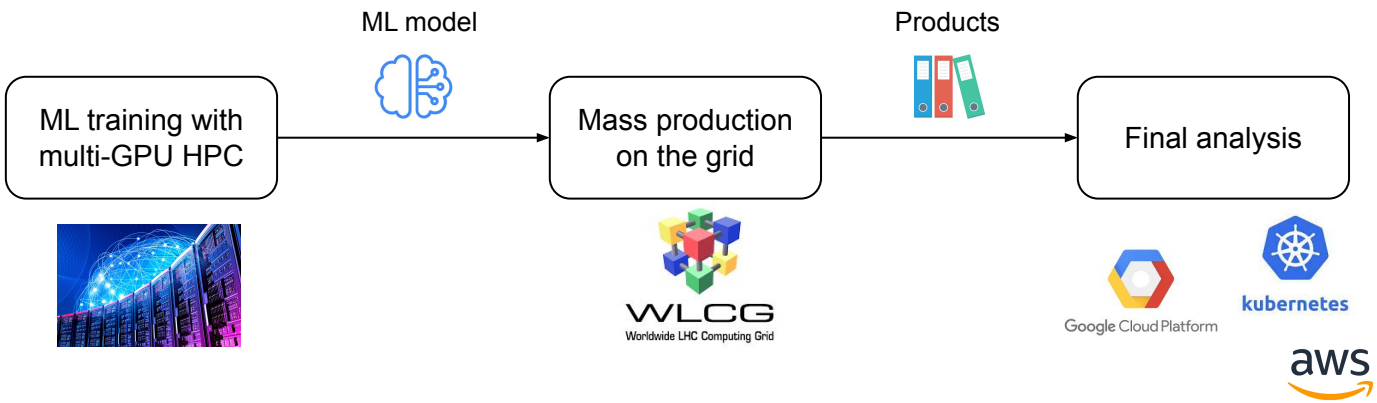
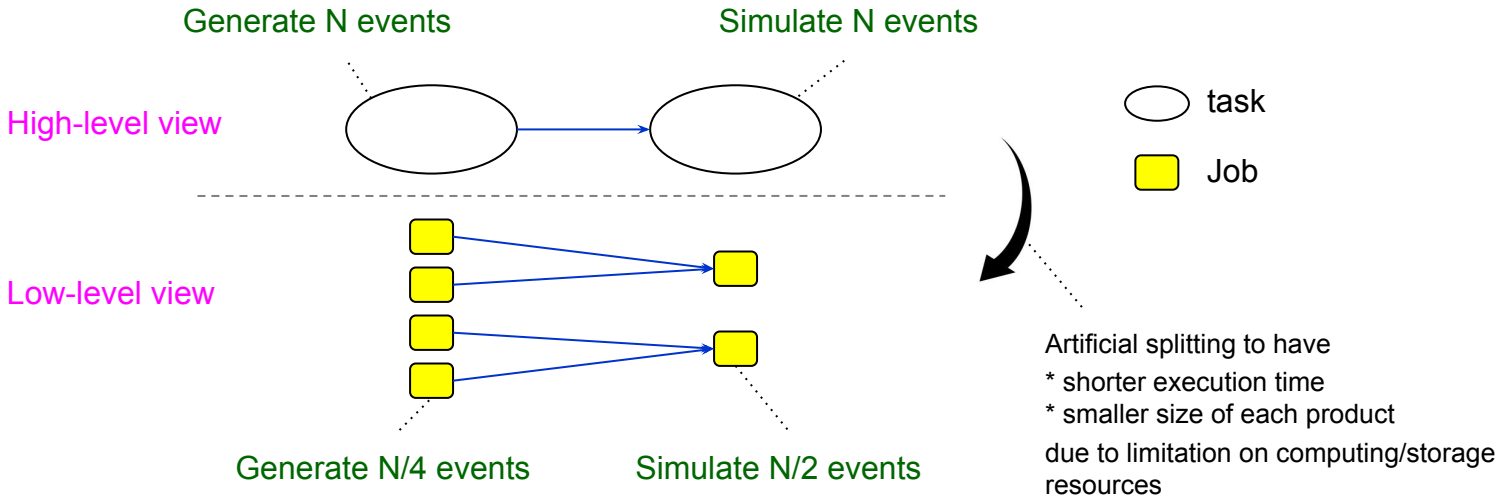
Running Workflow with PanDA and iDDS

- Thanks to iDDS, the PanDA system now has the capability to run workflow (chain of tasks) with
  - Conditional branching
  - Nested workflows
  - Parallelization with scatter
  - Loops (coming)
- Workflow description with directed acyclic graph (DAG)
  - Task-level
    - Description of relationship among tasks
  - Job-level
    - Description of relationship among jobs across task boundaries
- Leverages PanDA's capability to integrate heterogeneous and geographically-distributed resources



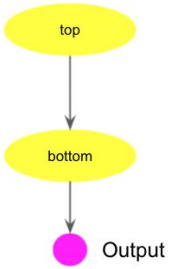
Running Workflow with PanDA and iDDS

- Issues with existing workflow automation systems like AirFlow, Luigi, Prefect, etc
  - Embedded resource management and workload scheduling
    - Doesn't fit well with PanDA's mechanisms
  - Low-level workload orchestration
    - Task in their terminology = Job in PanDA
      - Atomic entity running minimal workload on a real resource
    - Don't support high-level workflow descriptions
      - Not declarative so that users have to describe the details
      - Generally forbid the child step to get started before the parent steps are done
      - Possible to stream output from the parent while it is running, but it is just a pipe and tricky to run both parent and child steps properly in parallel

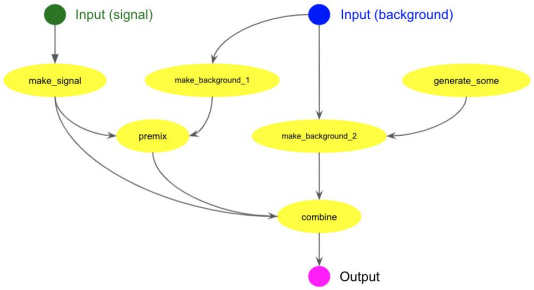


Running Workflow with PanDA and iDDS

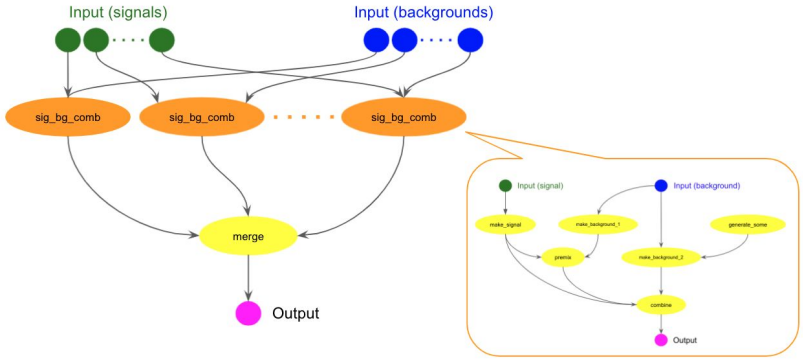
- A new client tool, pchain, in panda-client for users to describe arbitrary workflows in CWL
  - Documentation: Link
  - Available in 1.4.86 or higher on PyPI
  - Examples
    - Parent-child chain of prun tasks



    - More complicated chain

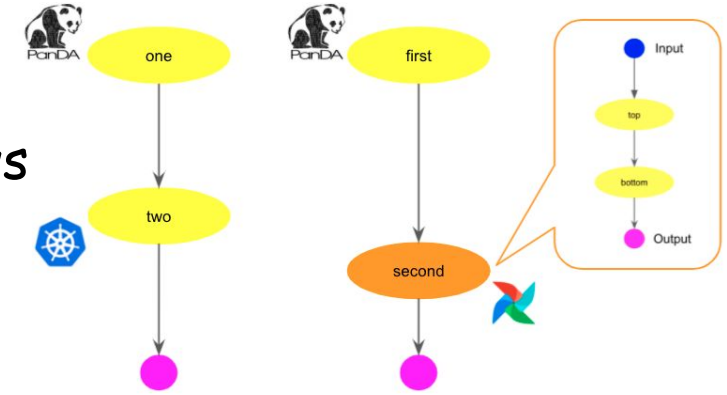


    - Parallelization with scatter and sub-workflow



Running Workflow with PanDA and iDDS

- Future plans
  - New plugins in iDDS for non-PanDA backends
    - E.g., running bare payloads on local batch systems or k8s clusters to have quick tasks using local resources, offloading sub-workflows to other automation systems, …



  - Workflow monitoring
  - Looping support for the workflows that require iterations with feedbacks from old execution, such as active learning
  - Add user interface to pchain for job-level DAG in addition to task-level DAG
  - Support for other workflow description languages
  - Visual interface to prepare CWL files rather than directly editing yaml