

iDDS Status and Plans

Wen Guan, Tadashi Maeno
Brian Bockelman, Torre Wenaus
Fernando Barreiro, Fahui Lin, Rui Zhang

Jan 21, 2021
ATLAS WFMS

iDDS Status

❖ New functions

- DAG based workflow management
 - High-level workflows specified by DAGs driving workload scheduling
- Task template support
 - To generate new tasks based on different parameters with same task template
- User Interface
 - Simply interface with reorganized APIs
 - A Command Line Interface is developed

❖ Use cases

- Data Carousel
 - Solved the issues with the delayed start of processing data on tape
- HPO (Hyper Parameter Optimization)
 - To provide a fully-automated platform for hyper-parameter optimization on top of geographically distributed GPU resources on the grid, HPC, and clouds
- LSST exercise
 - Using new DOMA PanDA instance with DAG support
- Active Learning
 - DAG based task chain management, new tasks generation based on a collection and learning task, prototype passed tests and deployed to test instance

❖ PyPI

- New version 0.1.0 is deployed to PyPI

Plans: Rubin LSST enhancement

❖ Function Improvements

- Current iDDS LSST plugin manages both task and files
 - Slow and the file info is not well management, the db usage is huge.
 - Enhancement of iDDS data components interface
 - Get rid of the file management from the LSST plugin and use the interface instead
- Errors propagation
 - Aggregation of errors from various sources to the original request, so that users who submit the request can easily understand what happens
- Monitoring of task chains
 - When there is a chain in the task, current interface is not well organized to show the relationship
 - Functions to generate DAG map
 - Functions to show tasks orderly based on DAG
- Improve documents

Plans: Integration of IRIS-HEP products

❖ **Close collaboration with other IRIS-HEP activities**

- To use those systems as data transformation/processing backends based on use cases
- Integration of ServiceX and Function-as-a-Service
 - Interface between iDDS and ServiceX
 - Task managements and data managements
 - Consistency
 - Data transformation definition
 - Output management and lifetime based deletion

Plans: others

❖ **Condition branching in DAG**

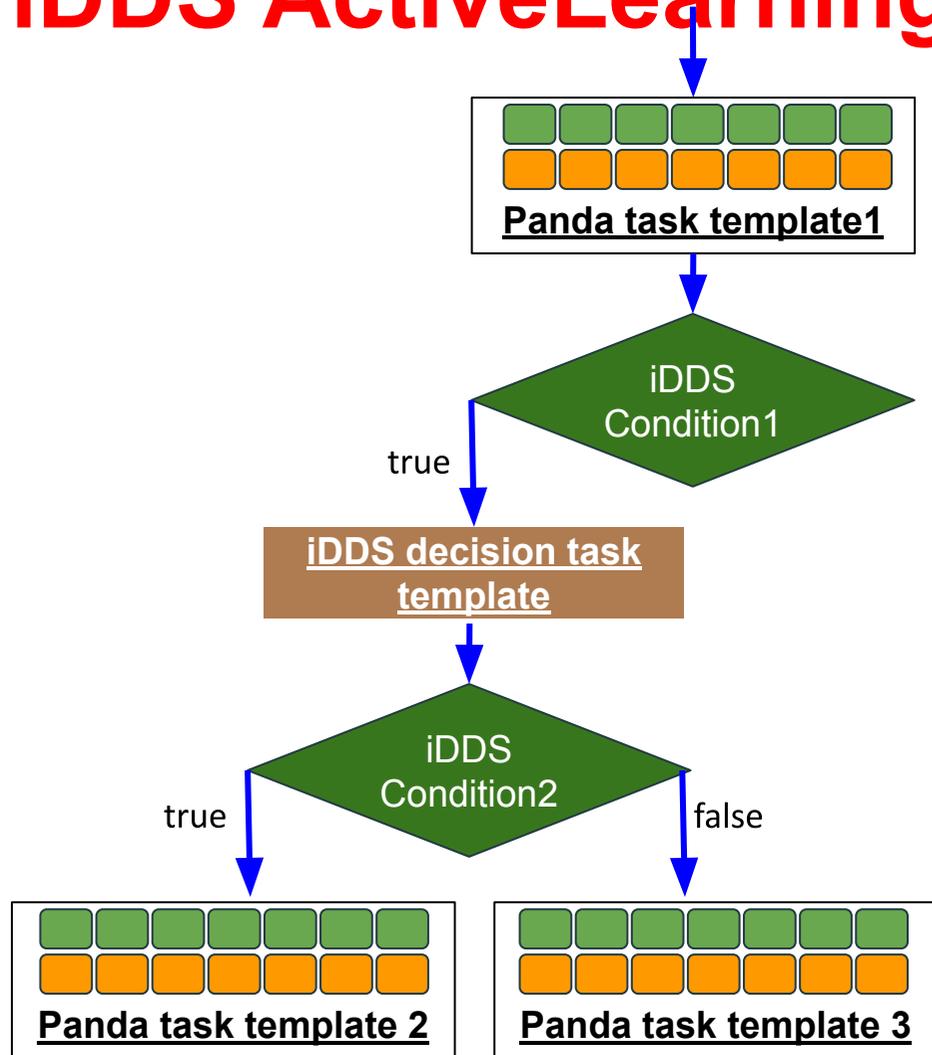
- Current condition branching is available only when a task is terminated. When a condition branch returns True or False, a `true_task` or a `false_task` is generated.
- Todo: New type of condition branching to support other operations, in addition to task generation.

❖ **Other Use Cases**

- Dynamic transformation and placement on demand, for example
Derivation on Demand
- Fine-grained data transformation and delivery, such as Event Streaming Service

backups

iDDS ActiveLearning



❖ Task template

- New task generation is based on template
- New parameters can be set when generating new tasks based on template
- iDDS decision task can generate parameters for following tasks
- Template1 and template 2 or 3 can be the same one, then a loop

❖ iDDS Condition

- Defines the relationship between different task
- The condition can be a function of its source.