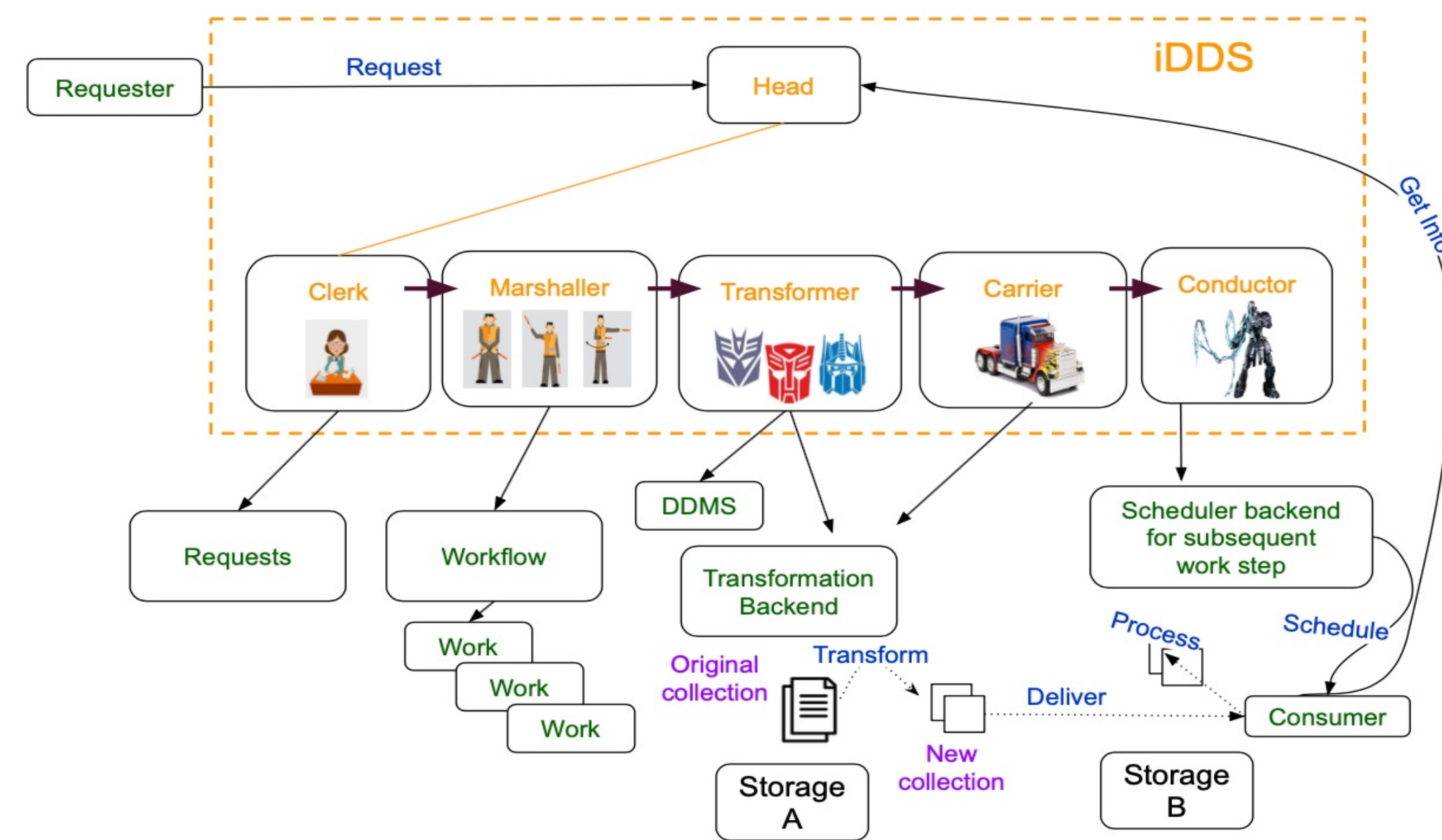


intelligent Data Delivery Service

A new service to decouple the expensive storage format and the processing format of the data, to intelligently select only required data for transferring and caching, and then deliver the required data to consumers in a fine-grained approach.

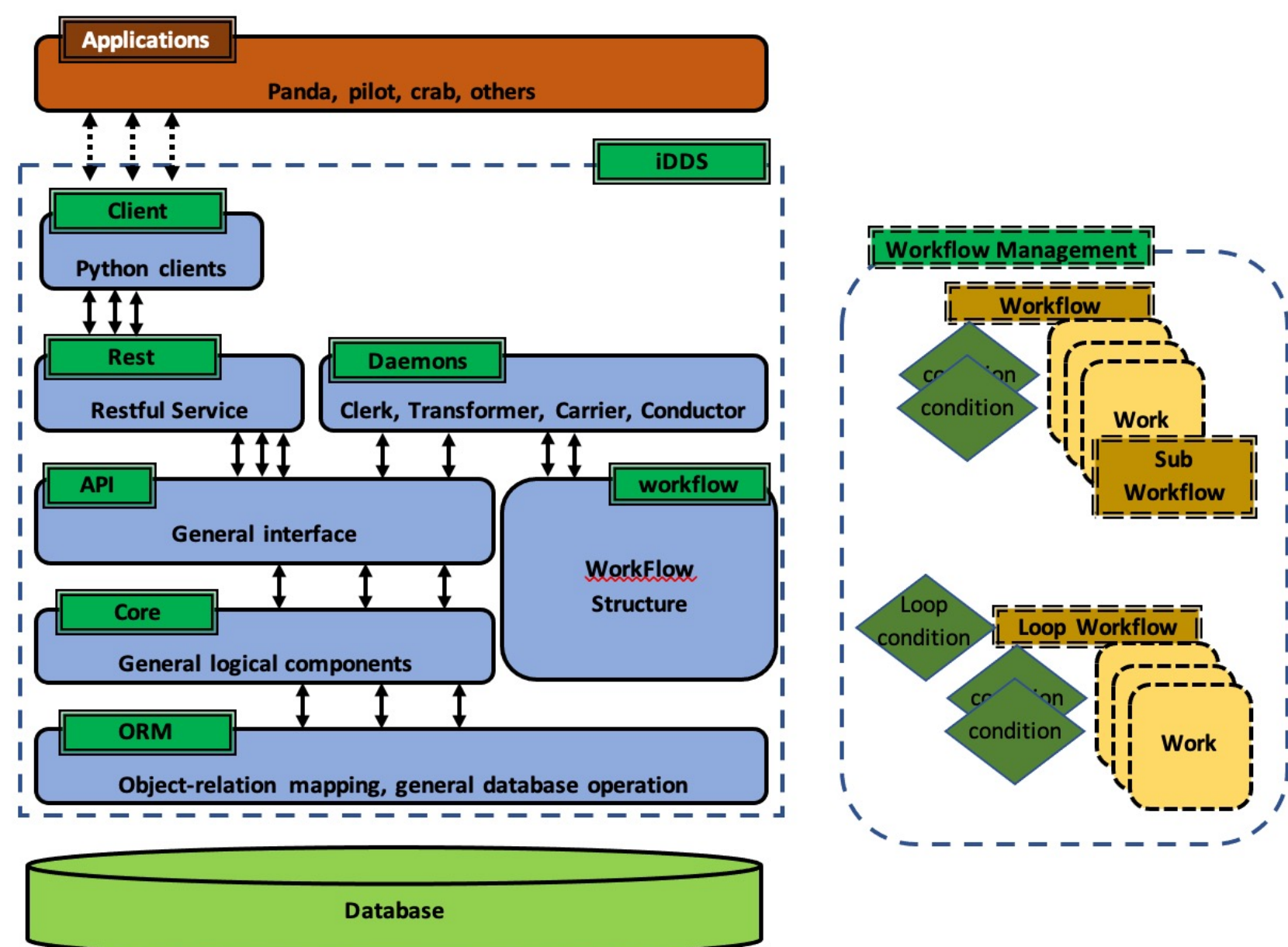
- **Transformation on demand:** Transform expensive data on demand at storage site to select only required data and then deliver the required format of data for processing to consumers.
- **Fine-grained delivery:** Coordinate with processing workflow to deliver data and remove data after processing in a fine-grained approach, without waiting for all data to be cached; Reduce local replica or cache usage and speedup processing.
- **Orchestration:** Orchestration between workflow management services and data management services for optimal usage of limited resources.
- **Intelligent:** Intelligent algorithms to apply data locality knowledge and processing requests to trigger on-demand transformations, fine-grained delivery and cache management to optimize the processing workflow and promote the cache reuse.

iDDS Components



- **HEAD:** Restful interface.
- **Marshaller:** Workflow management.
- **Transformer:** With different backends to transform data.
- **Carrier:** Manage transforms on different backends.
- **Conductor:** Notify/schedule consumers
- **Client:** To communicate with iDDS Restful service.

iDDS Architecture

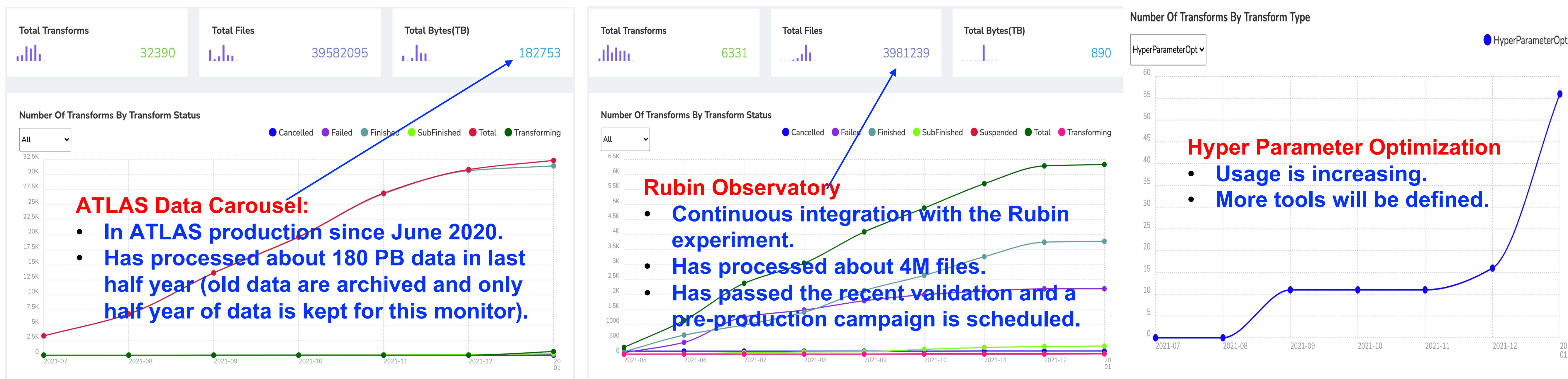


- **Experiment agnostic**
- **Abstract Layers for Generalization**
- **Workflow management**
 - DAG (Directed Acyclic Graph) management
 - Complicated workflow: SubWorkflow, LoopWorkflow, Template workflow, Dynamic Conditions

Applications

- **ATLAS Data Carousel:**
 - Orchestrate between data management system and workload management system to deliver data in proper granularities.
- **Rubin Observatory (LSST) exercise:**
 - Employ iDDS DAG to automate multiple-step processing.
 - Manage dependencies and trigger to release jobs incrementally.
- **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.
- **Automatic ToyMC generation:**
 - To provide an automated workflow to manage different steps of ToyMC generation.
- **Active Learning:**
 - Automate multiple-step processing based on some decision-making steps.
 - Task template to generate concrete tasks.
 - Condition branches to control the workflow.

Current Achievements



Status and Plans

- **Current Status**
 - Continuous support for current applications and improvements.
 - Complicated workflow improvements.
 - Deployment improvements and Cloud deployments.
- **Future plans**
 - Dynamic transformation and placement on demand.
 - Fine-grained data transformation and delivery.