

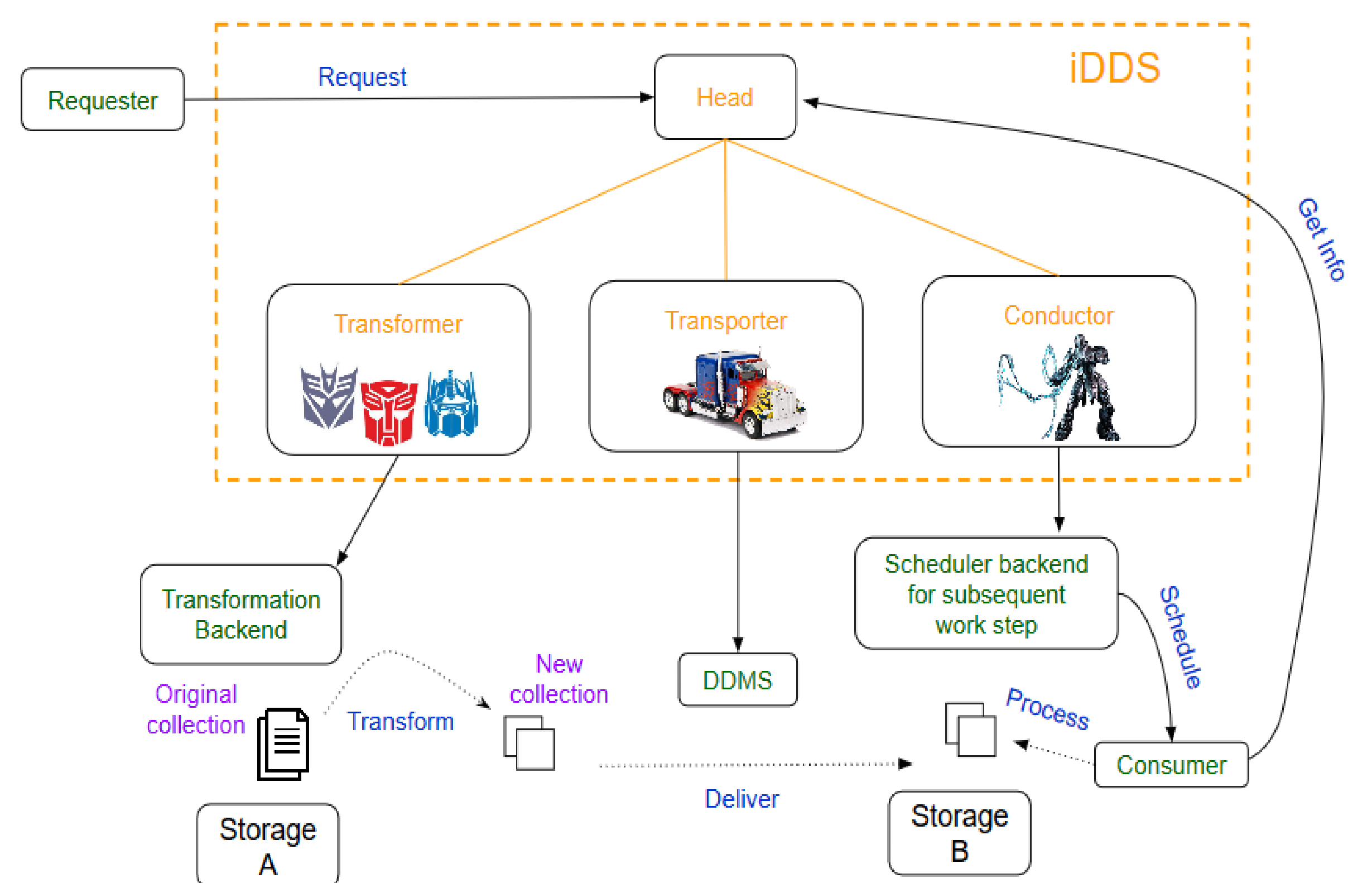
intelligent Data Delivery Service

- **A new service to decouple the expensive storage format and the processing format of the data, to intelligently transform and deliver needed data to consumers in a fine-grained approach.**
- **Transformation on demand:** Transform expensive data on demand at storage site and only deliver needed format of data for processing to consumers.
 - New format of data may improve overall processing throughput.
 - On-demand transformation will avoid produce unneeded data.
 - Transformation at storage site will avoid huge data replication.
 - Delivering cheaper transformed data will reduce local replica or cache usage.
 - Apply data locality knowledge and intelligence in caching process to promote cache reuse.
- **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, will 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.

Motivation

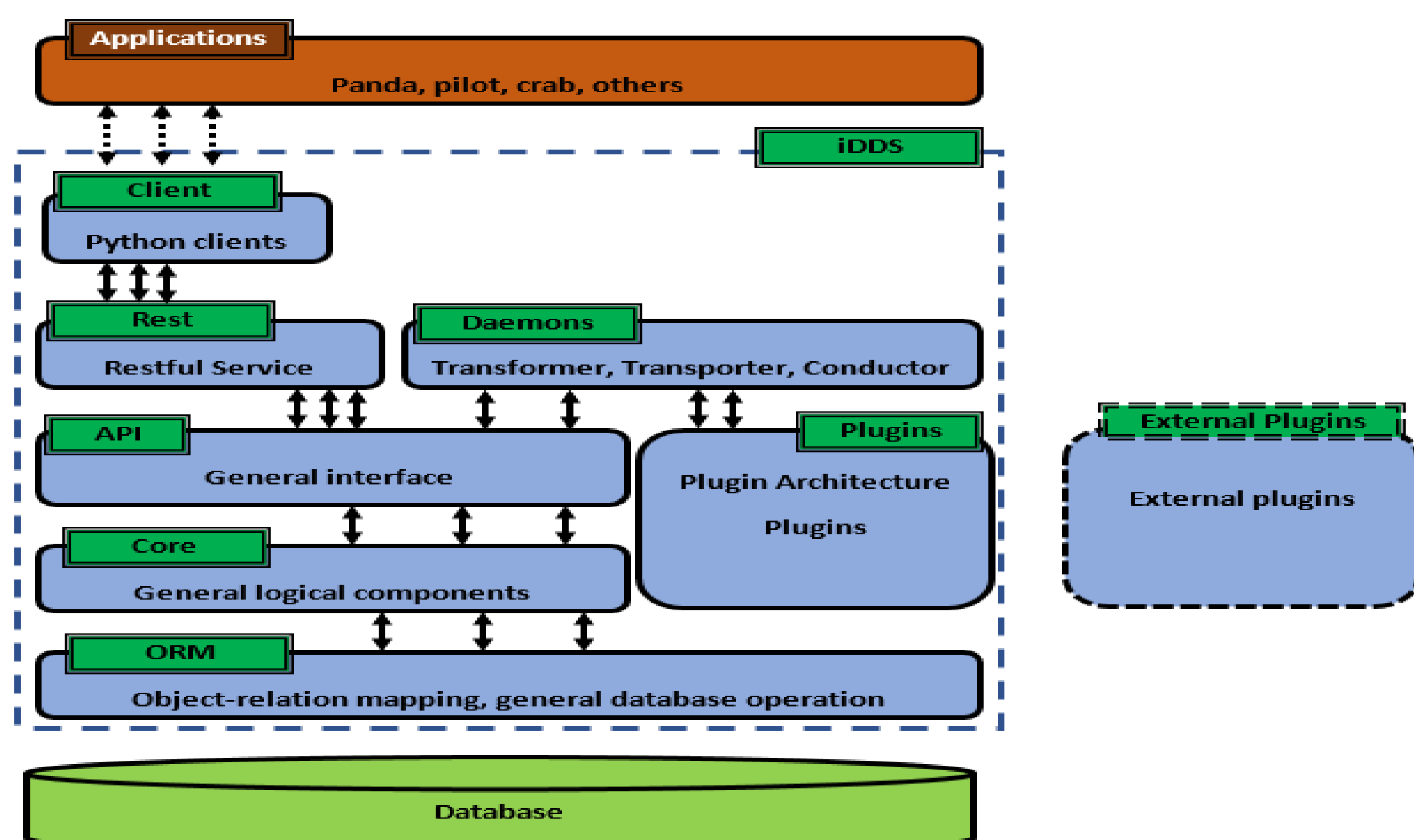
- **Fine-grained delivery:**
 - **ATLAS Event Streaming Service (ESS):** asynchronously deliver only the input events required by ES processing, with the aim to decrease data traffic over WAN, hide WAN latency and improve overall data processing throughput.
 - **ATLAS data carousel:** Without waiting for all data to be staged-in.
- **Transformation on demand:**
 - In current ATLAS computing model, some analysis format such as DAOD are centrally produced and stored for a long time. Some of them may never be used before a new version is produced; some of them are used just a few times in a short period and then removed.
 - The ATLAS Event Streaming Service (ESS) is an approach to preprocess data based on task demands.
- **Orchestration:**
 - Some existing services developed with accumulated requirements are housed in inappropriate components, which increases the difficulty to maintain them and to optimize the workflow.

iDDS Components



- **HEAD: Restful interface**
 - To register and query requests
 - To provide catalog service for consumers to get/list required collections and contents(files or events).
- **Transformer**
 - To transform input data from one collection to another collection
 - With plugin structure to support different transform type and different backends.
- **Transporter**
 - To manage collections with DDM backends
- **Conductor**
 - To notify/schedule consumers to consume new transformed data in a fine-grained granularity
- **Others**
 - **Client:** To communicate with iDDS Restful service.

iDDS Architecture



- **Experiment agnostic**
- **Abstract Layers**
 - Generalization
 - Hide the complexity of different logics to implement general API.
- **Plugin Architecture**
 - Flexible to support new transform types
 - Flexible to support new backends

Status and Plans

- **Current Status**
 - Main components are implemented and deployed.
 - An ATLAS data carousel workflow is also implemented with plugins and integrated to ATLAS production system, integration tests underway.
- **Advanced version (~ by the end of 2020)**
 - The requester defines only workflow and input, while iDDS dynamically optimizes granularity and destination of output collection and triggers the subsequent step by using own decision making engine