

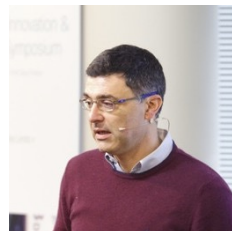
PLAS

Platformed Workflows

Andrea Detti: andrea.detti@uniroma2.it

Luca Petrucci: luca.petrucci@uniroma2.it

Ludovico Funari: ludovico.funari@uniroma2.it



Project supported by GÉANT Innovation Programme (July – December 2021)

GÉANT Cloud Flow



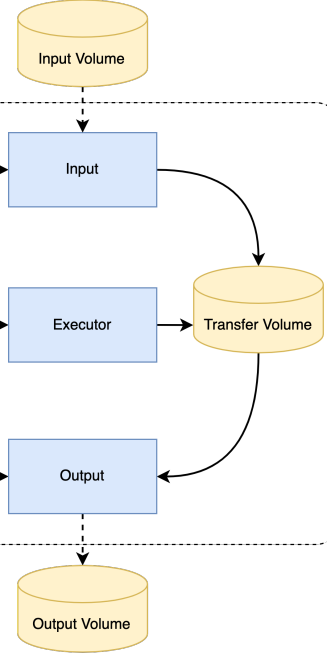
```
1  cwlVersion: v1.0
2  class: CommandLineTool
3  hints:
4  - class: DockerRequirement
5    dockerPull: kubler/openssl
6  inputs:
7  - id: input
8    type: File
9    doc: "original content"
10  inputBinding:
11    position: 1
12  outputs:
13  - id: output
14    type: stdout
15    stdout: sha
16  baseCommand: ["openssl", "dgst"]
17  arguments: ["-sha1"]
```

CWL

cwl-TES
client

JSON
TESK-API
server

taskmaster



- GÉANT Cloud Flow (GCF) provides researchers with the ability to run workflows composed of analysis tasks exploiting GÉANT cloud facilities (e.g. based on Kubernetes)
- Workflows and tasks described with CWL language
- A task is command line tool run in a Linux Container...*containerized-tasks*

GCF possible limitations

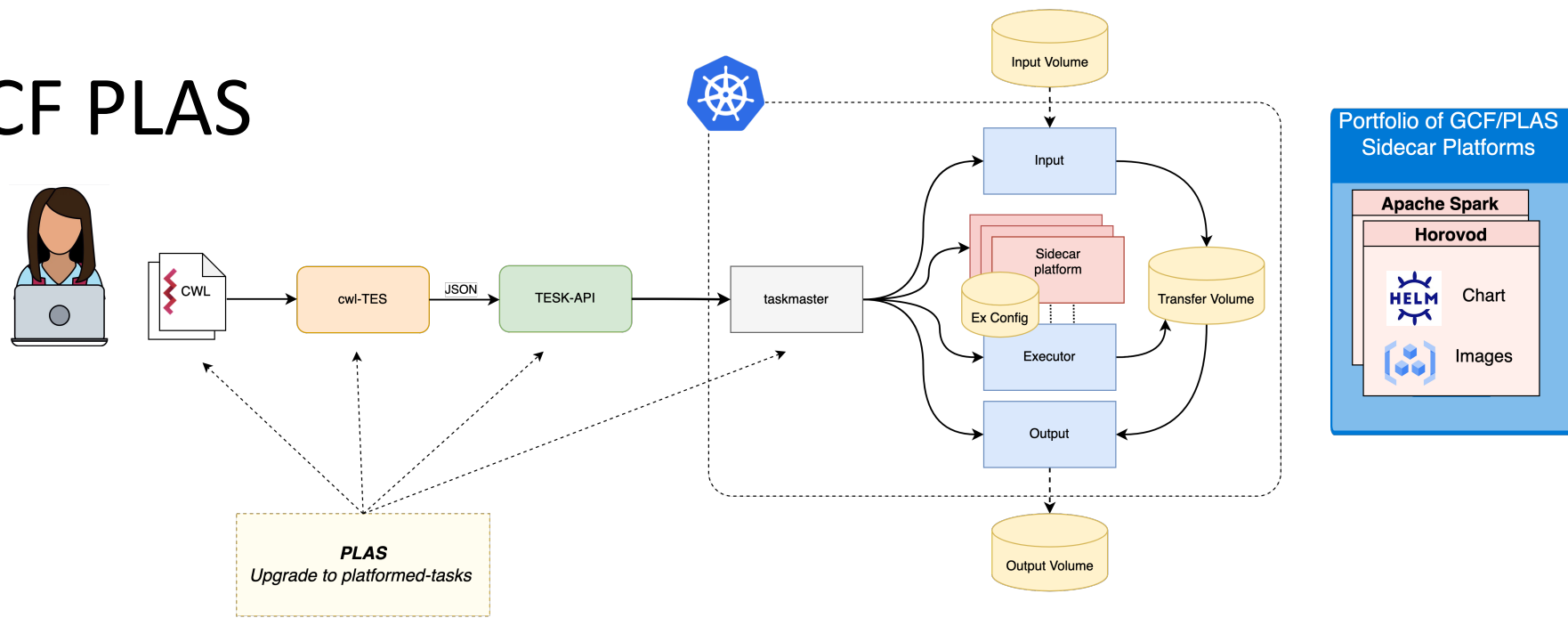
No horizontal scalability

- A task runs in a single Linux Container
- A task can use the resources of a single host of the cloud

Limited comfortability

- A task is a command line tool, like a python file, packaged in a self-consistent Container
- Data analysts are used to using the services offered by analytics platforms such as Apache Spark, Horovod, etc. rather than developing self-consistent code

GCF PLAS



- Extension to execute *platformed-tasks*: tasks that run in a *sidecar platform*
 - **GCF offers platform-as-a service**
- Task's Containers deployed alongside the *sidecar platform*, installed via Helm
- The portfolio of sidecar platforms now includes : Apache Spark and Horovod for accelerating ML training jobs

GCF PLAS at Work

Task master

NAME	READY	STATUS
task-a963dd6a--1-8zhrq	1/1	Running
task-a963dd6a-ex-00--1-57c9w	0/1	Completed
task-a963dd6a-inputs-filer--1-hmmqd	0/1	Completed
task-a963dd6a-outputs-filer--1-phhc6	0/1	Completed
task-a963dd6a-platform-horovod-0	1/1	Terminating
task-a963dd6a-platform-horovod-1	1/1	Terminating

GCF PLAS at Work

Executor, Input, Output

NAME	READY	STATUS
task-a963dd6a--1-8zhrq	1/1	Running
task-a963dd6a-ex-00--1-57c9w	0/1	Completed
task-a963dd6a-inputs-filer--1-hmmqd	0/1	Completed
task-a963dd6a-outputs-filer--1-phhc6	0/1	Completed
task-a963dd6a-platform-horovod-0	1/1	Terminating
task-a963dd6a-platform-horovod-1	1/1	Terminating

GCF PLAS at Work

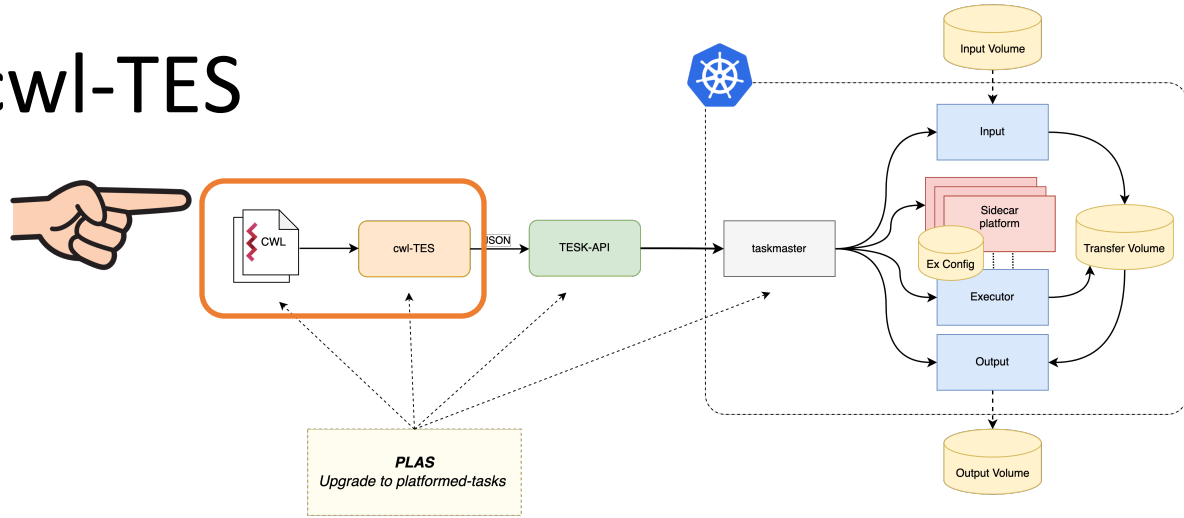
Sidecar platform

NAME	READY	STATUS
task-a963dd6a--1-8zhrq	1/1	Running
task-a963dd6a-ex-00--1-57c9w	0/1	Completed
task-a963dd6a-inputs-filer--1-hmmqd	0/1	Completed
task-a963dd6a-outputs-filer--1-phhc6	0/1	Completed
task-a963dd6a-platform-horovod-0	1/1	Terminating
task-a963dd6a-platform-horovod-1	1/1	Terminating

Extension Challenges

New CWL and cwl-TES

Extend the cwl-tes to add the support to new Common Workflow Language (CWL) **HelmRequirement** class.



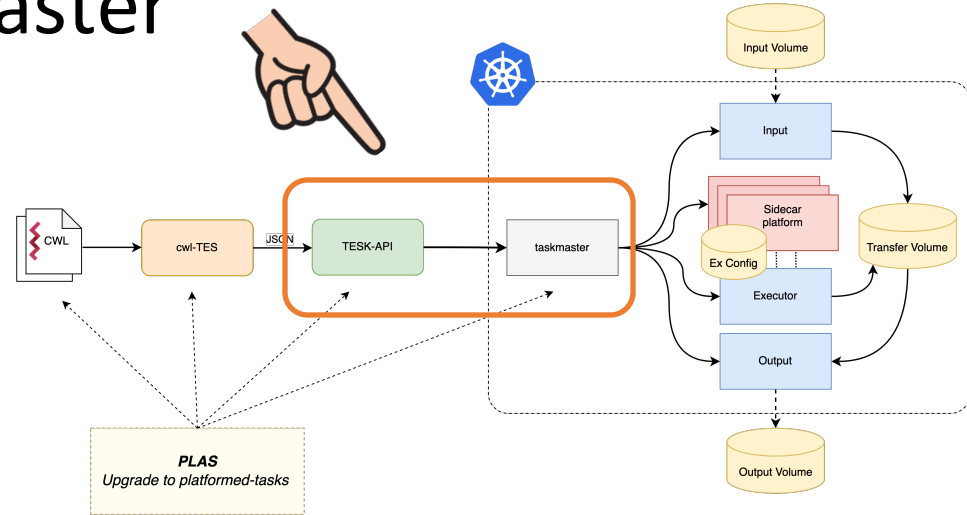
```
1  cwlVersion: v1.0
2  class: CommandLineTool
3  requirements:
4  - class: HelmRequirement
5    chartRepo: "https://link_to_horovod_chart_repo/charts"
6    chartVersion: "3.0.0"
7    chartName: "horovod"
8    executorImage: "horovod:latest"
```

Note: Supported chart repository link and its specific executor image is needed

New TESK and Taskmaster

Extend the TESK-API component to forward changes and generate the taskmaster pod with additional volume and extra necessary details

Improve the capability of the taskmaster to deploy the platform using **Helm**, in addition to the creation of input, output and executor containers.



Charts and executors for sidecar platforms

Charts of platformed workflow (at the moment *horovod* and *spark*) and corresponding Executor Docker Images

The screenshot shows the PlatformedTasks dashboard. At the top, there is a navigation bar with the following items: Overview (selected), Repositories (8), Packages, People (3), Teams, Projects, and Settings. Below the navigation bar, there is a 'Pinned' section with three cards: 'Documentation' (Public), 'PLAS-TESK' (Public), and 'PLAS-charts' (Public). The 'PLAS-charts' card has a 'Mustache' icon. Below the pinned section, there is a 'Repositories' section with a search bar and filters for Type, Language, and Sort. A 'New' button is also present. The main content area displays a list of repositories, each with a title, a language icon, a star icon, a license icon, a version icon, a build icon, and an update date. Each repository also has a green line chart on the right side. The repositories listed are: Documentation (Public), PLAS-horovod (Public), PLAS-spark (Public), PLAS-charts (Public), PLAS-cwl-tes (Public), PLAS-TESK (Public), PLAS-tesk-api (Public), and PLAS-tesk-core (Public). Three blue arrows point from the text on the left to the 'PLAS-horovod', 'PLAS-spark', and 'PLAS-charts' repository entries.

Repository Name	Language	Stars	License	Version	Builds	Issues	Updated
Documentation	Public	0	Apache-2.0	0	0	0	Updated on 31 Jan
PLAS-horovod	Python	0	Apache-2.0	0	0	0	Updated on 31 Jan
PLAS-spark	Python	0	Apache-2.0	0	0	0	Updated on 31 Jan
PLAS-charts	Mustache	0	Apache-2.0	0	0	0	Updated on 31 Jan
PLAS-cwl-tes	Python	0	Apache-2.0	0	0	0	Updated on 31 Jan
PLAS-TESK	Public	0	Apache-2.0	0	0	0	Updated on 27 Jan
PLAS-tesk-api	Java	0	Apache-2.0	0	0	0	Updated on 26 Jan
PLAS-tesk-core	Python	0	Apache-2.0	0	0	1	Updated on 26 Jan

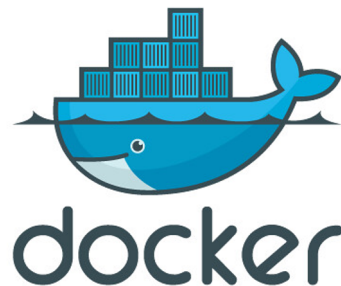
<https://github.com/PlatformedTasks>

We have a GitHub repository for each module:

- cwl-TES
- TESK-API
- Taskmaster
- Platform helm charts:
 - Horovod
 - Spark

We also have developed the Docker Images of:

- TESK-API
- Taskmaster
- Horovod Executor
- Spark Executor



Conclusion and Future works

GCF can now offers platform-as-a-service to speed up computation and make researcher more comfortable with their instruments



Future works

Increase platform portfolio
e.g. with the HPC platforms

Support QoS

Security assessment

DEMO TIME