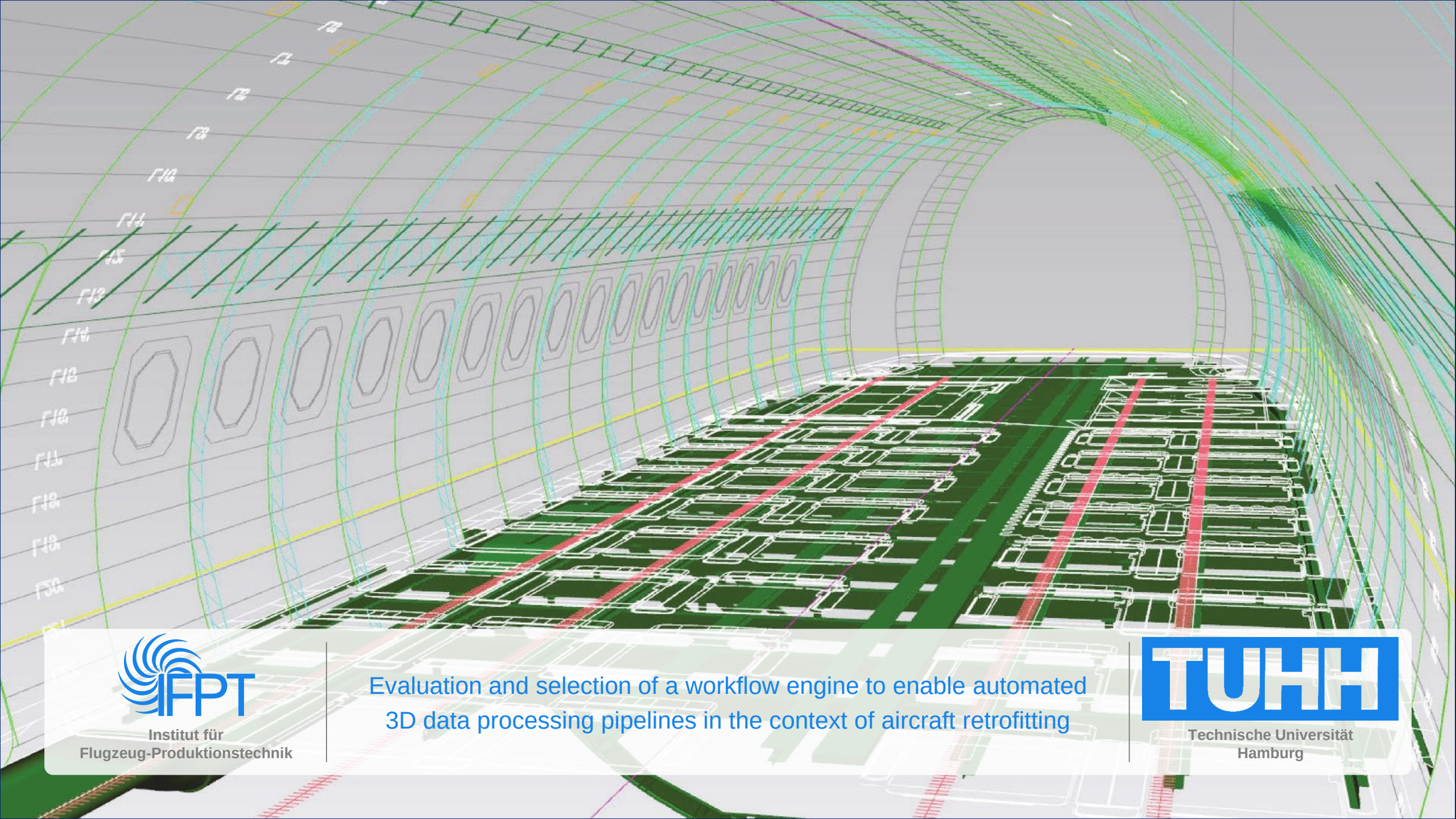
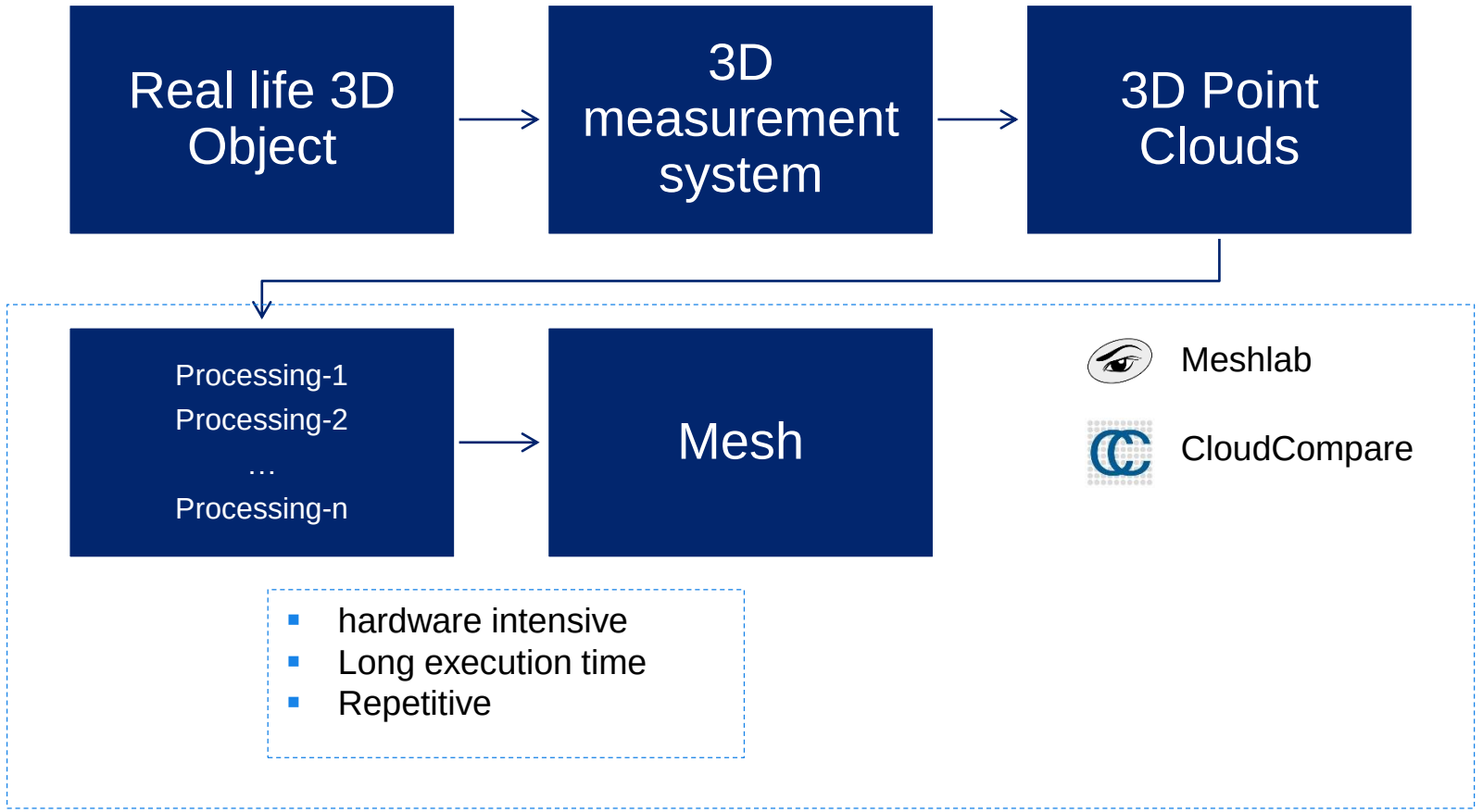


Evaluation and selection of a workflow engine to enable automated 3D data processing pipelines in the context of aircraft retrofitting

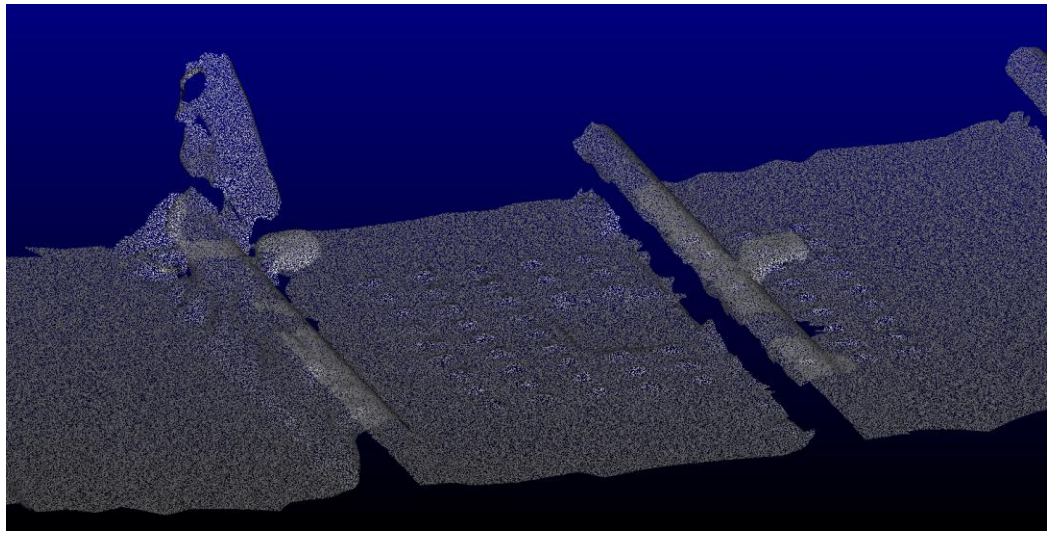
Institut für Flugzeug-Produktionstechnik

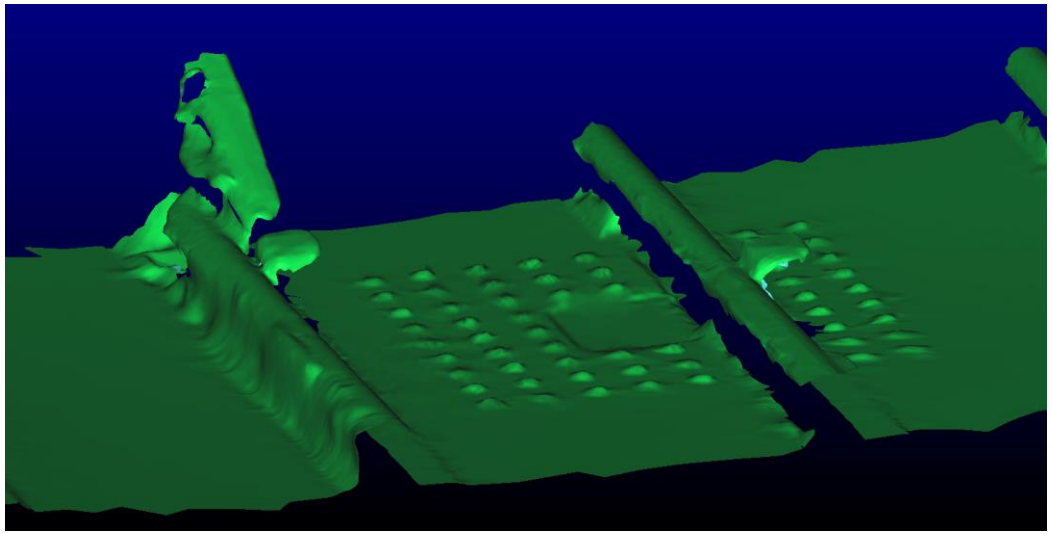
Technische Universität Hamburg


Real life 3D Object
3D measurement system
3D Point Clouds
Processing-1
Processing-2
…
Processing-n
Mesh
Meshlab
CloudCompare
hardware intensive
Long execution time
Repetitive

Artifacts generated by a 3D data processing pipeline

[1]





3D Data Processing using Workflow Engine

[2]

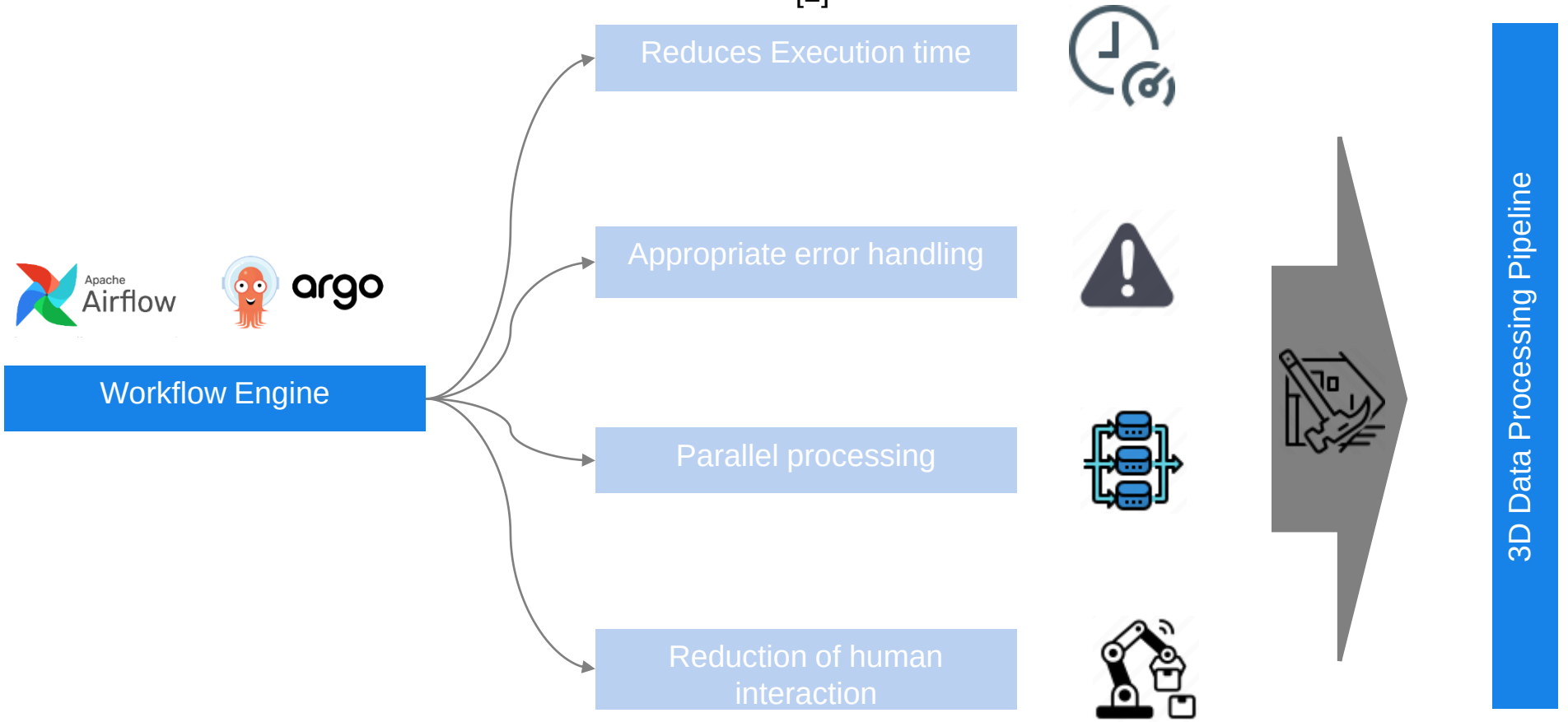


Table of Contents

Core Project Tasks

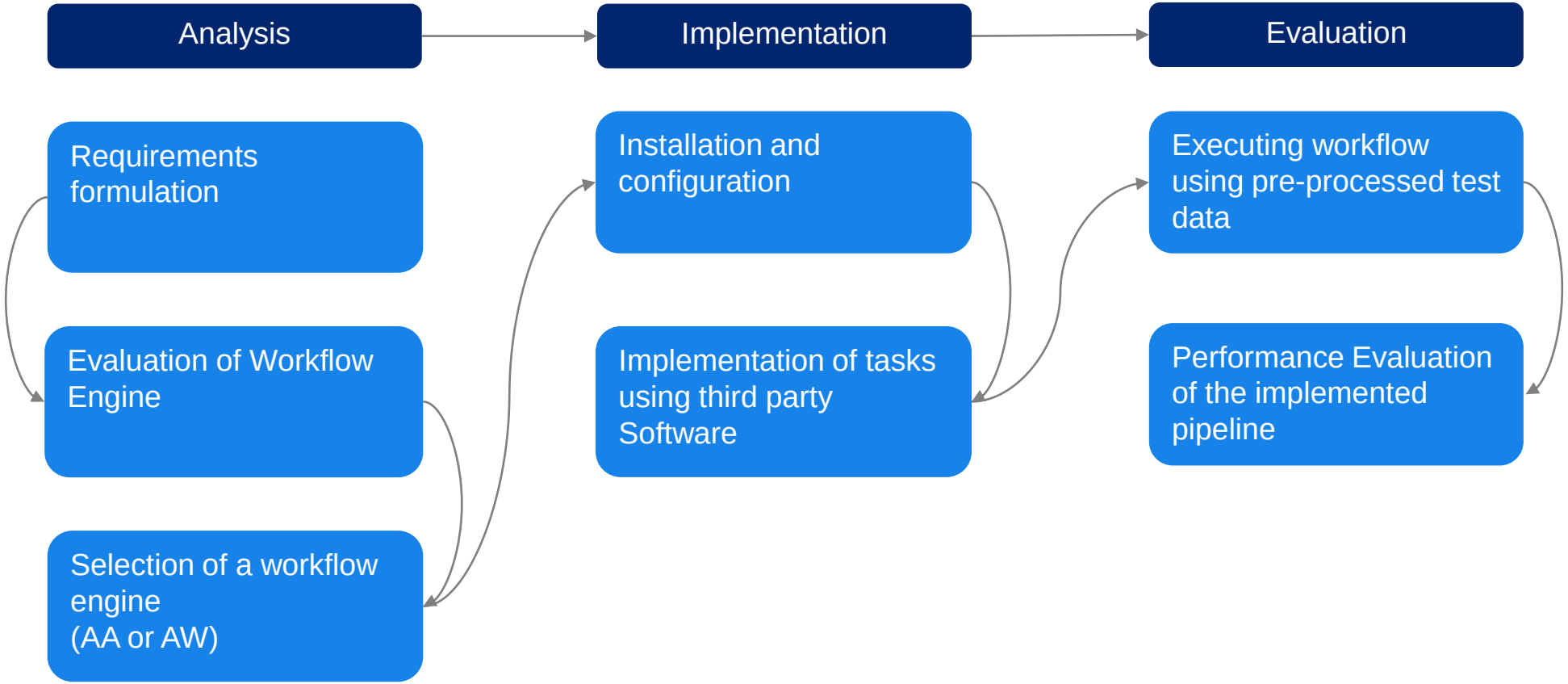
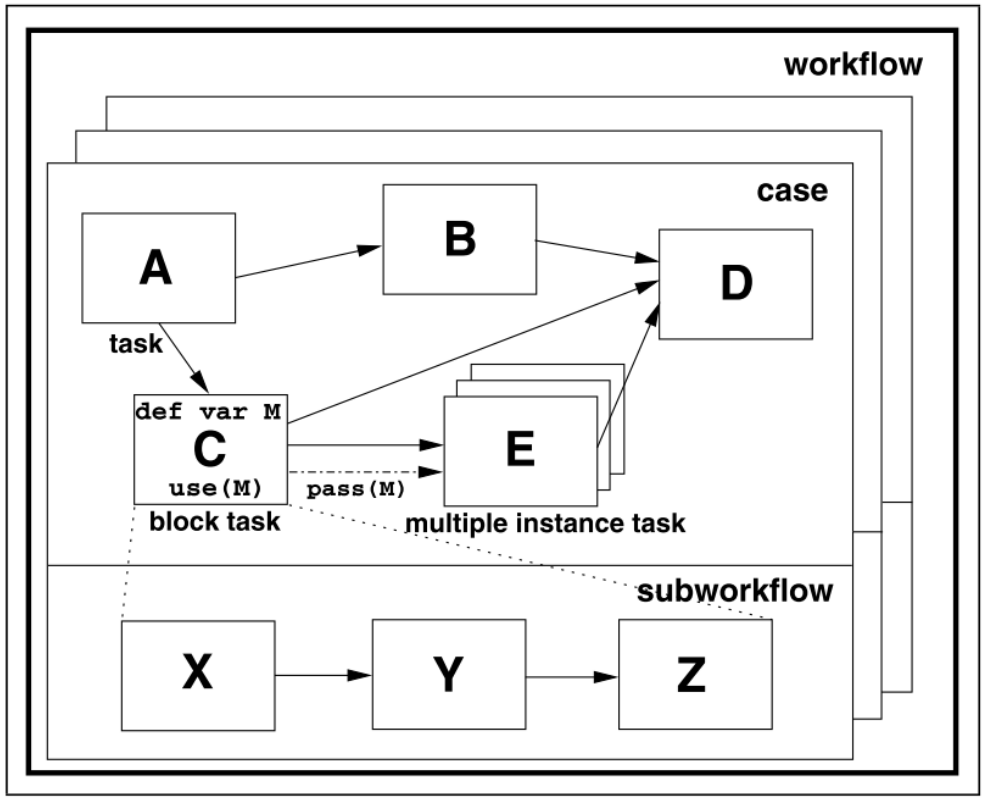


Table of Contents

So what is a Workflow Engine?

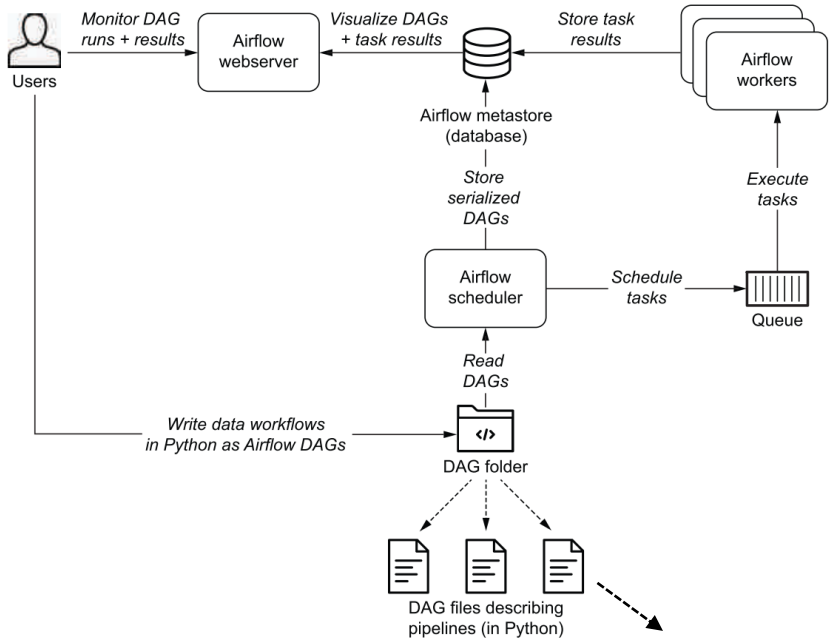
- Software systems that facilitate the automation of a workflow
- A workflow consists of several processes
- Defines Workflows in DAGs



[3]

[4]

Apache airflow

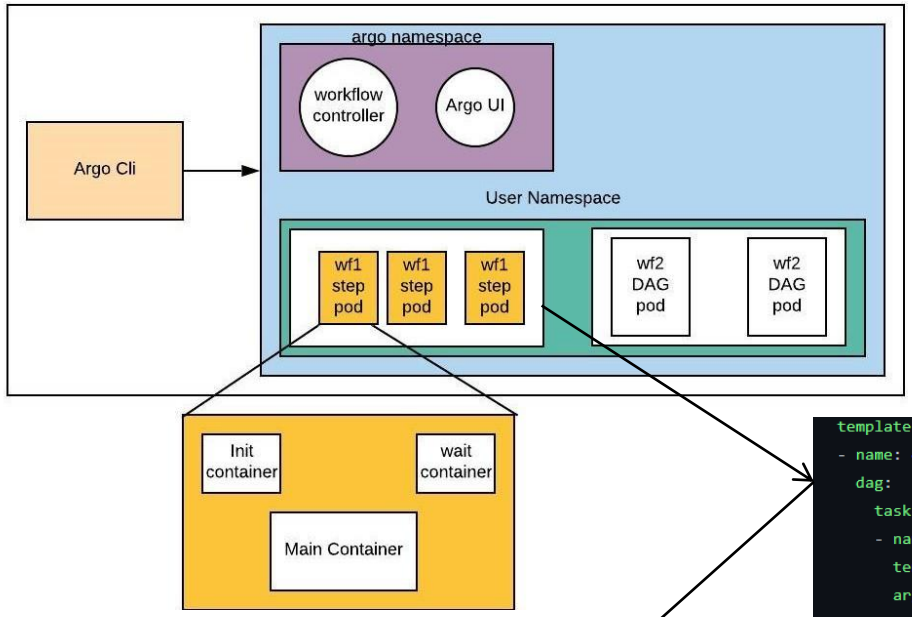



```python
t1 = BashOperator(
    task_id='print_date',
    bash_command='date',
)

t2 = BashOperator(
    task_id='sleep',
    depends_on_past=False,
    bash_command='sleep 5',
    retries=3,
)
```

Argo Workflow

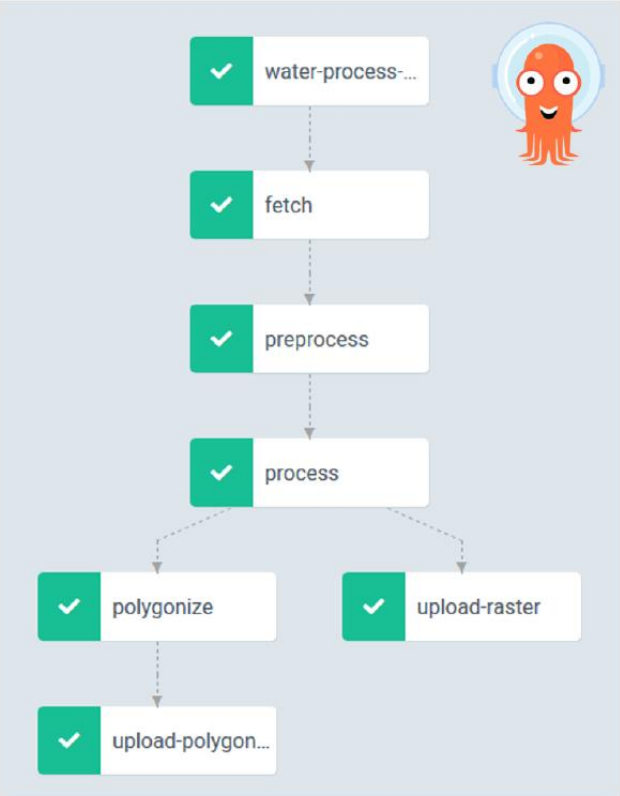
[5]



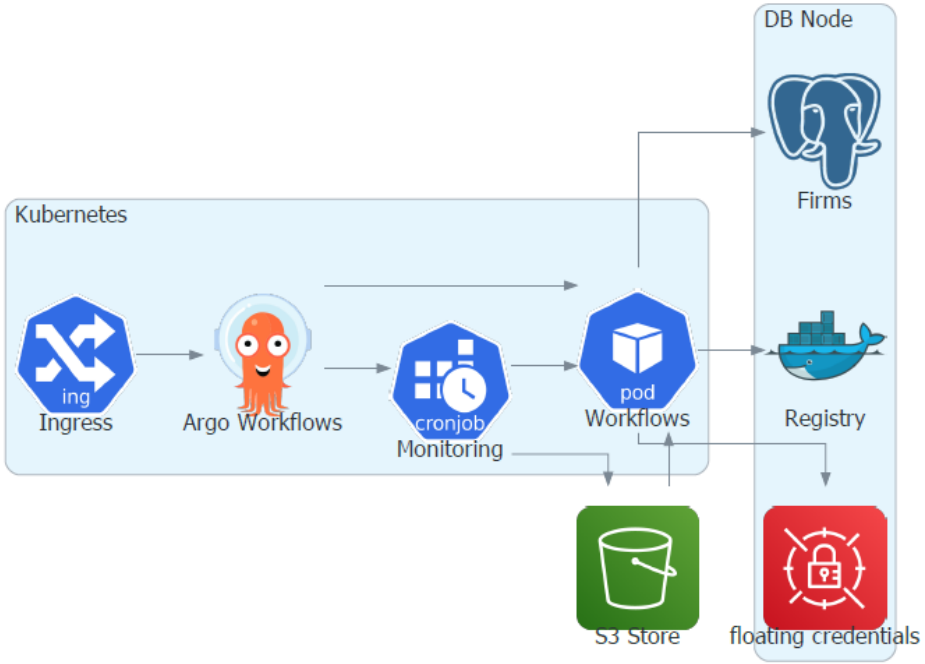

```yaml
- name: echo
  inputs:
    parameters:
    - name: message
  container:
    image: alpine:3.7
    command: [echo, "{{inputs.parameters.message}}"]
```

```yaml
templates:
- name: diamond
  dag:
    tasks:
    - name: A
      template: echo
      arguments:
        parameters: [{name: message, value: A}]
    - name: B
      depends: "A"
      template: echo
      arguments:
        parameters: [{name: message, value: B}]
    - name: C
      depends: "A"
      template: echo
      arguments:
        parameters: [{name: message, value: C}]
    - name: D
      depends: "B && C"
      template: echo
      arguments:
        parameters: [{name: message, value: D}]
```

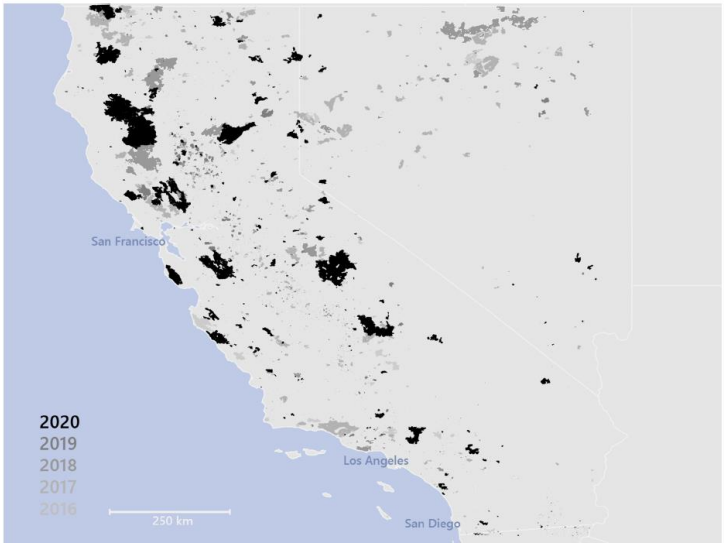
Scalable Processing Of Copernicus Sentinel satellite Images Using Argo Workflows [6]



DAG Structure



Processing Architecture



Results

Table of Contents

[7]

Functional Requirements

- Loading and saving 3D data object
- Containerization of tasks
- Artifact storage system
- Visual workflow editor
- Reusable task template
- Deployment in kubernetes

Non-Functional Requirements

- Notification system
- Detailed logs of the pipeline execution
- Version controlled DAG/Task editing
- Appropriate error handling
- Appropriate UI usability for non-expert

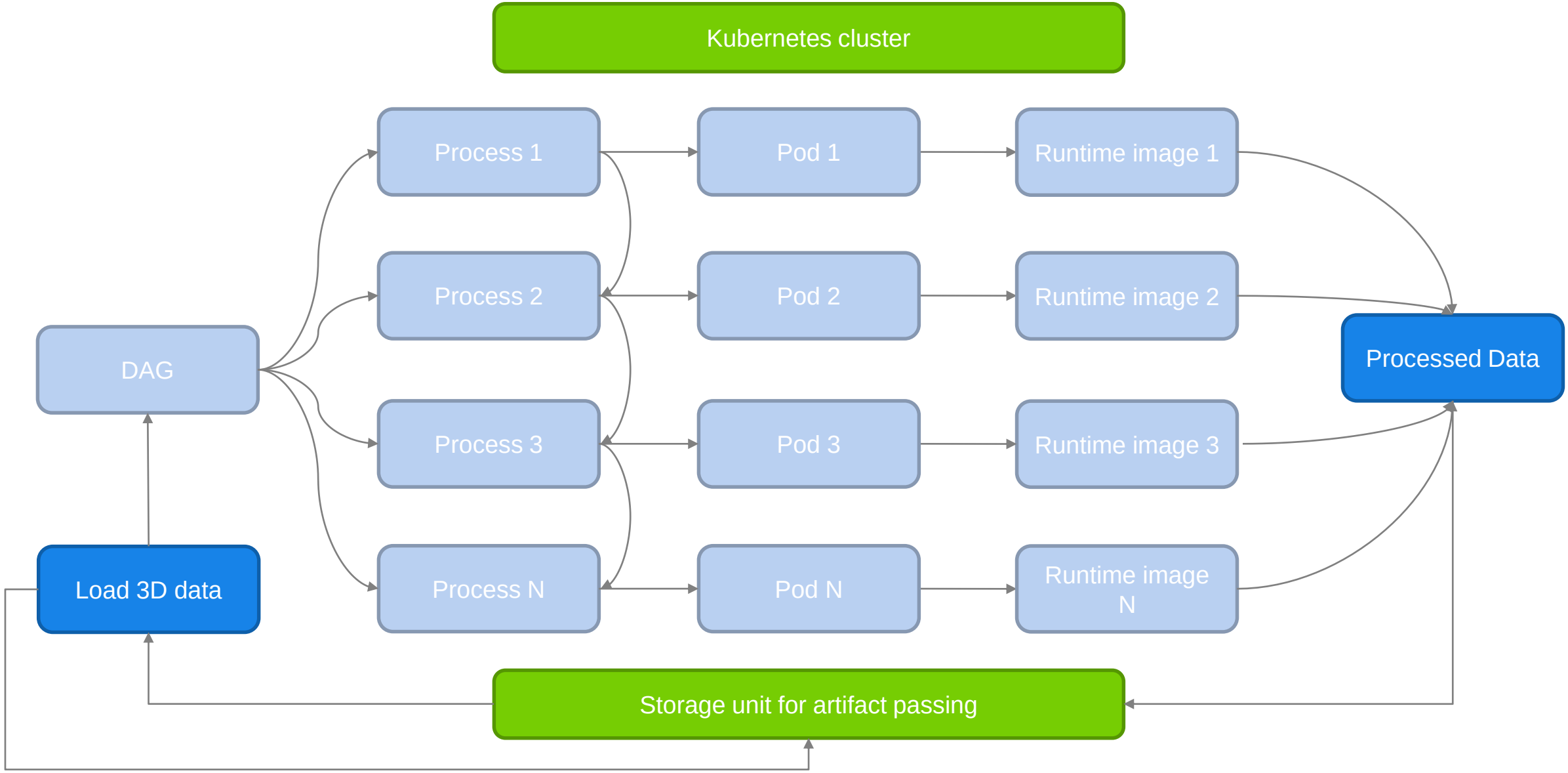
Table of Contents

Average Weight-score analysis

	Apache Airflow	**Argo Workflow**
Functional Requirements	3.4	4.2
Non-Functional Requirements	4.1	4.12

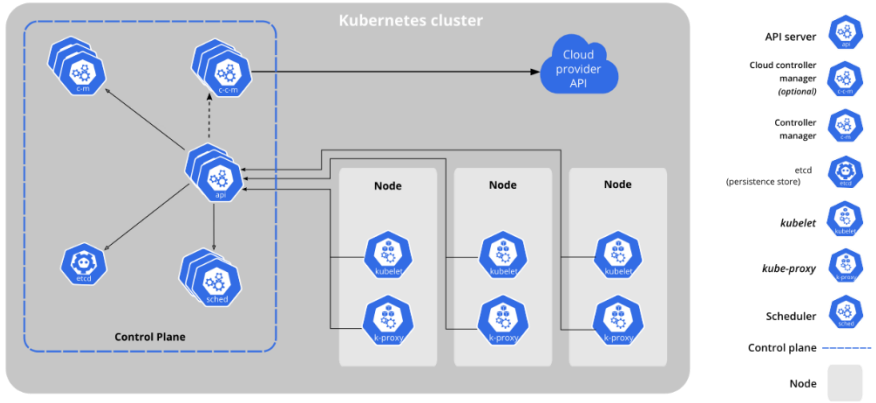
Table of Contents

Conceptual Pipeline

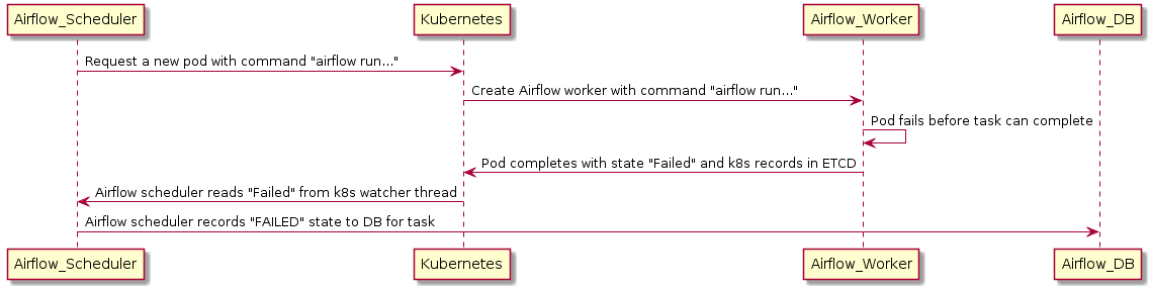


[8, 9, 10, 11]

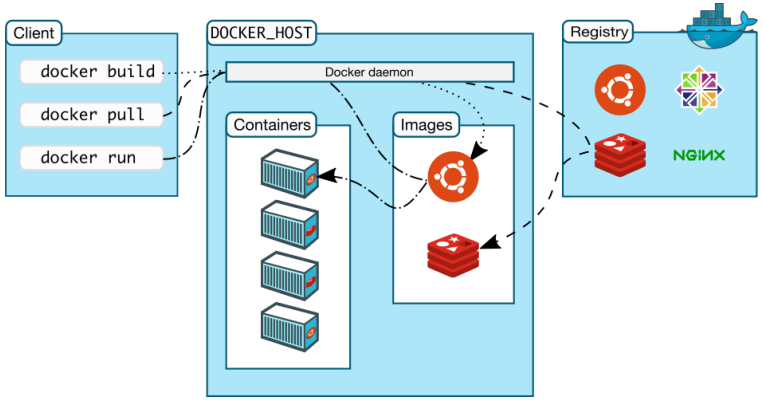
Kubernetes Cluster



Fault Handling



Containerization of tasks



Artifact handling/Storage unit (s3)

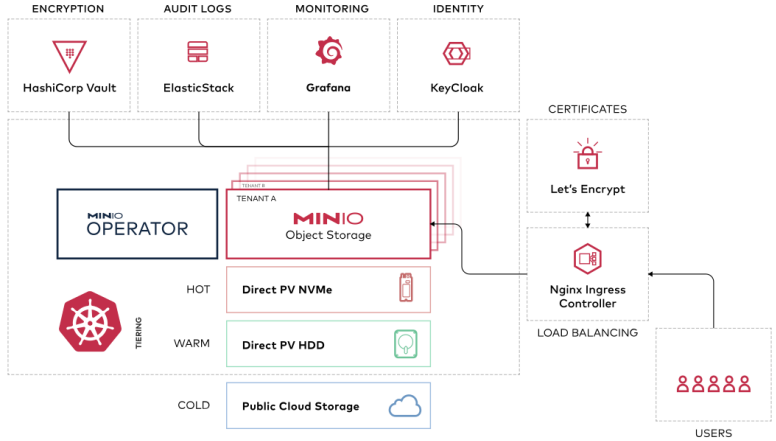
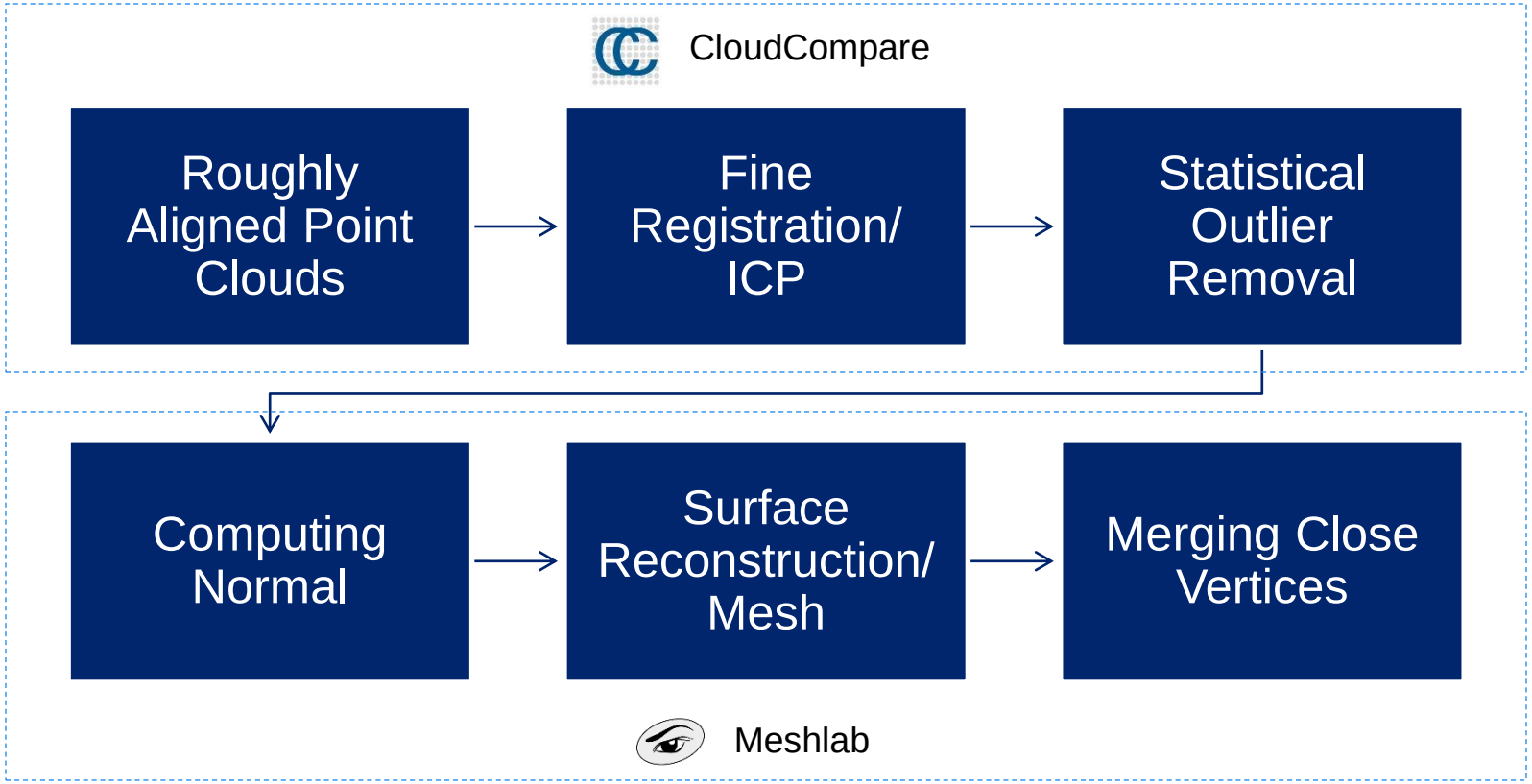
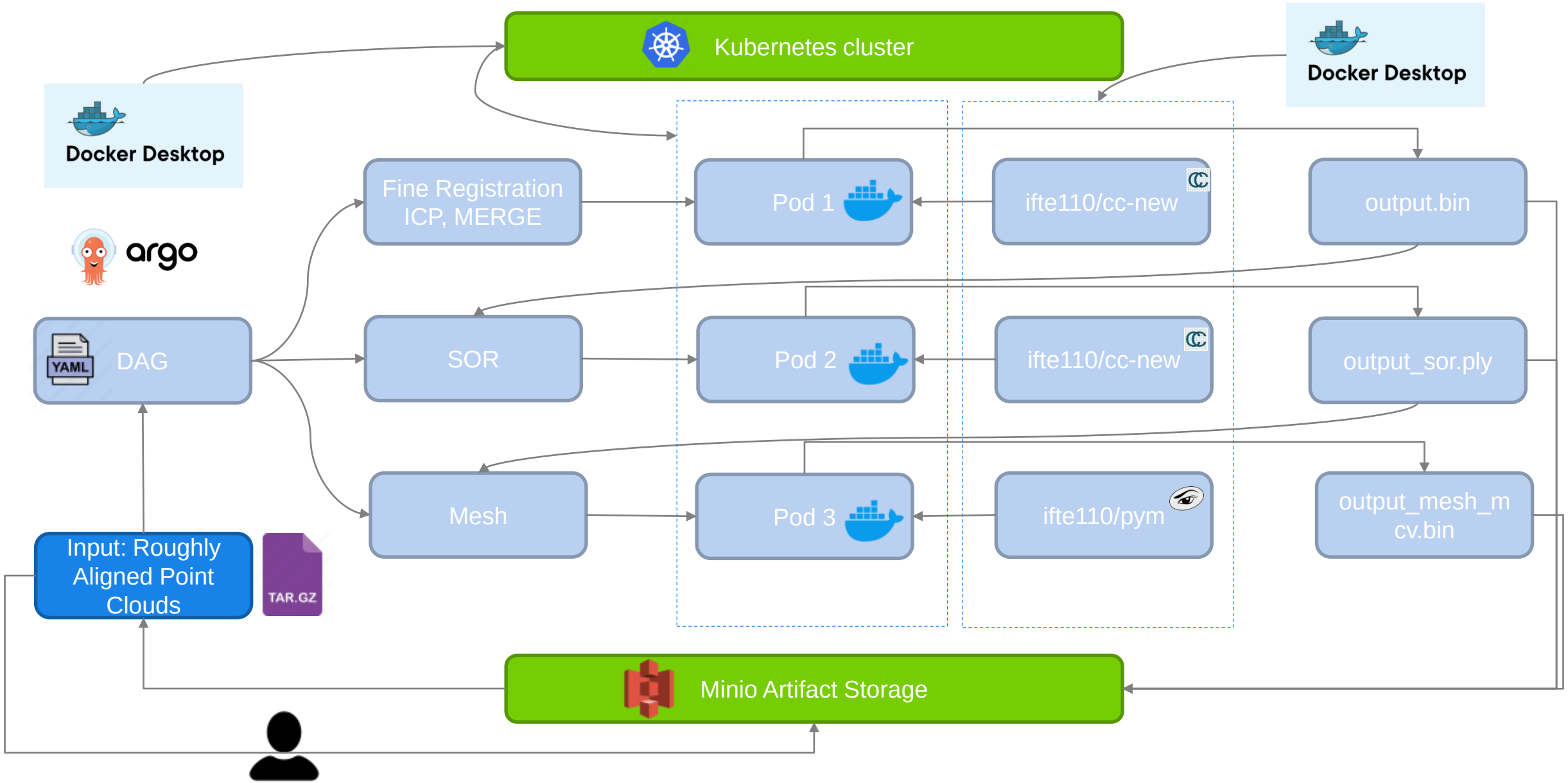


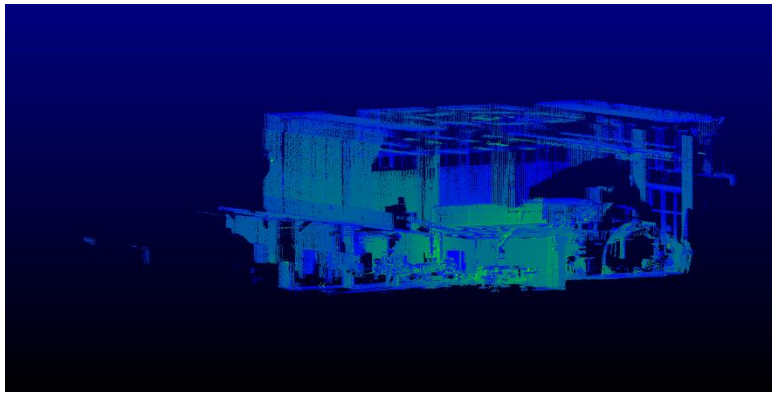
Table of Contents

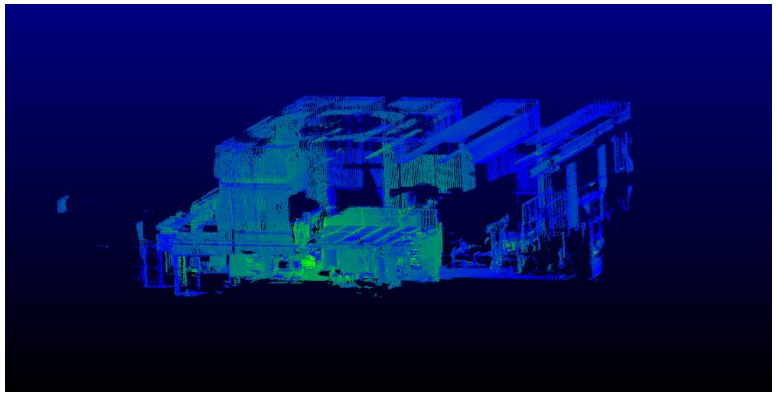

CloudCompare
Roughly Aligned Point Clouds
Fine Registration/ ICP
Statistical Outlier Removal
Computing Normal
Surface Reconstruction/ Mesh
Merging Close Vertices
Meshlab

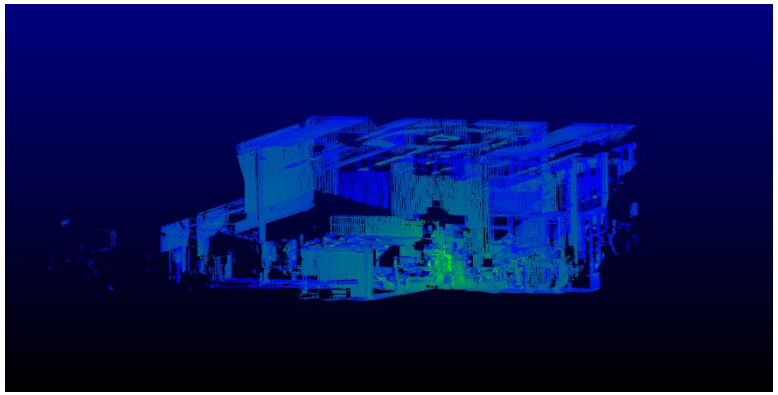
The automated pipeline



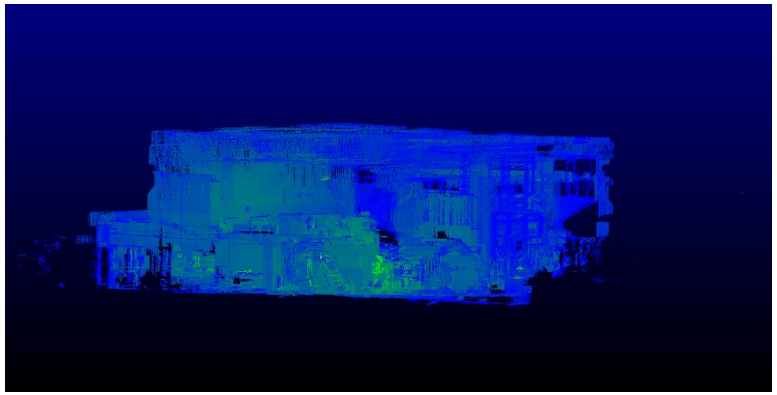
Input files (Roughly Aligned Point Clouds)

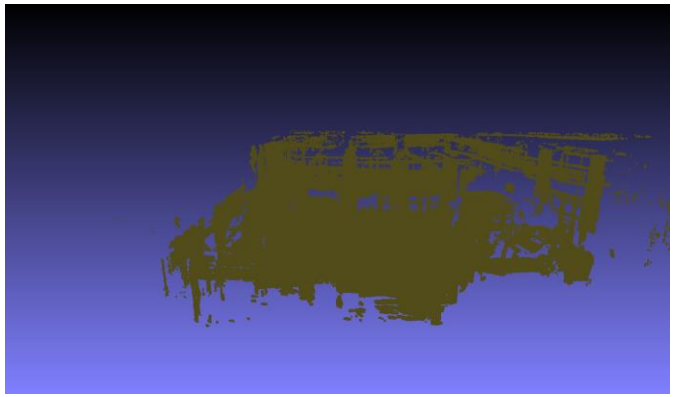


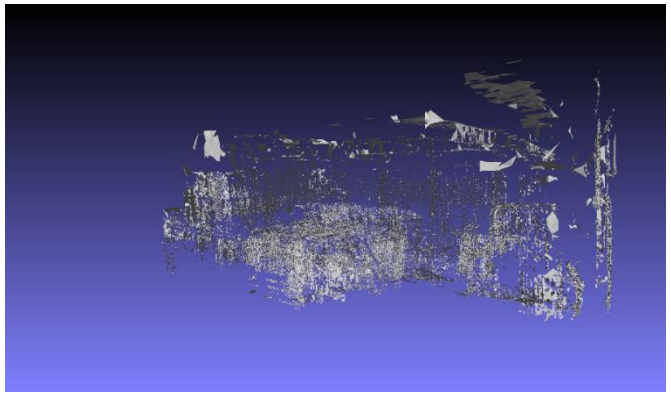


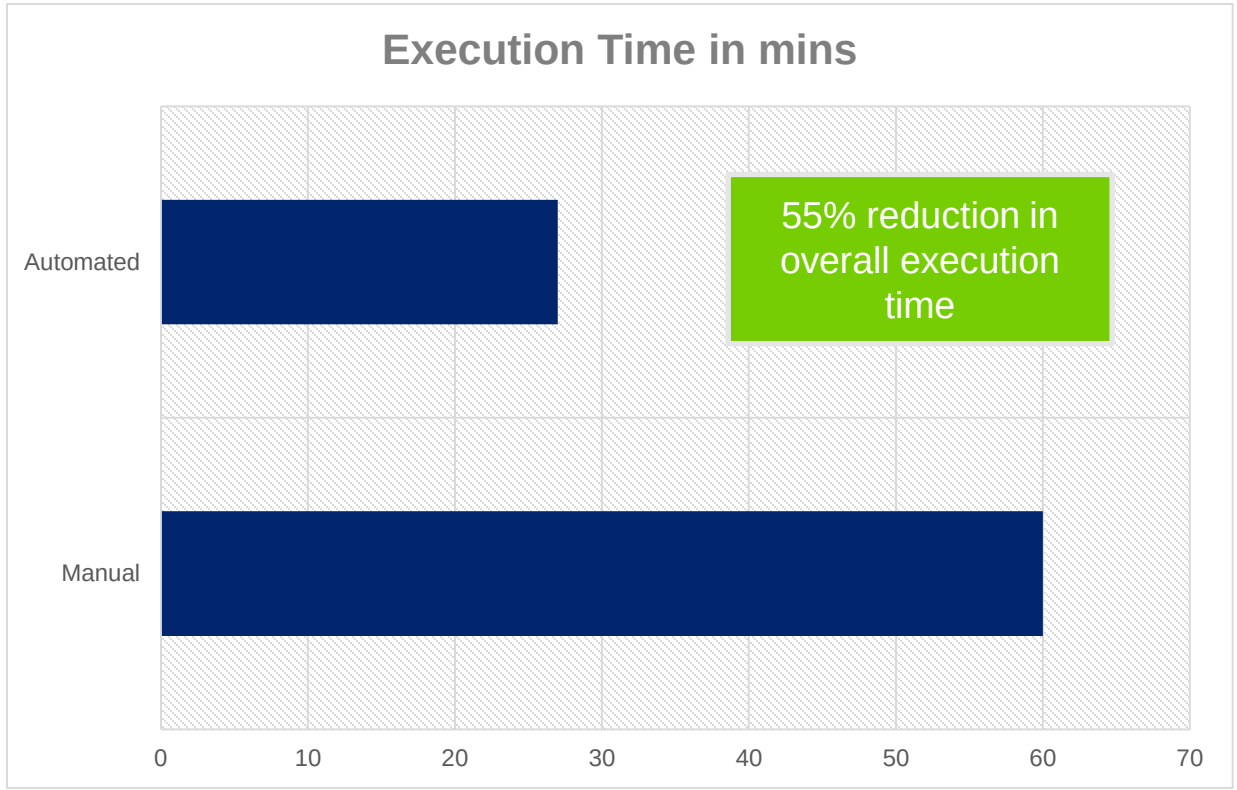


Output files









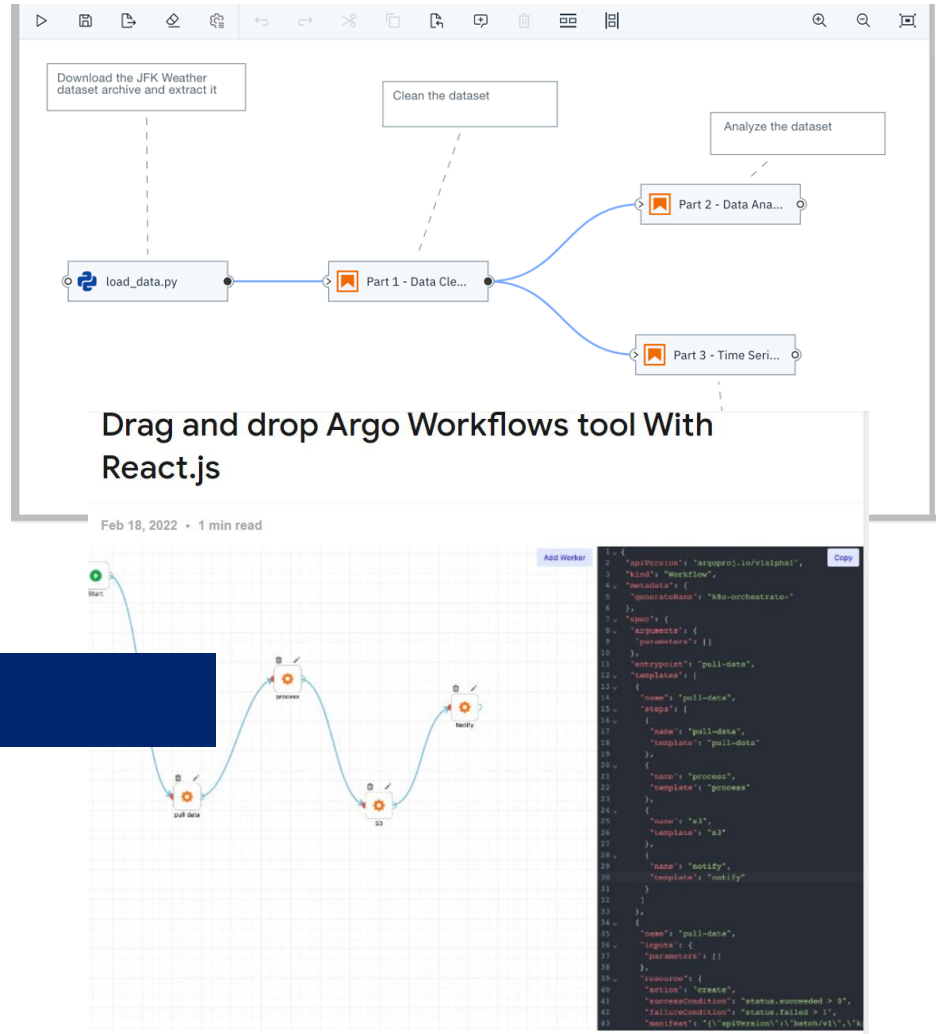
- Reduces Execution time
- Appropriate error handling
- Parallel processing
- Reduction of human interaction
- Third Party Software Integration
- Easy to use reusable task template

Table of Contents

Level of Automation

- Ease of use
- Overall Exe Time
- Parallel Processing and Error Handling

Visual Workflow Editor [12, 13]



More computing power

TUHH

References

[1] Institut für Flugzeug-Produktionstechnik

[2] Hapke, Hannes (2020): Building Machine Learning Pipelines : Automating Model Life Cycles with TensorFlow. Automating model life cycles with TensorFlow. 1st edition. Beijing, Boston, Farnham: O'Reilly Media, Incorporated.

[3] Nick Russell et al. “Workflow Data Patterns: Identification, Representation and Tool Support”. In: Springer, Berlin, Heidelberg, 2005, pp. 353–368. DOI: 10.1007/11568322{_}23. URL: https://link.springer.com/chapter/10.1007/11568322_23.

[4] Bas Harenslak and Julian de Ruiter. Data pipelines with Apache Airflow. Shelter Island, NY: Manning Publications Co, 2021. ISBN: 9781617296901.

[5] Argo Workflows - The workflow engine for Kubernetes. 5/26/2022. URL: https://argoproj.github.io/argo-workflows/.

[6] Florian Fichtner, Nico Mandery, Maximilian Schwinger, Jonas Eberle, Michael Nolde, Torsten Riedlinger(2021): Scalable Processing Of Copernicus Sentinel satellite Images Using Argo Workflows. In: Proceedings of the 2021 conference on Big Data from Space, 18-20 May 2021, S. 77–80. DOI: 10.2760/125905.

[7] Aurum, Aybüke; Wohlin, Claes (2005): Engineering and managing software requirements. Berlin: Springer.

[8] Kubernetes. Kubernetes Documentation. 5/29/2022. URL: https://kubernetes.io/docs/home/.

[9] Docker Documentation. Docker overview. 5/26/2022. URL: https://docs.docker.com/get-started/overview/.

[10] MinIO, Inc. MinIO | MinIO for Kubernetes. 5/27/2022. URL: https://min.io/product/kubernetes.

[11] Apache Airflow Documentation — Airflow Documentation. 5/25/2022. URL: https://airflow.apache.org/docs/apache-airflow/stable/.

[12] Elyra Documentation — Elyra 3.8.1 documentation. 5/4/2022. URL: https://elyra.readthedocs.io/en/stable/.

[13] React. “Drag and drop Argo Workflows tool With React.js”. In: React.js Examples (2/18/2022). URL: https://reactjsexample.com/drag-and-drop-argo-workflows-tool-with-react-js/.