

INCEPTION

OR

A very brief introduction to Argo Workflows on
Kubernetes

I'm Illya

Senior Geospatial Data Engineer

FATMAP 

by **STRAVA**

- Elevation Data

- Elevation Data
- Winter and Summer Satellite Imagery

- Elevation Data
- Winter and Summer Satellite Imagery
- Vector Map Data - roads, paths, pistes, lifts

- Elevation Data
- Winter and Summer Satellite Imagery
- Vector Map Data - roads, paths, pistes, lifts
- Routes, POIs

DATA PIPELINES



ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

- Define workflows where each step in the workflow is a container

ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

- Define workflows where each step in the workflow is a container
- Model multi-step workflows or capture dependencies between tasks using a directed acyclic graph (DAG)

ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

- Define workflows where each step in the workflow is a container
- Model multi-step workflows or capture dependencies between tasks using a directed acyclic graph (DAG)
- Flexible enough to use for a vast range of processing jobs

ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

- Define workflows where each step in the workflow is a container
- Model multi-step workflows or capture dependencies between tasks using a directed acyclic graph (DAG)
- Flexible enough to use for a vast range of processing jobs
- Implemented as a Kubernetes CRD (Custom Resource Definition)

ORCHESTRATING PARALLEL JOBS ON KUBERNETES.

- Define workflows where each step in the workflow is a container
- Model multi-step workflows or capture dependencies between tasks using a directed acyclic graph (DAG)
- Flexible enough to use for a vast range of processing jobs
- Implemented as a Kubernetes CRD (Custom Resource Definition)
- Define workflows and templates as YAML or Python (Hera)

**SO WHAT DOES A WORKFLOW LOOK LIKE IN
CODE?**

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20 templates:
21   - name: loop-rasters
22     parallelism: 50
23     inputs:
24       parameters:
25       - name: rasters
26     steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20  templates:
21    - name: loop-rasters
22      parallelism: 50
23      inputs:
24        parameters:
25        - name: rasters
26      steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```



```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20 templates:
21   - name: loop-rasters
22     parallelism: 50
23     inputs:
24       parameters:
25       - name: rasters
26     steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20 templates:
21   - name: loop-rasters
22     parallelism: 50
23     inputs:
24       parameters:
25       - name: rasters
26     steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20 templates:
21   - name: loop-rasters
22     parallelism: 50
23     inputs:
24       parameters:
25       - name: rasters
26     steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   generateName: "TantalusRange-DigitalGlobe-50cm-08062011-w-"
5 spec:
6   securityContext:
7     runAsUser: 0
8   entrypoint: loop-rasters
9   arguments:
10    parameters:
11    - name: log-level
12      value: debug
13    - name: catalogue-dsn
14      value: "postgres://some-postgres-connection/catalogue"
15    - name: pipeline-bucket
16      value: "a-bucket-name"
17    - name: rasters
18      value: '[{"dataset-id":"TantalusRange-DigitalGlobe-50cm-08062011-w","raster-path":"s3://coremapdata/raw/imag
19
20  templates:
21    - name: loop-rasters
22      parallelism: 50
23      inputs:
24        parameters:
25        - name: rasters
26      steps:
```

```
27     -- name: process-rasters
28     template: process-raster
29     arguments:
30     parameters:
```



```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```



```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet/url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet/url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```



```
1 apiVersion: argoproj.io/v1alpha1
2 kind: Workflow
3 metadata:
4   name: osm-import-replicate
5   generateName: osm-download-import-replicate-
6 spec:
7   entrypoint: main-workflow
8   onExit: workflow-exit-handler
9   securityContext:
10     runAsUser: 0
11   volumes:
12     - name: workdir
13       persistentVolumeClaim:
14         claimName: osm-file-cache
15   arguments:
16     parameters:
17       - name: osm-url
18         value: https://planet.url/planet-latest.osm.pbf
19       - name: replication-server
20         value: https://replication/day/
21       - name: pbf-file
22         value: /data/osm_download/planet-latest.osm.pbf
23       - name: imposm-config
24         value: /import_configs/planet_config.json
25       - name: mapping-file
26         value: /import_mappings/poi_tiles_mapping.yml
```

```
27 templates:
28 # DAG
29 - name: main-workflow
30   dag:
```

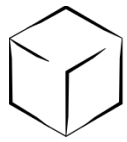

- OSM Planet Imports and Updates

- Satellite Imagery Processing e.g. Pan-sharpening, Orthorectification, Color Balancing

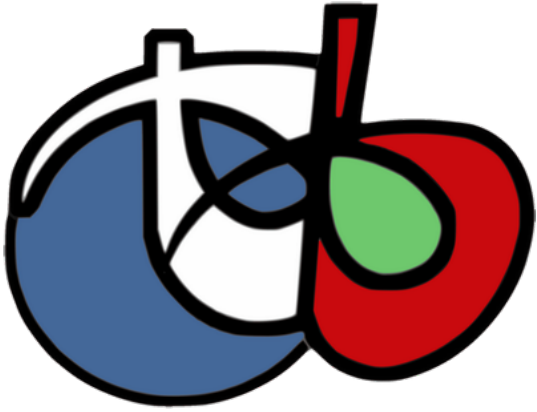
- Parallelised Tile generation - imagery and terrain

- Database syncs and maintenance operations

- Metadata ingestion and cataloging



WhiteboxTools™









SO WHY...

INCEPTION

THANKS!

QUESTIONS