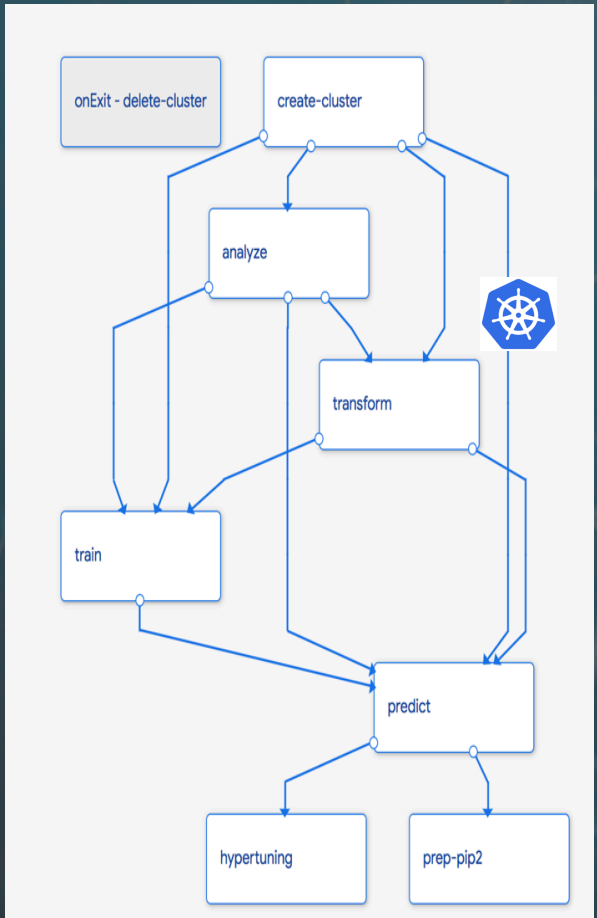
Ness Digital Engineering

NextGenML Platform



Revolutionary container based hybrid cloud solution for ML

Radu Moldovan
Europe Lead Big Data & Machine Learning

Making ML simple portable and scalable

Ness Machine Learning Platform (MLP)

- Robust pipeline container oriented to:
  - ✓ to **train a model**
  - ✓ provide **metrics** and generic **evaluation**
- **Measure** the **performance** of the trained model
- Deploy the **model as a service**

… of course all integrated into a **Kubernetes** container solution

Ness Machine Learning Platform (MLP)

Ness MLP act as a hub around which all data science work takes place at enterprise scale level.

Ness' data science platform puts the entire data modelling process in the hands of data science teams so they can focus on gaining insights from data and communicating them to key stakeholders in your business.

• Infrastructure Agnostic (Docker)

Not vendor locked, entire technology stack is opensource

•was deployed **on prem** but also **on cloud** (Azure Kubernetes Service)

•still easy to **integrate with cloud services** (Databricks – Spark)

•used **cutting edge devOps technologies** in container area (Dockerize, K8s)

• Azure **heterogenous cluster (CPU&GPU)** for dynamic GPU allocation for ML trainig

In client's production we run in parallel more than 20 ML models for organ-segmentation, each one took around 20-24 hours. The GPU nodes were allocated on the fly minimaxing the time and costs.

• ML pipeline (Kubeflow)

Build, deploy and manage multi-step ML workflows based on Docker containers

- Pipeline creation is done programmatically using **Python DSL (low level control)**
- Real time perception of **progress, execution, logs, in/outs of parameters**
- **Unified all the logs** from pipeline containers and also system containers so it's easy for all the users(data engineers/scientist) to check progress
- Easy **tool to compare** in parallel different parameters and hyper-parameters for scientific experiments.
- Also keep track of entire **history of all runs and re-runs**

• Asset Catalog

Build as ML/AI collaboration (Records, Models, Experiments, Pipelines, Deployments)

- built by Ness us from scratch as **single source of truth** built around ML concepts
- is the centric part for data scientist, **the reference data integration** also for data engineers
- **data traceability** UI/UX designed to easily drill down the data and ML execution
- **REST API and SDK(Python)** to manage records, experiment, pipelines during ML execution
- “experiment metadata” provided through UI => no need to ssh the machine for info

• Asset Catalog

Build as ML/AI collaboration (Records, Models, Experiments, Pipelines, Deployments)

- built by Ness us from scratch as **single source of truth** built around ML concepts
- is the centric part for data scientist, **the reference data integration** also for data engineers
- **data traceability** UI/UX designed to easily drill down the data and ML execution
- **REST API and SDK(Python)** to managerecords, experiment, pipelines during ML execution
- “experiment metadata” provided through UI => no need to ssh the machine for info

• ETL (Spark)

Build for easy integration with the ML platform

- full integration with Asset Catalog
- created own Spark adapters, Spark Operator, to deploy and run **Spark on Kubernetes**
- created own Spark Python SDK to create Databricks Spark Cluster, deploy and execute spark job on Databricks
- integration of **Azure Data Factory** with Databricks and **Asset Catalog** to support **Azure Service Bus**(distributed queue)

• Automation (CI/CD)

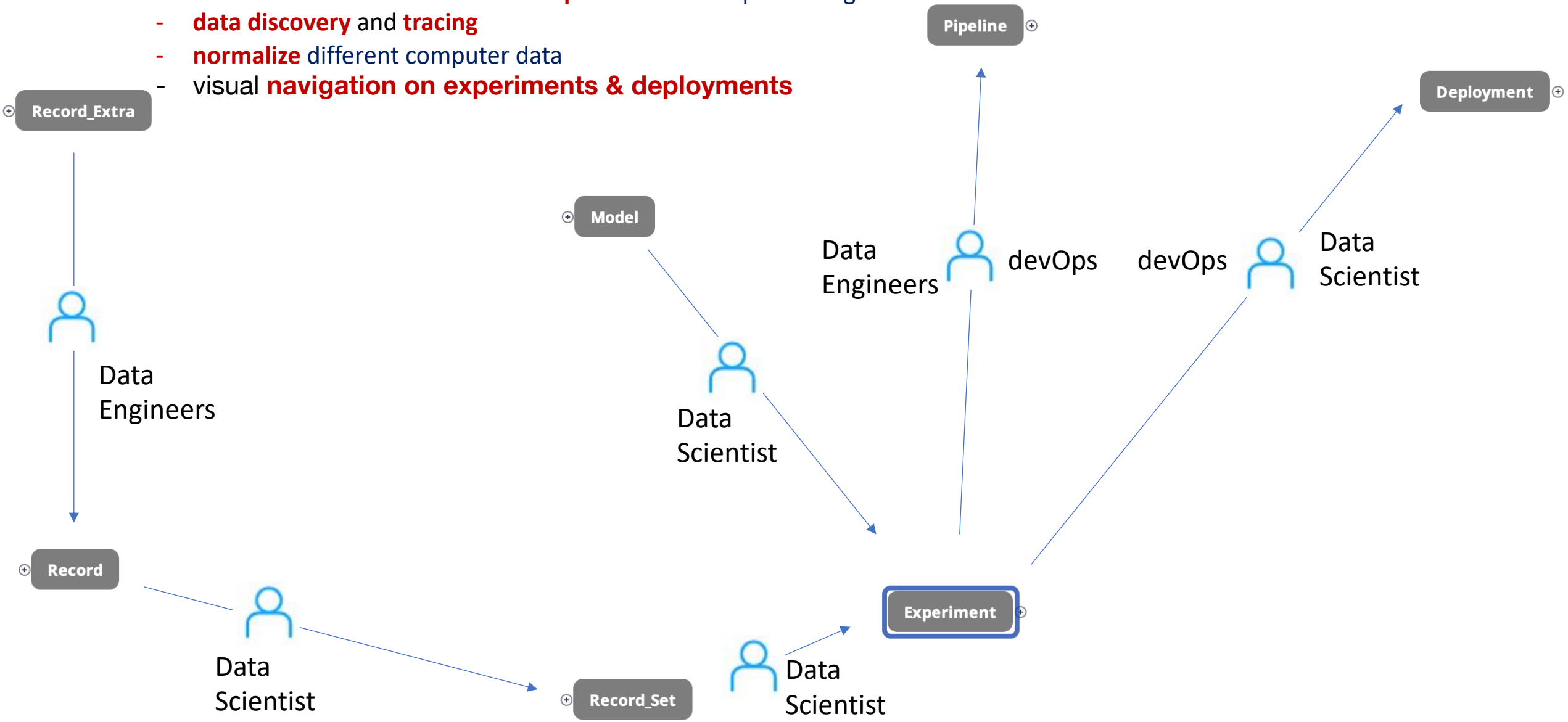
Created for an easy code and flow integration with user's demands

- smooth **propagation of released code**
- deploy easily "exact the same Docker image" in **different environments** like development, integration system, production
- built on top of Jenkins and **devOps best practices**

NextGenML Platform

Developing a cognitive interface between engineers and technology – **data catalog**

- **unify the work** for data scientists across different Line of Business
- to automate most of their **manual process** for data processing
- **data discovery** and **tracing**
- **normalize** different computer data
- visual **navigation on experiments & deployments**



Kubeflow

```
import kfp.dsl as dsl
from kubernetes import client as k8s_client

training_op = dsl.ContainerOp(
    name='training',
    image=AZURE_ML___GPU_IMAGE,
    command=['bash', '-c'],
    arguments=[...],
    file_outputs={'output': '....'})

training_op.add_pod_annotation('....', 'false')
training_op.set_gpu_limit("1")
```

Input parameters

Output parameters

dataingestion-output /tmp/output

Arguments

```
du -h /mnt/disk/ ; python3 main.py -
```

Command

```
sh
-c
```

Image

nproduction.azurecr.io/ -ingestion:307

devOps

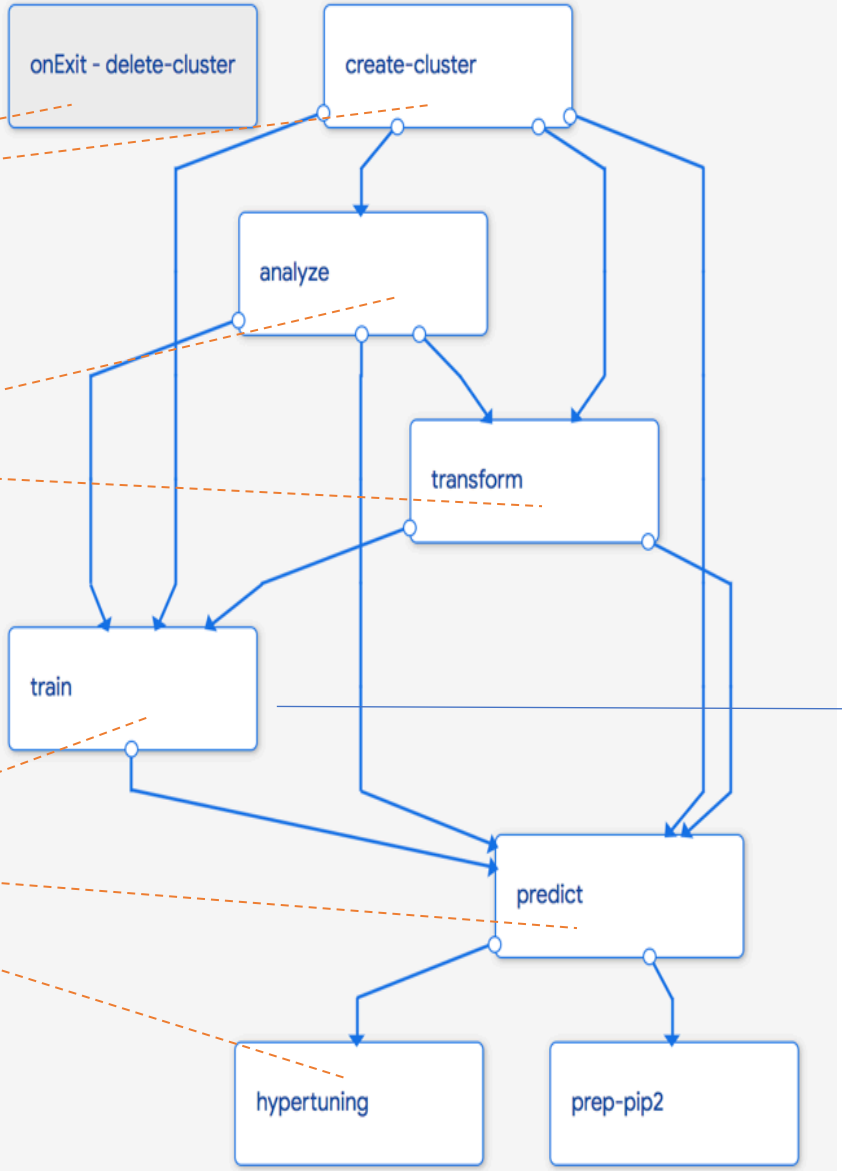
Data Engineers

Data Scientist

CPU Cluster

Ness Digital Engineering

NextGenML Platform



TensorFlow

Keras

GPU

TensorFlow

Keras

GPU

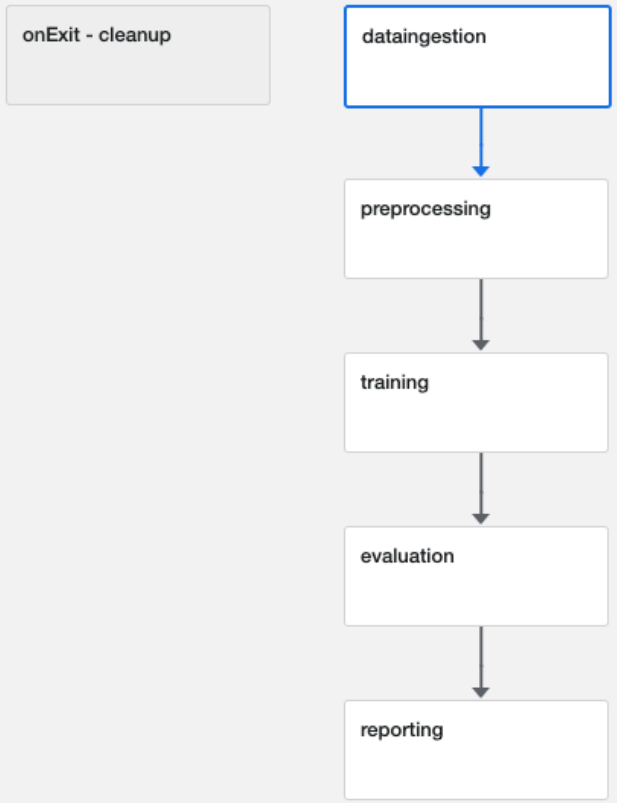
User implementation

System automation

Data Ingestion

Ness Digital Engineering — NextGenML Platform

Graph | Source



/dataingestion

Input parameters

mount-path-disk-preprocessed	
record-set-name	
store-name	

Output parameters

dataingestion-output	/tmp/output

Arguments

```
echo WorkflowId: {{workflow.uid}}; python3 main.py --store-name {{inputs.parameters.store-name}} --datacatalog-manager
http://10.97.102.65:8000/api --destination-path {{inputs.parameters.mount-path-disk-preprocessed}} --record-set-name
{{inputs.parameters.record-set-name}}
```

Command

```
sh
-c
```

Image

```
rotsrlxdv22.ness.com:8445/ml-ingestion:12
```

Volume Mounts

/mnt/vnet/data	ml-ingestion
/mnt/ml-experiments	ml-ingestion

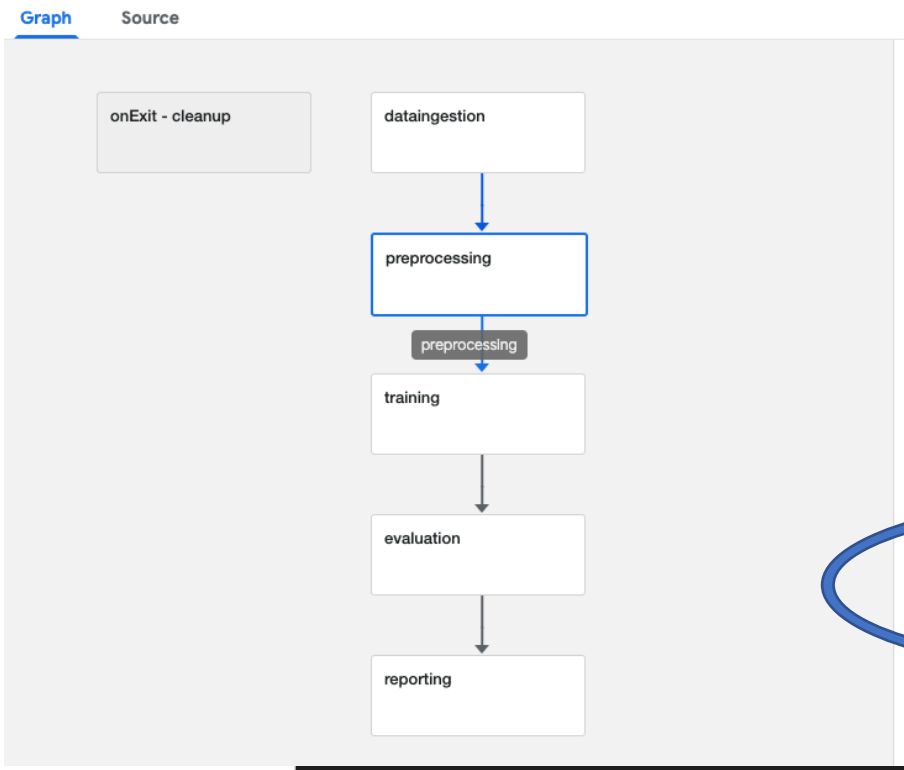
Run parameters

record-set-name	lung_data_set
store-name	/mnt/vnet/data/data_lake/clinics
mount-path-disk-preprocessed	/mnt/vnet/data/data_sets
training-epochs	--epochs 2

Data Preparation\Transformation

Ness Digital Engineering

NextGenML Platform



/preprocessing

Input parameters

dataingestion-output

Output parameters

preprocessing-output	/tmp/output

Arguments

```
set -o pipefail ; python preprocess.py --input-path={{inputs.parameters.dataingestion-output}}
```

Command

```
bash
-c
```

Image

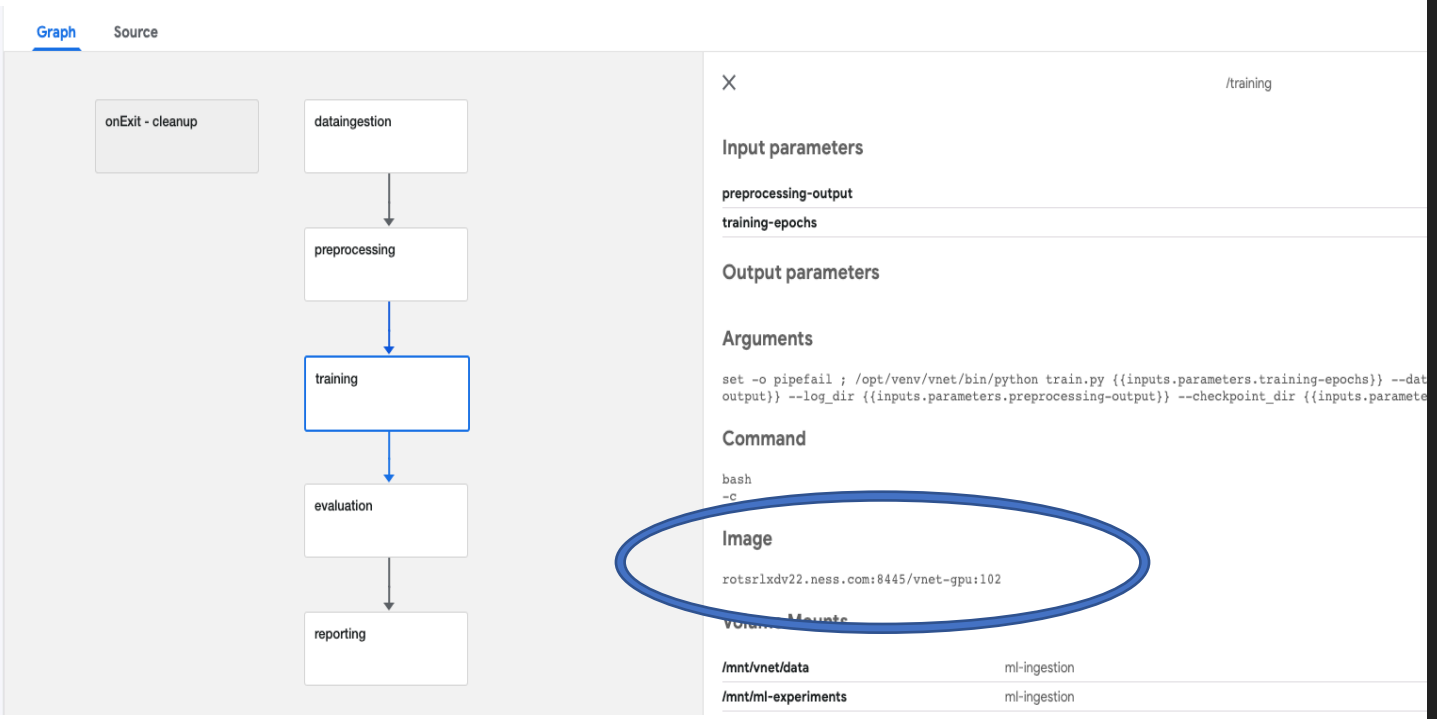
```
rotsrlxdv22.ness.com:8445/ml-preprocess:2
```

Volume Mounts

/mnt/vnet/data	ml-ingestion
/mnt/ml-experiments	ml-ingestion

```
1495 INFO:preprocess.dicom_transform:******* structureid: [structure, ROINum, ContourIdx] ********
1496 INFO:preprocess.dicom_transform:{0: ['Adrenal Left', "1", 0], 1: ['Adrenal Right', "2", 1]}
1497 INFO:preprocess.dicom_transform:****** Preprocessing completed for dicomdir/mnt/disk/chw_total_training_set/_27_CHWCT_2019_04_29-12_54_57_114/DICOM
1498 ****** Resample and Crop volumes for dicomdir/mnt/disk/chw_total_training_set/_27_CHWCT_2019_04_29-12_54_57_114/DICOM
1499 08-09 07:09 INFO : ****** Preprocessing completed for dicomdir/mnt/disk/chw_total_training_set/_27_CHWCT_2019_04_29-12_54_57_114/DICOM
1500 INFO:preprocess.preprocess_dicom:[Preprocessing finished successfully]
1501 INFO:__main__:[preprocessing finished successfully...]
1502 file counter #   78    /mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/nifti/img77.nii.gz    /mnt/share/output/results/PROD_Abdomen_A
1503 08-09 07:09 INFO : [Preprocessing finished successfully]
1504 08-09 07:09 INFO : [preprocessing finished successfully...]
1505 INFO:root:[updating catalog for preprocessing]
1506 INFO:root:**************** Configuration ***************
1507 INFO:root:{'input_folder': '/mnt/disk/', 'output_folder': '/mnt/share/output/results', 'kubeflow_pipeline_url': 'http://10.0.232.44/pipeline', 'tb_log_path': '/mnt/
1508 INFO:root:**************** preprocess: updating an Experiment into Asset Catalog ***************
1509 08-09 07:09 INFO : [updating catalog for preprocessing]
1510 08-09 07:09 INFO : **************** Configuration ***************
1511 08-09 07:09 INFO : {'input_folder': '/mnt/disk/', 'output_folder': '/mnt/share/output/results', 'kubeflow_pipeline_url': 'http://10.0.232.44/pipeline', 'tb_log_path
1512 08-09 07:09 INFO : **************** preprocess: updating an Experiment into Asset Catalog ***************
1513 Saved input to /mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/configuration.json
1514 08-09 07:09 INFO : **************** preprocess: updating an Experiment into Asset Catalog finished ***************
1515 INFO:root:**************** preprocess: updating an Experiment into Asset Catalog finished ***************
1516 Saved input to input/configuration.json
```

Training Model




```
2467  1/19 [>.............................] - ETA: 45s - loss: 0.6598 - categorical_accuracy: 1.0000
2468  2/19 [==>...........................] - ETA: 42s - loss: 0.6606 - categorical_accuracy: 1.0000
2469  3/19 [===>..........................] - ETA: 40s - loss: 0.6589 - categorical_accuracy: 1.0000
2470  4/19 [=====>........................] - ETA: 37s - loss: 0.6572 - categorical_accuracy: 1.0000
2471  5/19 [======>.......................] - ETA: 35s - loss: 0.6573 - categorical_accuracy: 1.0000
2472  6/19 [========>.....................] - ETA: 32s - loss: 0.6554 - categorical_accuracy: 1.0000
2473  7/19 [==========>...................] - ETA: 30s - loss: 0.6564 - categorical_accuracy: 1.0000
2474  8/19 [==========>...................] - ETA: 27s - loss: 0.6572 - categorical_accuracy: 1.0000
2475  9/19 [=============>................] - ETA: 25s - loss: 0.6576 - categorical_accuracy: 1.0000
2476 10/19 [==============>...............] - ETA: 22s - loss: 0.6579 - categorical_accuracy: 1.0000
2477 11/19 [===============>..............] - ETA: 20s - loss: 0.6579 - categorical_accuracy: 1.0000
2478 12/19 [=================>............] - ETA: 17s - loss: 0.6582 - categorical_accuracy: 1.0000
2479 13/19 [==================>...........] - ETA: 15s - loss: 0.6578 - categorical_accuracy: 1.0000
2480 14/19 [====================>.........] - ETA: 12s - loss: 0.6581 - categorical_accuracy: 1.0000
2481 15/19 [=====================>........] - ETA: 10s - loss: 0.6579 - categorical_accuracy: 1.0000
2482 16/19 [======================>.....] - ETA: 7s - loss: 0.6575 - categorical_accuracy: 1.0000
2483 17/19 [=======================>....] - ETA: 5s - loss: 0.6578 - categorical_accuracy: 1.0000
2484 18/19 [=========================>..] - ETA: 2s - loss: 0.6580 - categorical_accuracy: 1.0000
2485 19/19 [==============================] - 54s 3s/step - loss: 0.6583 - categorical_accuracy: 1.0000 - val_loss: 0.6626 - val_categorical_accuracy: 1.0000
2486 Epoch 100/100
```

Ness Digital Engineering

NextGenML Platform

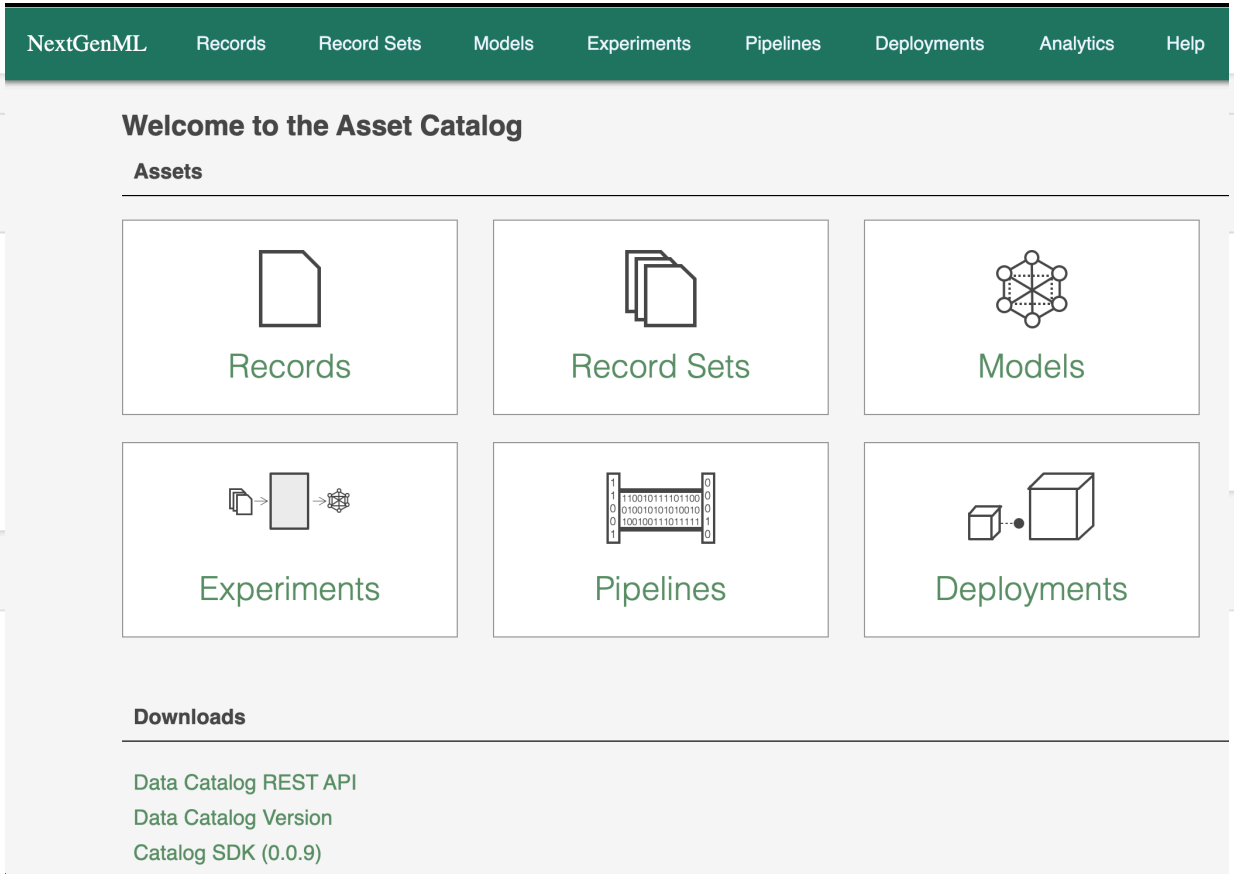
```
08-09 08:44 INFO : [updating catalog for training]
INFO - __main__ - 2019-08-09 08:44:38,970:
Configuration after training:
{'asset_catalog_experiment_id': 1093,
 'augmentation_parameters': {'change_intensity': 0.1,
                             'label_symmetry_map': [1
                             'lr_flip': False,
                             'rotate_random': 10.0,
                             'scale_random': 0.4,
                             'translate_random': 15.0},
 'conn_file_name': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/conn_file_name',
 'evaluation_parameters': {'dice_filename': 'Dice_evaluation.csv',
                           'msd_filename': 'MSD_evaluation.csv',
                           'save_volumes': True},
 'experiment_name': 'PROD_Abdomen_AdrenalLeft',
 'experiment_results_path': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30',
 'experiment_run': 'chw_run17',
 'input_folder': '/mnt/disk/',
 'kubeflow_pipeline_url': 'http://10.0.232.44/pipeline',
 'kubeflow_run_id': 'ffb0d61d-ba6f-11e9-bae7-1a3459758944',
 'model_parameters': {'asymmetric_decoder': 'True',
                      'cropped_vol_size': [192, 192, 168],
                      'dropout_rate': 0.1,
                      'init_filters': 4,
                      'input_size': [192, 192, 168],
                      'kernel_size': 3,
                      'model_name': 'PROD_Abdomen_AdrenalLeft_chw_run17_19-08-09T06-36-30.h5',
                      'normalization_layer': 'group',
                      'num_layers_per_block': [1, 3, 3]},
 'normalization_parameters': {'normalization_divisionby_value': 1000,
                              'normalization_low_high_pecentile': [5, 95],
                              'normalization_mean_range': [180, 360],
                              'normalization_technique': 'divisionbyvalue'},
 'output_folder': '/mnt/share/output/results',
 'preprocessing_parameters': {'clamp_range': [-900, 4096],
                              'image_size': [256, 256, 168],
                              'input_folder': '/mnt/disk/chw_total_training_set',
                              'output_folder': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T
                              'padding': [128, 128, 128],
                              'resolution': [2.0, 2.0, 2.0],
                              'structures_to_train': ['Adrenal Left',
                                                      'Adrenal Right'],
                              'temp_dir': '/mnt/disk/CHWCT/'},
 'reporting_parameters': {'report_file_name': 'report.html'},
 'start_time': '19-08-09T06-36-30',
 'stats': {'loss_results_path': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/train_
           'model_output_path': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/train_
 'tb_log_path': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30',
 'test_record_sets': 'chw_total_training_set',
 'train_parameters': {'acc_function': <function categorical_accuracy at 0x7fc07448eae8>,
                      'batch_size': 3,
                      'input_folder': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/
                      'input_normalization': 'simplect',
                      'labels_to_train': [1, 2],
                      'loss_function': <function dice_loss at 0x7fc0785650d0>,
                      'model_name': 'Live1.h5',
                      'num-training-epochs': 70,
                      'num_cpus': 16,
                      'num_gpus': 1,
                      'num_training_epochs': 100,
                      'optimizer': 'adam',
                      'optimizer_parameters': {'beta_1': 0.9,
                                               'beta_2': 0.999,
                                               'decay': 0.0,
                                               'epsilon=': 1e-08,
                                               'lr': 0.001},
                      'train_output_folder': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06
                      'train_test_split': 0.9,
                      'train_test_split_seed': 1234},
 'training_record_sets': 'chw_total_training_set',
 'uid': 'default_usr',
 'varian_manager': 'http://220.0.0.97:8000/api',
 'varian_manager_file_name': '/mnt/share/output/results/PROD_Abdomen_AdrenalLeft/chw_run17_19-08-09T06-36-30/varian_ma
 'workflow_id': 'ffb0d61d-ba6f-11e9-bae7-1a3459758944'}
```

NextGenML | Records | Record Sets | Models | Experiments | Pipelines | Deployments | Analytics | Help

Welcome to the Ness Asset Catalog

Asset Catalog

- Records
- Record Sets
- Models
- Experiments
- Pipelines
- Deployments



Data Catalog- Record

Ness Digital Engineering — NextGenML Platform

Powered by Ness

Records | Record Sets | Models | Experiments | Pipelines | Deployments | Analytics | Help

[Pipeline][Count: 54]

Search

Name: | State: | Category:

- AUTO SEGMENTATION
- cat
- Category
- Category

Version: | Uri: | Comments: | Description:

Framework: | Structure: | Metadata: | Creation Date:

« previous 1 2 3 next »

Name\Version\Category\Creation Date	Uri\Description\Comments\State	Framework\Structure\Metadata
Auto - Pipeline9ac9c134-6c49-40b8-bea4-3bcab08b4a57 **Version:** **Category:** AUTO SEGMENTATION **Create date:** Aug. 9, 2019, 8:42 a.m. **Updated at :** Aug. 9, 2019, 8:42 a.m.	**Uri:** **Description:** **Comments:**	**Framework:** `{'executor': 'kubeflow'}`

v10_y_test_9.npy

Parent:
Uri: mnist/data/test/y_test_9.npy
Description: mnist data for CNN training & validation & testing ML model
Logic info: `{}`
Metadata:

```
{ 'file_hash': '8a61469f7ea1b51cbae51d4f78837e45',
  'original_source': 'https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz'}
```

Ref Link: /tmp/mnist/y_test_9.npy
Ref Date: Aug. 30, 2019, 2:30 p.m.
Ref Source: Collaborator 0
Ref MD5:
Ref Size: 0
Creator: Ness Guest
Date: Oct. 3, 2019, 11:04 a.m.

Records (70)

Filters | Identifier: y_test | Types: CT Slice | Sources: Collaborator 0 | Creation Date After: 2019-10-01 12:32 | Reset | Filter

Download | Create record set

ID	Type	Description	Last update	Metadata	References	Logic Info	Link	More
v10_y_test_9.npy	CT Slice	mnist data for CNN training & validation & …		0	3	0		–
v10_y_test_8.npy	CT Slice	mnist data for CNN training & validation & …						+

Creation Date: 2019-10-03 14:04

Date: 2019-08-30T14:30:00Z
Link: /tmp/mnist/y_test_9.npy
Source: 1

Data Catalog- Experiment

Ness Digital Engineering — NextGenML Platform

Powered by Ness

Records | Record Sets | Models | Experiments | Pipelines | Deployments | Analytics | Help

[Experiment]
[Count: 54]

+ Create New

Search

Name:

Description:

Parameters:

Configuration:

Metadata:

Training State:

Training Output link:

Training Report Link:

Training Stats:

Training Metrics:

Creation Date:

« previous | 1 | 2 | 3 | next »

Name\Description\Training State	Training (Stats\Metrics\Output Link\ Report Link)	Parameters\Configuration\Metadata
Experiment-autobc45c0e3-07c1-4e2a-8242-099e39401d25 Model: model1_varian_sdk58982 Training Set: record_set_variant_sdk_training16863 record_set_variant_sdk_training55341 Test Set: No data	Training Stats: Training Metrics: Training Output: Training Report:	Metadata: {'gender': 'female'} Parameters: {'param': '1'} Configuration: {'config': '1'} **State:** PREPARING **Description:** **Create date:** Aug. 9, 2019, 8:42 a.m. **Updated at:** Aug. 9, 2019, 8:42 a.m.

Training (Stats\Metrics\Output Link\ Report Link)

Training Stats:


```
{ '100': { 'Preparing for ingestion.': { 'Update date': '2019-09-30 '
                                                        '13:25:42.521
693'}},
  '101': { 'Ingestion finished successfully': { 'Update date': '2019-
09-30 '
                                                        '13:2
5:42.521719'}},
  '200': { 'Preparing for training.': { 'Update date': '2019-09-30 '
                                                        '13:25:59.0662
60'}},
  '201': { 'Model training finished successfully.': { 'Update date':
'2019-09-30 '

'13:25:59.066295'}},
  '300': {'Started Evaluation.': {'Update date': '2019-09-30 13:28:4
7.943689'}},
  '301': { 'Evaluation finished successfully.': { 'Update date': '201
9-09-30 '
                                                        '13:
28:47.943715'}}}
```


Training Metrics:


```
{ 'epoch-0': { 'acc': 0.9291874766349792,
               'loss': 0.2338664118591696,
               'val_acc': 0.9712499976158142,
               'val_loss': 0.0957180842316399},
  'epoch-1': { 'acc': 0.9712708592414856,
               'loss': 0.09555063702126964,
               'val_acc': 0.9789166450500488,
               'val_loss': 0.0700550024897481},
  'epoch-2': { 'acc': 0.9793750047683716,
               'loss': 0.06580551405716688,
               'val_acc': 0.9800833463668823,
               'val_loss': 0.0672117559503143},
  'epoch-3': { 'acc': 0.984083354473114,
               'loss': 0.04865670974641883,
               'val_acc': 0.9832500219345093,
               'val_loss': 0.05802248525169368},
  'epoch-4': { 'acc': 0.9884583353996277,
               'loss': 0.035777118205636116,
               'val_acc': 0.9835833311080933,
               'val_loss': 0.06274347158684396},
  'evaluation_acc': 0.982200026512146,
  'evaluation_loss': 0.06102500319739629}
```


Training Output:
/mnt/ml-experiments/9c4e3768-baf2-4a7a-bdd6-52701d8c0559
Training Report:

Parameters\Configuration\Metadata

Metadata:


```
{'kfp_run_id': '9c4e3768-baf2-4a7a-bdd6-52701d8c0559'}
```


Parameters:


```
{ 'common-data-set-path': '/mnt/data-sets/mnist',
  'epochs': '5',
  'record-set-name': 'mnist_data_set',
  'validation-set': '0.2'}
```


Configuration:


```
{}
```


State: DONE
Description: Evaluation finished successfully.
Create date: Sept. 30, 2019, 1:25 p.m.
Updated at: Sept. 30, 2019, 1:25 p.m.

Data Catalog- Model

Ness Digital Engineering

NextGenML Platform

Powered by Ness

Records | Record Sets | Models | Experiments | Pipelines | Deployments | Analytics | Help

[Model][Count: 101]

Search

Name:

Version:

VCS Link:

VCS Branch name:

VCS Commit Id:

Framework:

Hyper parameters:

Metadata:

Creation Date:

« previous 1 2 3 4 5 next »

Name\Version\State\Creation Date	VCS (Link\Branch Name\Commit Id)	Metadata/Framework/Hyper parameters

model1_sdk649b77f0-8bef-4e54-9e61-cadeef7f9787
Version: 1.0-alpha-sdk
Create date: Aug. 9, 2019, 8:42 a.m.
Updated at: Aug. 9, 2019, 8:42 a.m.
	VCS Link:
	VCS Branch Name:
	VCS Commit Id:
		Metadata:
		`{'json_field_int': 2, 'json_field_string': 'CT'}`
		Framework:
		`{'key': 'value1_sdk'}`
		Hyper Parameters:
		`{'level': 1, 'string_value': 'this is a string value'}`

Data Catalog- Record Set

record_set_variant_sdk_training16863

Download

Type: TRAINING SET

Description:

Metadata: {}

Summary: {}

Histogram: {}

Creator: 1

Date: July 3, 2019, 9:01 a.m.

Update

Record Updates

[CT Image] [Count: 1]

Identifier	Creator	Metadata
demo ct_image 1318199616322289009740643896298390956869441 03 **Desc:** sdk_ctimage_description **Uri:** /Clinical Sites/Patient/Public Websites	**Creator:** Varian Guest **Creation date:** 07/03/2019 **Reference Info:**	**Logic Info:** **Metadata:**

[CT Slice] [Count: 1]

[RT Structure Set] [Count: 1]

demo ct_image 131819961632 44103

Ness Digital Engineering

NextGenML Platform

Parent:

Uri: /Clinical Sites/Patient/Public Websites

Description: sdk_ctimage_description

Logic info:

```
{ 'institution': 'hospital',
  'patient': 10020,
  'patient_id': 1,
  'tags': ['tag1', 'tag2', 'tag3']}
```

Metadata:

```
{ 'contrast': '',
  'curator_name': '',
  'number_of_slices': 729,
  'patient_age': '',
  'patient_id': '07DC0024',
  'scan_length': 455.0,
  'slice_spacing': 0.6241426611796982,
  'structure_names': [ 'Adrenal Left',
                       'Adrenal Right',
                       'Bladder',
                       'Breast Left',
                       'Breast Right',
                       'Duodenum',
                       'Esophagus',
                       'Femoral Head Left',
                       'Femoral Head Right',
                       'Gall Bladder',
                       'Gonads',
                       'Heart',
                       'Kidney Left',
                       'Kidney Right',
                       'Large Intestine',
                       'Liver',
                       'Pancreas',
                       'Prostate',
                       'Rectum',
                       'Small Intestine',
                       'Spinal Canal',
                       'Spleen',
                       'Stomach',
                       'Thymus',
                       'UteroCervix'],
  'study_date': '20151231',
  'upload_date': '',
  'z_max': -255.5,
  'z_min': -710.5}
```

Ref Link: /landing_zone/area_1

Ref Date: Feb. 10, 2019, 1:01 a.m.

Ref Source: Collaborator 0

Ref MD5:

Ref Size: 0

Creator: Varian Guest

Date: July 3, 2019, 9:01 a.m.

Data Catalog- Analytics

Ness Digital Engineering

NextGenML Platform

Powered by Ness

Records | Record Sets | Models | Experiments | Pipelines | Deployments | Analytics | Help

Analytics

06/01/2019 - 07/31/2019

Record | Record Set | Model | Experiment | Pipeline | Deployment

Record Types:

- [x] CT Image
- [ ] CT Slice
- [x] RT Structure Set
- [x] RT Image
- [x] RT Plan

Sources:

- [ ] Collaborator 0
- [ ] Internal
- [ ] Public
- [ ] Collaborator 1
- [ ] Collaborator 2

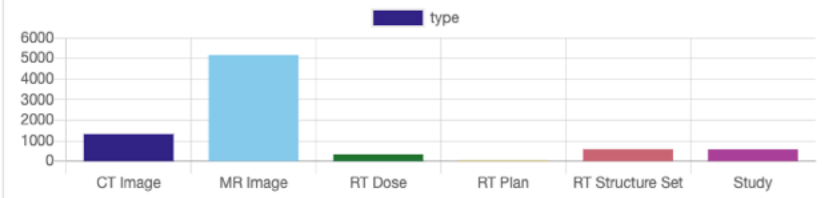
Filter

Records

Count 8046

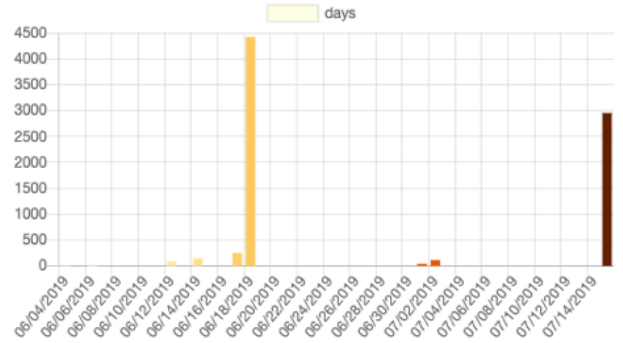
Segments

[type]		
	CT Image	1324
	MR Image	5155
	RT Dose	338
	RT Plan	53
	RT Structure Set	591
	Study	585

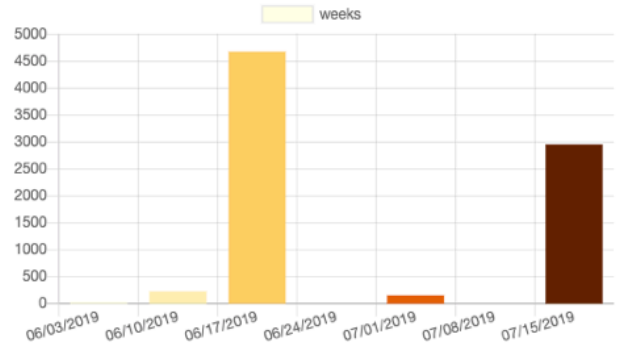


Series

Days		
	2019-06-04	15
	2019-06-06	21
	2019-06-12	87
	2019-06-14	143
	2019-06-17	253
	2019-06-18	4417
	2019-07-01	43
	2019-07-02	114
	2019-07-15	2953



Weeks		
	2019-06-03	36
	2019-06-10	230
	2019-06-17	4670
	2019-07-01	157
	2019-07-15	2953



Continuous Integration (**CI**) / Continuous Deployment (**CD**)

Ness Digital Engineering **NextGenML Platform**

Data Scientist

Data Engineers

Testing

devOps

MASTER

git

DEV

git

Team City

Azure DevOps

build

Docker

Helm

Azure Container Registry

kSonnet

CD

PRODUCTION

2 x Azure Kubernetes Cluster

CPU

GPU

CI

DEV & UAT

2 x Azure Kubernetes Cluster

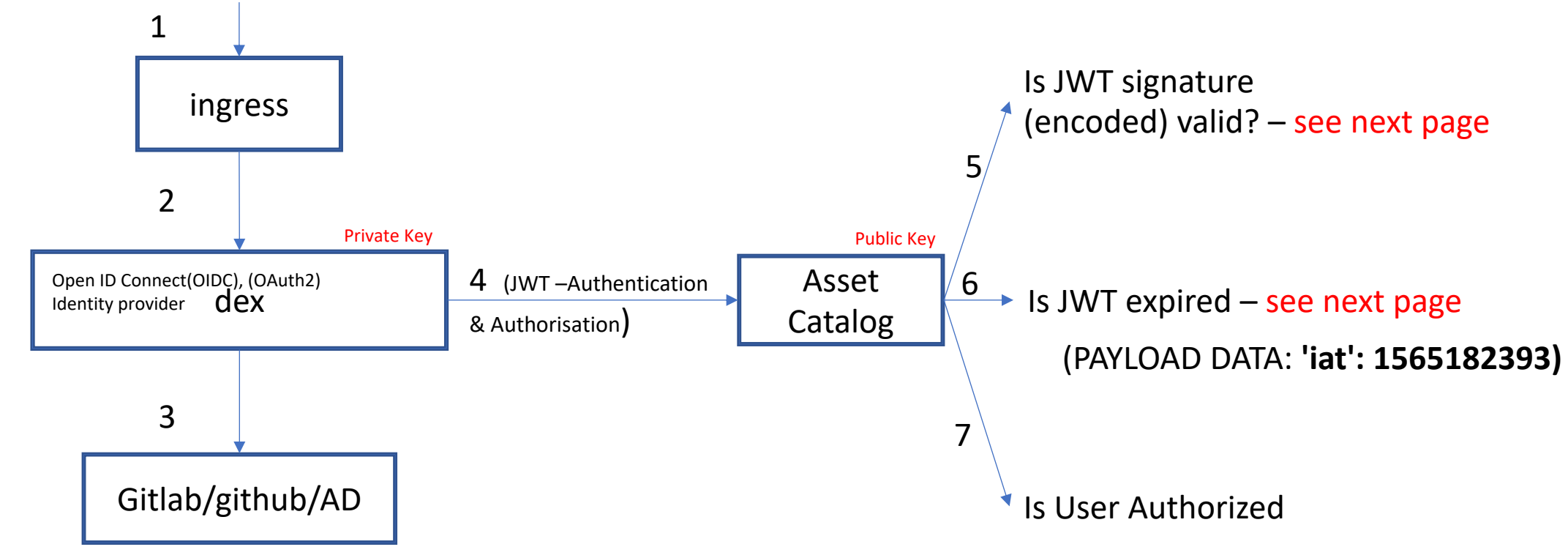
CPU

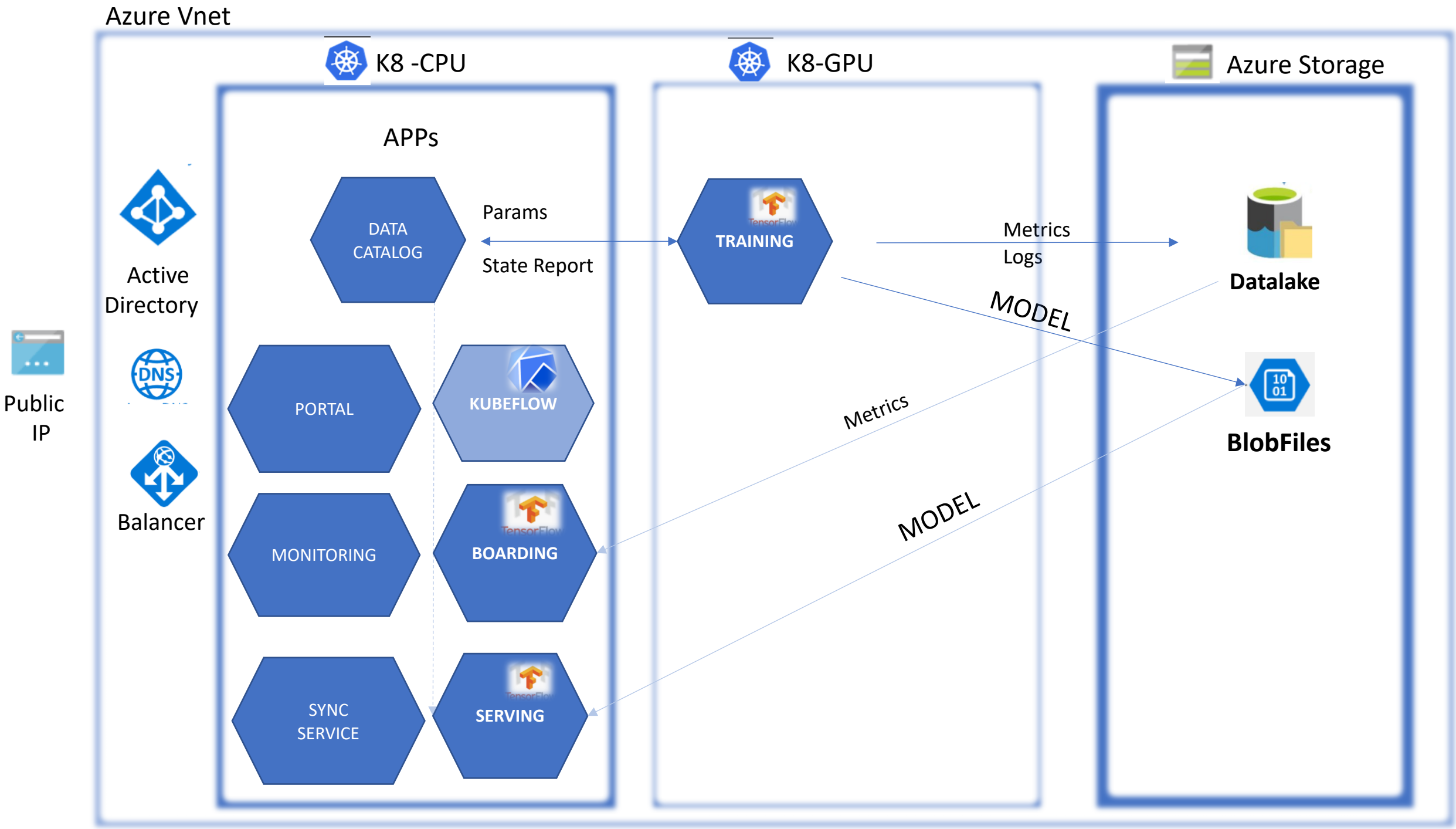
GPU

User initiated

System initiated

http://assetcatalog.rotsrlxdv29.ness.com/




Azure Vnet
K8 -CPU
K8-GPU
Azure Storage
APPs
Active Directory
DNS
Balancer
Public IP
DATA CATALOG
Params
State Report
TRAINING
Metrics
Logs
Datalake
MODEL
PORTAL
KUBEFLOW
Metrics
BlobFiles
MONITORING
BOARDING
MODEL
SYNC SERVICE
SERVING

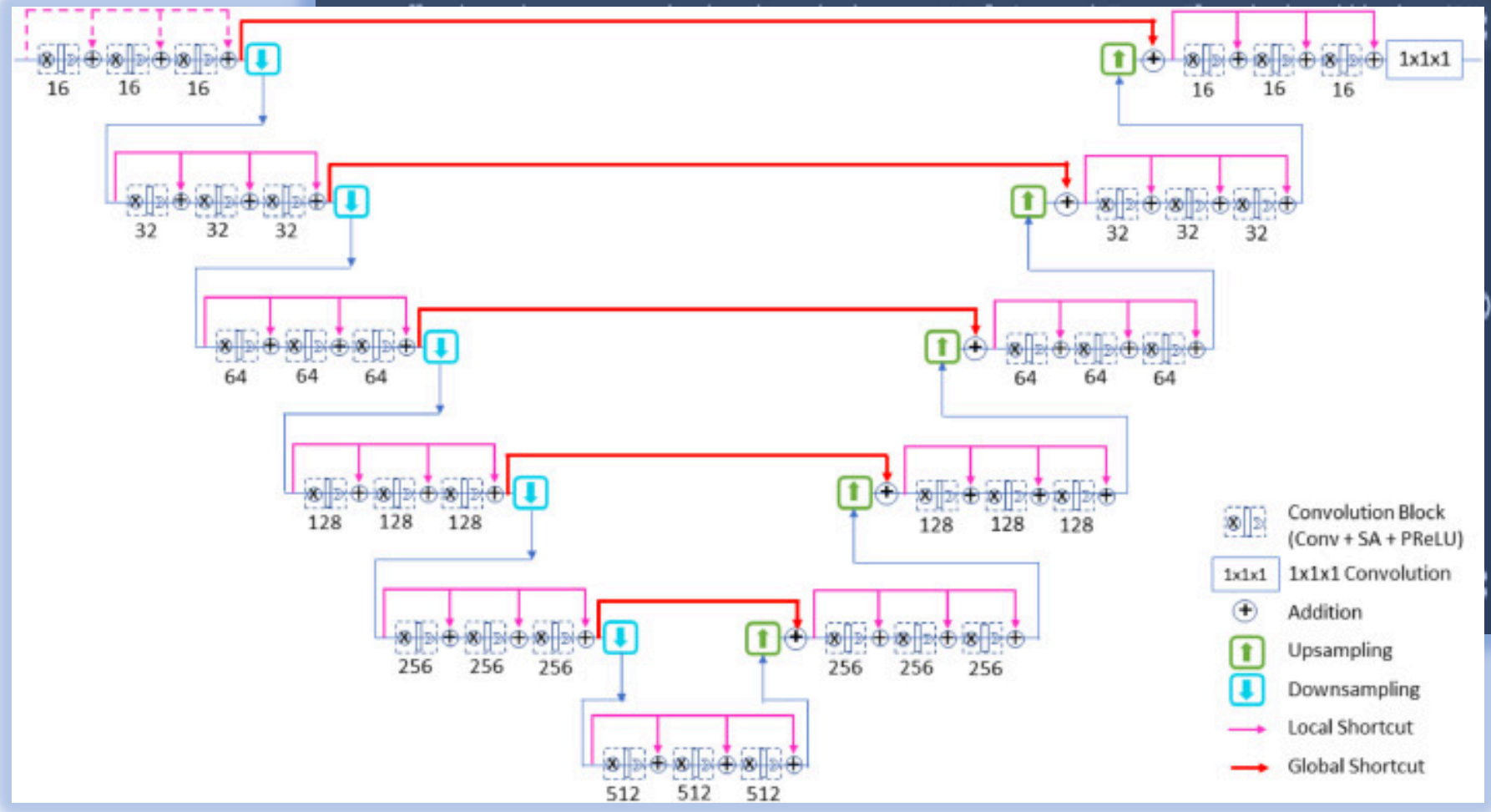
TRAINING

TensorFlow

Ness Digital Engineering

NextGenML Platform

```
tensorflow/core/common_runtime/gpu/gpu_device.cc:1511] Adding visible gpu devices: 0
tensorflow/core/common_runtime/gpu/gpu_device.cc:982] Device interconnect StreamExecutor with strength 1 edge matrix:
tensorflow/core/common_runtime/gpu/gpu_device.cc:988]      0
tensorflow/core/common_runtime/gpu/gpu_device.cc:1001] 0:   N
```




```
0 with 11486 MB memory) -> physical GPU

strength 1 edge matrix:

ost/replica:0/task:0/device:GPU:0 with 11

strength 1 edge matrix:

0 with 11486 MB memory) -> physical GPU
```

ND6s instance

BOARDING

TensorFlow

Ness Digital Engineering

NextGenML Platform

Ignore outliers in chart scaling

Tooltip sorting method: default

Smoothing

0

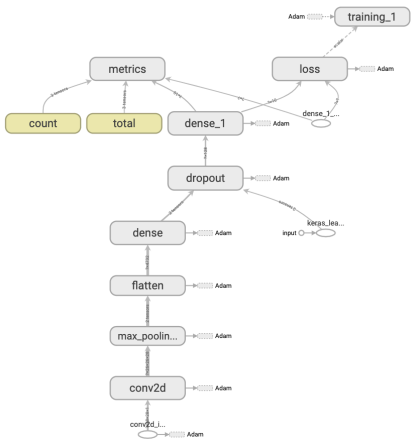
Horizontal Axis

STEP RELATIVE WALL

Runs

Write a regex to filter runs

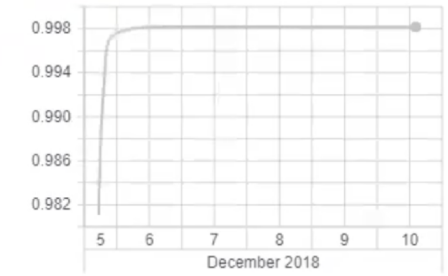
v1_hvd.h5



TOGGLE ALL RUNS

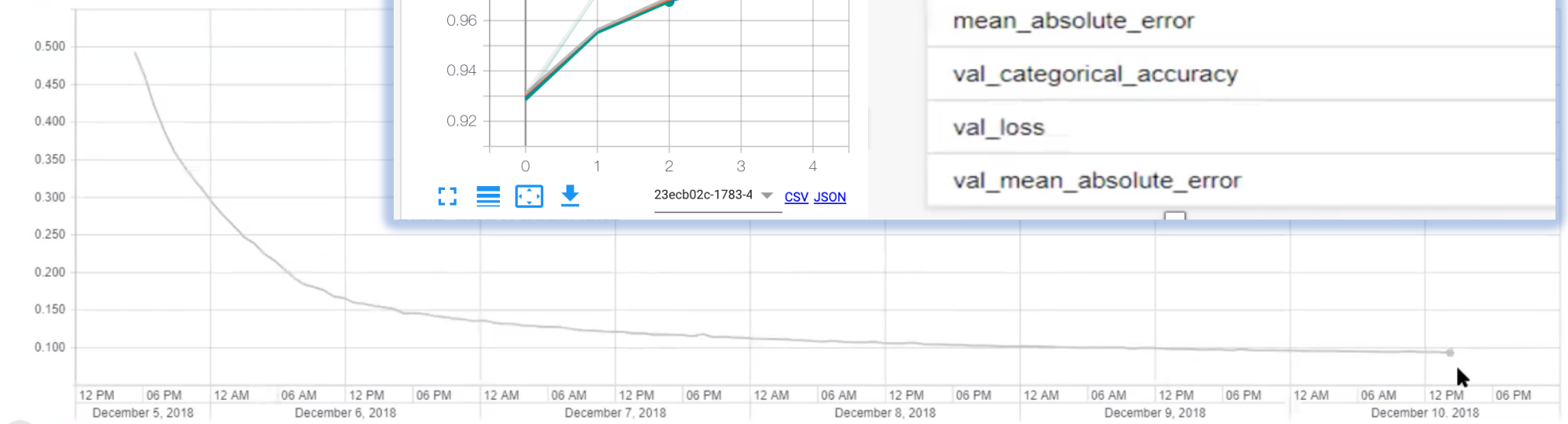
autoseg:/mnt/data/src/vmsseg/logs,petr:/mnt/data/src/logs,phil:/mnt/data/Phil_SOI/Limited_

categorical_accuracy



loss

loss



Name	Smoothed	Value	Step	Time	Relative
autoseg/pancreas_v3_2mm_120518.h5	0.09334	0.09334	132.0	Mon Dec 10, 14:14:02	4d 20h 59m 19s

autoseg/vmsnet_pancreas_small2_hvd.h5
autoseg/pancreas_1mm.h5
autoseg/pancreas_2mm.h5
autoseg/female_pelvis_2mm.h5
autoseg/female_pelvis_1mm.h5
autoseg/female_pelvis_2mm_hvd.h5
autoseg/female_pelvis_1mm_hvd.h5

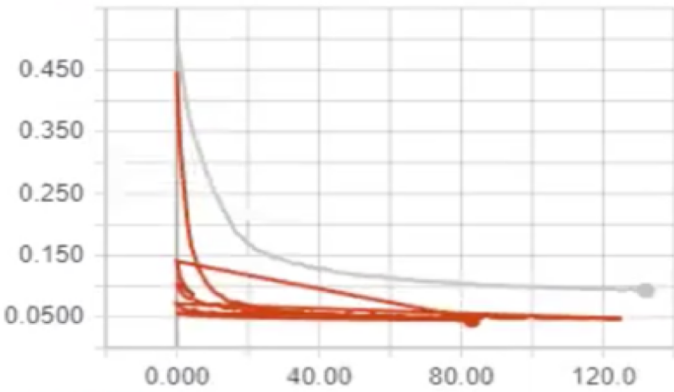
epoch_acc

epoch_acc

23ecb02c-1783-4 CSV JSON

loss

loss



mean_absolute_error

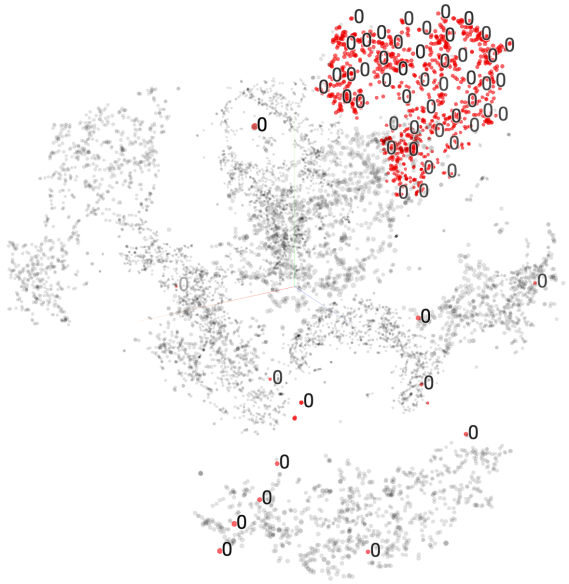
val_categorical_accuracy

val_loss

val_mean_absolute_error

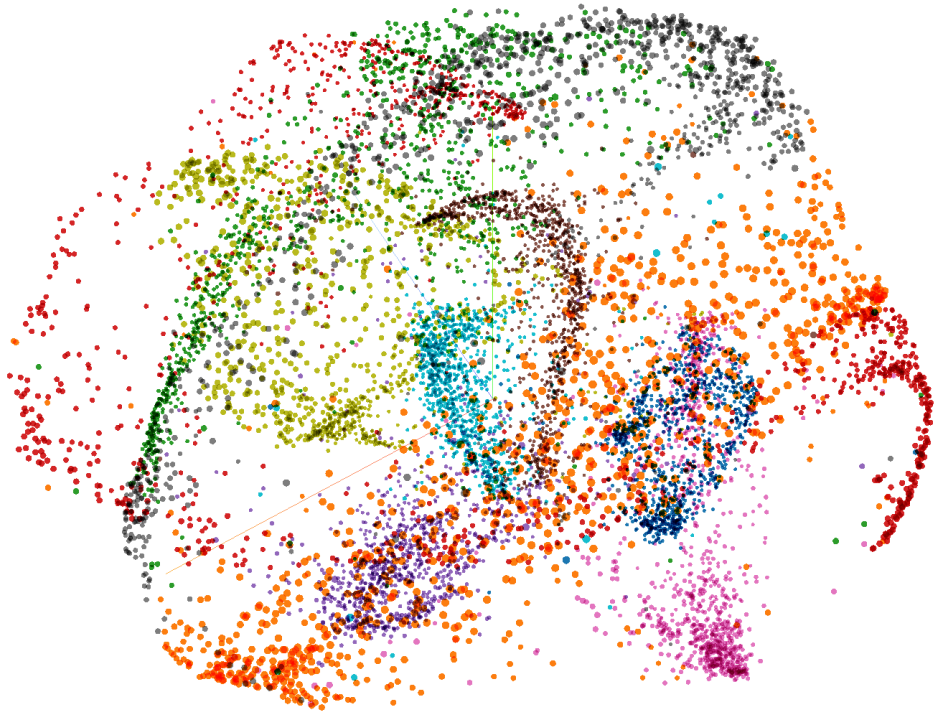
VISUALISATION

TensorFlow



Ness Digital Engineering

NextGenML Platform



TensorBoard PROJECTOR

DATA

1 tensor found
images

Label by
label

Color by
label

Supervise with
label

No ignored label

Edit by
label

Tag selection as

Load Download Label

UMAP T-SNE PCA CUSTOM

Dimension 2D 3D

Perplexity 25

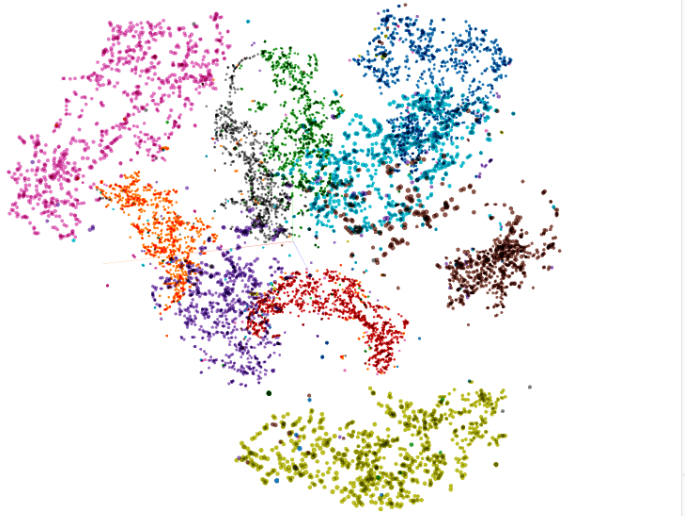
Learning rate 10

Supervise 0

Re-run Pause Perturb

Iteration: 1018

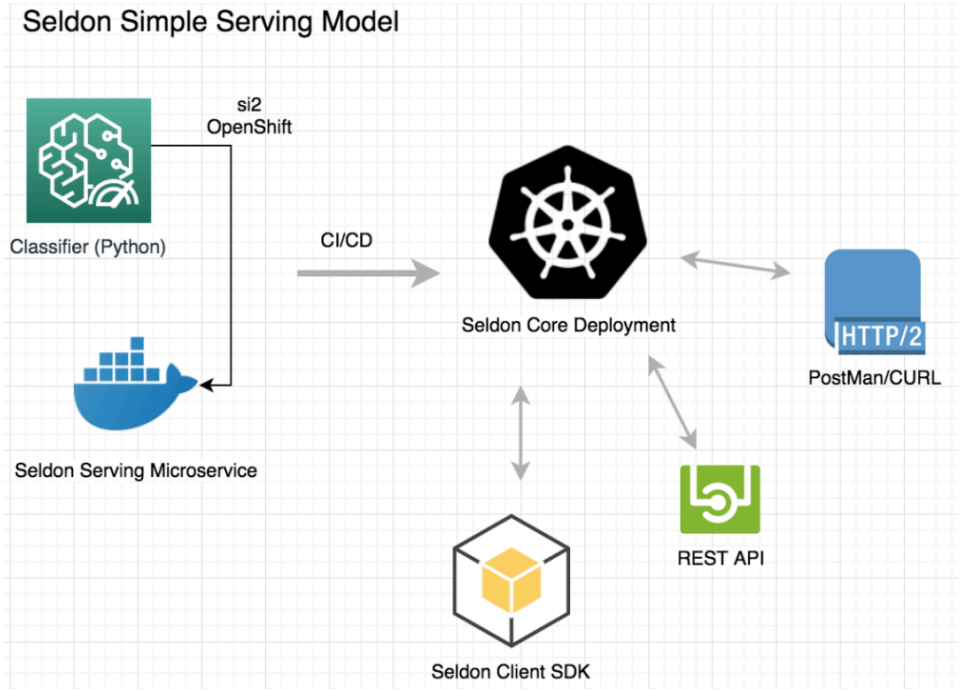
How to use t-SNE effectively.



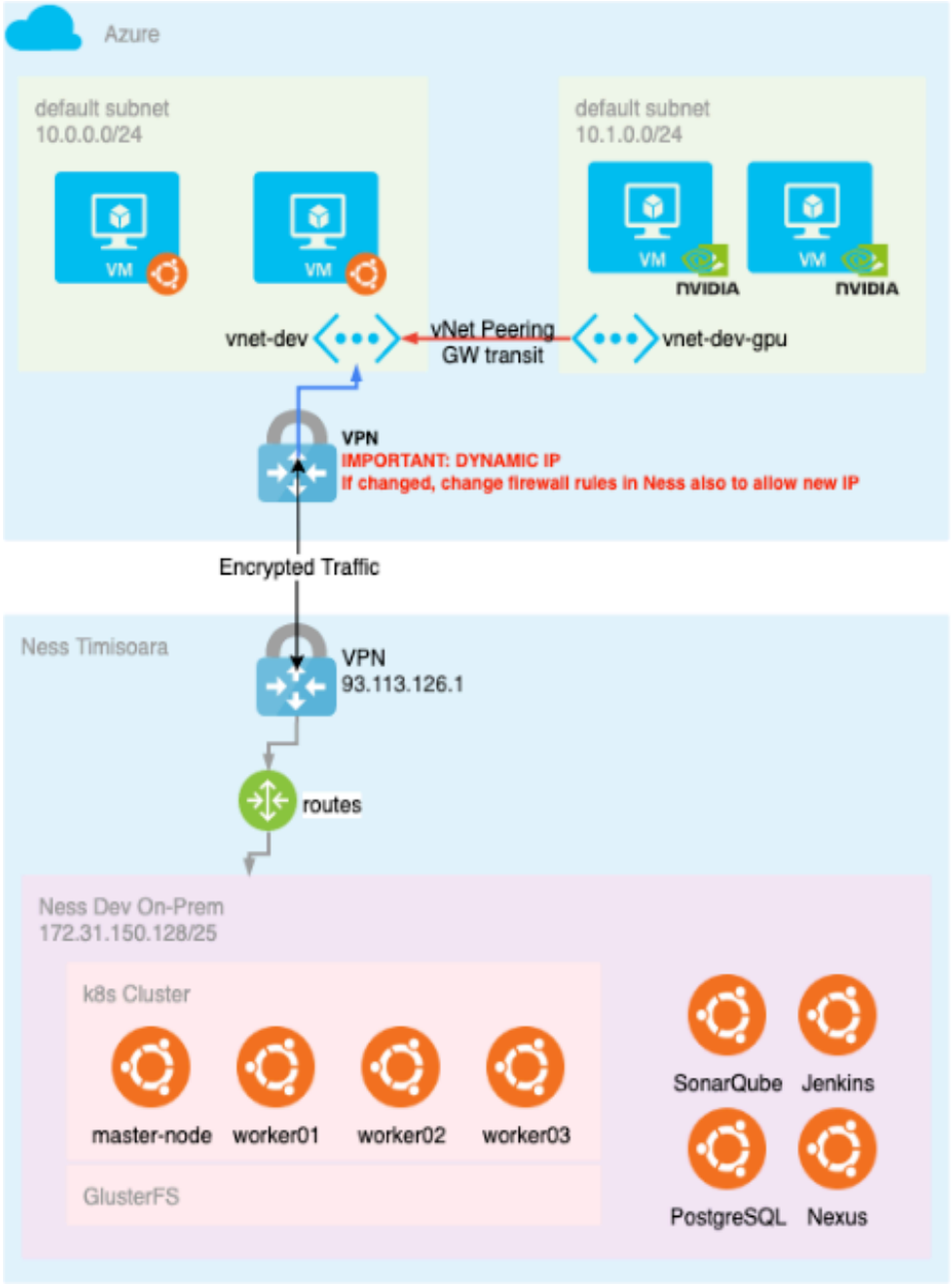
TensorFlow Serving REST API (https://www.tensorflow.org/tfx/serving/api_rest)

POST http://host:port/v1/models/${MODEL_NAME}[/versions/${MODEL_VERSION}]:(classify|regress)

```
#preparation payload
encoded_image = base64.b64encode(THE_IMAGE)

payload_json{
                    “image”:encoded_image,
                    ….
        }

# post request to Tensorflow Serving
json_response = requests.post(url=rest_url, json= payload_json{})

# decode the response and extract marked organ
response = extract_from_json_THE_2D_PREDICTION (numpy)
organ_countur = png.Reader(bytes=y.tostring())
```



Hybrid Cloud Topology



Ness Digital Engineering

NextGenML Platform

Azure Service

VPN + KeyVault
AZ DevOps Pipelines (Brigade)

Kubernetes Services (AKS)
HDInsight (Kafka & Spark)

Storage Account
DataLake Storage Gen 1 & 2

Data Factory (ADF)
Function App(lambda)
Batch Account (VM Pools)
Databricks + external metastore

Service Bus (ASB)
Cache for Redis

VM (GPU)

Machine Learning

Heterogeneous cluster (CPU+GPU)
Kubeflow(on K8s)
Tensorboard(Keras)- training/boarding/serving*
Seldon*

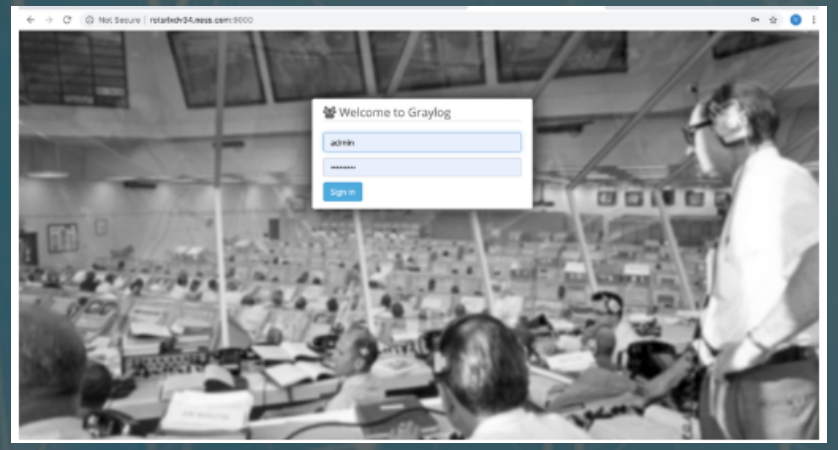
K8s
Gangway+Dex(GitLab/AD)+RBAC
Ingress(Nginx)/Istio +Services

KubeFlow
Spark
Asset Catalog
GlusterFS
Graphana + Prometeus

CI / CD

Kibana Logstash Filebit
GrayLog
Jenkins
Nexus
SonarQube**

graylog



Ness Digital Engineering

NextGenML Platform

Messages

Previous 1 Next

Timestamp	source
2019-08-09 10:58:49.834	kube-gelf-mjkcf

```
time="2019-08-09T10:58:49Z" level=error msg="Permanent failure while syncing resource (kubeflow/nextgenml-auto-segmentation-pipeline-59p5q): CustomError (code: 1): Syncing Workflow (nextgenml-auto-segmentation-pipeline-59p5q): permanent failure: CustomError (code: 1): Error while reporting workflow resource (code: InvalidArgument, message: Report workflow failed.: Invalid input error: Failed to update run 33c20cc6-e3cc-4ece-9a74-157a915bf4ef. Row not foun
```

b0e14430-ba94-11e9-af6e-005056873be6

Permalink | Copy ID | Show surrounding messages | Test against stream

Received by
Gelf UDP on
6f0e38b4 / localhost.localdomain

Stored in index
graylog_1

Routed into streams

docker
{"container_id":"c300b57ff335972553ac24905f0911863bd699bc6827adb7b11a00d161d5b370"}

facility
fluentd

kubernetes
{"container_name":"ml-pipeline-persistenceagent","namespace_name":"kubeflow","pod_name":"ml-pipeline-persistenceagent-69f558486c-7n5b8","pod_id":"cb3ee7a8-006a-47ad-b818-e4dcaee3bc8d","labels":{"app":"ml-pipeline-persistenceagent","pod-template-hash":"69f558486c"},"host":"worker01","master_url":"https://10.96.0.1:443/api","namespace_id":"be0b46cf-308e-43a0-9e0d-65746bd325c3"}

level
1

message
time="2019-08-09T10:58:49Z" level=error msg="Permanent failure while syncing resource (kubeflow/nextgenml-auto-segmentation-pipeline-59p5q): CustomError (code: 1): Syncing Workflow (nextgenml-auto-segmentation-pipeline-59p5q): permanent failure: CustomError (code: 1): Error while reporting workflow resource (code: InvalidArgument, message: Report workflow failed.: Invalid input error: Failed to update run 33c20cc6-e3cc-4ece-9a74-157a915bf4ef. Row not found.): rpc error: code = InvalidArgument desc = Report workflow failed.: Invalid input error: Failed to update run 33c20cc6-e3cc-4ece-9a74-157a915bf4ef. Row not found., &{TypeMeta:{Kind:Workflow APIVersion:argoproj.io/v1alpha1} ObjectMeta:{Name:nextgenml-auto-segmentation-pipeline-59p5q GenerateName:nextgenml-auto-segmentation-pipeline- Namespace:kubeflow SelfLink:/apis/argoproj.io/v1alpha1/namespaces/kubeflow/workflows/nextgenml-auto-segmentation-pipeline-59p5q UID:33c20cc6-e3cc-4ece-9a74-157a915bf4ef ResourceVersion:13123300 Generation:1 CreationTimestamp:2019-08-09 09:08:44 +0000 UTC DeletionTimestamp:<nil> DeletionGracePeriodSeconds:<nil> Labels:map[] Annotations:map[] OwnerReferences:[] Initializers:nil Finalizers:[] ClusterName:} Spec:{Templates:[{Name:cleanup Inputs:{Parameters:[] Artifacts:[]} Outputs:{Parameters:[] Artifacts:[{Name:mlpipeline-ui-metadata Path:/mlpipeline-ui-metadata.json Mode:<nil> From: ArtifactLocation:{ArchiveLogs:<nil> S3:<nil> Git:<nil> HTTP:<nil> Artifactory:<nil> HDFS:<nil> Raw:<nil>} GlobalName: Archive:<nil> Optional:true} {Name:mlpipeline-metrics Path:/mlpipeline-metrics.json Mode:<nil> From: ArtifactLocation:{ArchiveLogs:<nil> S3:<nil> Git:<nil> HTTP:<nil> Artifactory:<nil> HDFS:<nil> Raw:<nil>} GlobalName: Archive:<nil> Optional:true}] Result:<nil>} NodeSelector:map[] Affinity:nil Metadata:{Annotations:map[cluster-autoscaler.kubernetes.io/safe-to-evict:false] Labels:map[]} Daemon:<nil> Steps:[] Container:&Container{Name:,Image:rotsrlxdv22.ness.com:8445/vnet-gpu:102,Command:[sh -c],Args:[echo cleanup],WorkingDir:,Ports:[],Env:[],Resources:ResourceRequirements{Limits:ResourceList{},Requests:ResourceList{},},VolumeMounts:[],LivenessProbe:nil,ReadinessProbe:nil,Lifecycle:nil,TerminationMessagePath:,ImagePullPolicy:,SecurityContext:nil,Stdin:false,StdinOnce:false,TTY:false,EnvFrom:[],TerminationMessagePolicy:,VolumeDevices:[],} Script:<nil> Resource:<nil> DAG:<nil> Suspend:<nil> Volumes:[] InitContainers:[] Sidecars:[] ArchiveLocation:<nil> ActiveDeadlineSeconds:<nil> RetryStra

protocol
0

source
kube-gelf-mjkcf

stream
stderr

tag
kubernetes.var.log.containers.ml-pipeline-persistenceagent-69f558486c-7n5b8_kubeflow_ml-pipeline-persistenceagent-c300b57ff335972553ac24905f0911863bd699bc6827adb7b11a00d161d5b370.log

timestamp
2019-08-09 10:58:49.834 +00:00 i

---

graylog | Search | Streams | Alerts | Dashboards | Sources | System | In 7 / Out 7 msg/s | Help | Administrator

Search result

Found **140 messages** in 42 ms, searched in 1 index.
Results retrieved at 2019-08-07 08:55:26.

Add count to dashboard | Saved search | More actions

Fields | Decorators

Default | All | None | Filter fields

- docker
- facility
- kubernetes
- level
- message
- protocol
- source
- stream
- tag
- timestamp

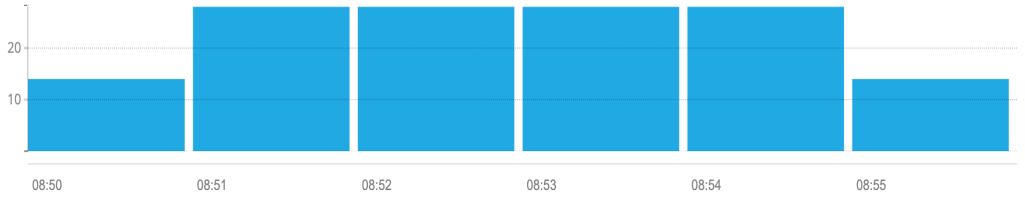
List fields of current page or all fields.

Highlight results

Histogram

Add to dashboard

Year, Quarter, Month, Week, Day, Hour, **Minute**



Messages

Previous 1 Next

Timestamp	source
2019-08-07 08:55:12.614	kube-gelf-nwl4g

```
I0807 08:55:12.613128 9 controller.go:205] SparkApplication spark/spark-python-pydicom was updated, enqueueing it
```

114f2766-b8f1-11e9-af6e-005056873be6

Permalink | Copy ID | Show surrounding messages | Test against stream

Received by
Gelf UDP on
6f0e38b4 / localhost.localdomain

Stored in index
graylog_1

Routed into streams
- All messages

docker
{"container_id":"cc3459998999c650c3c003584000a97c54575031b0a2a92c7cdecf7c74773751"}

facility
fluentd

kubernetes
{"container_name":"sparkoperator","namespace_name":"spark","pod_name":"so-sparkoperator-868574d8c4-lc6jj","pod_id":"ef0fec84-272a-440d-9be4-957467272306","labels":{"pod-template-hash":"868574d8c4","app_kubernetes_io/name":"sparkoperator","app_kubernetes_io/version":"v2.4.0-v1beta1-0.8.2"},"host":"worker01","master_url":"https://10.96.0.1:443/api","namespace_id":"790ab7f4-d2e0-4267-80ab-57bf72c97d0f"}

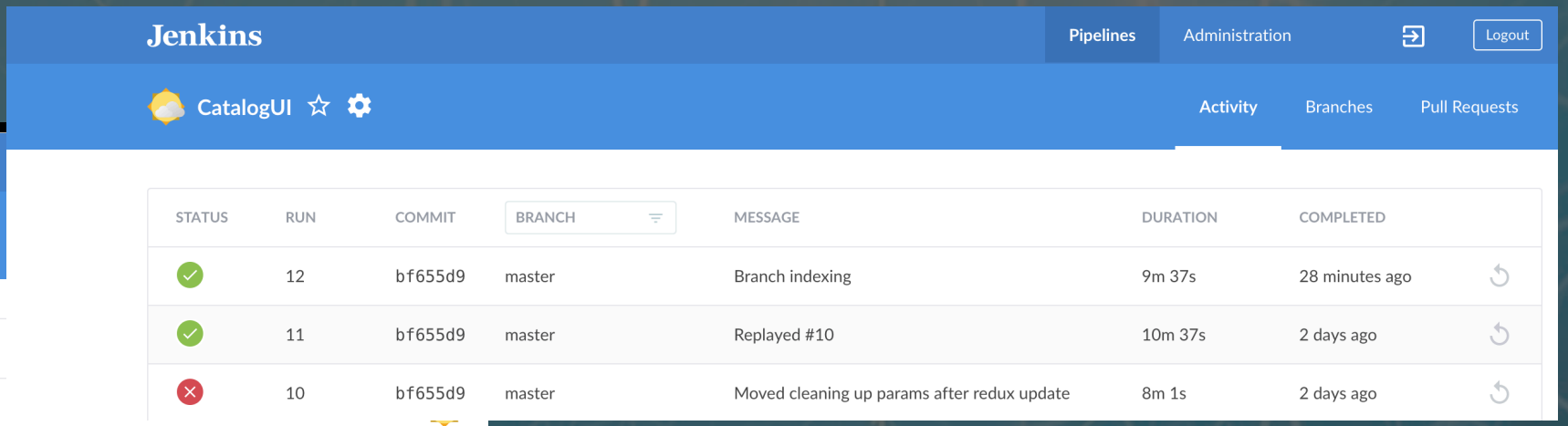
level
1

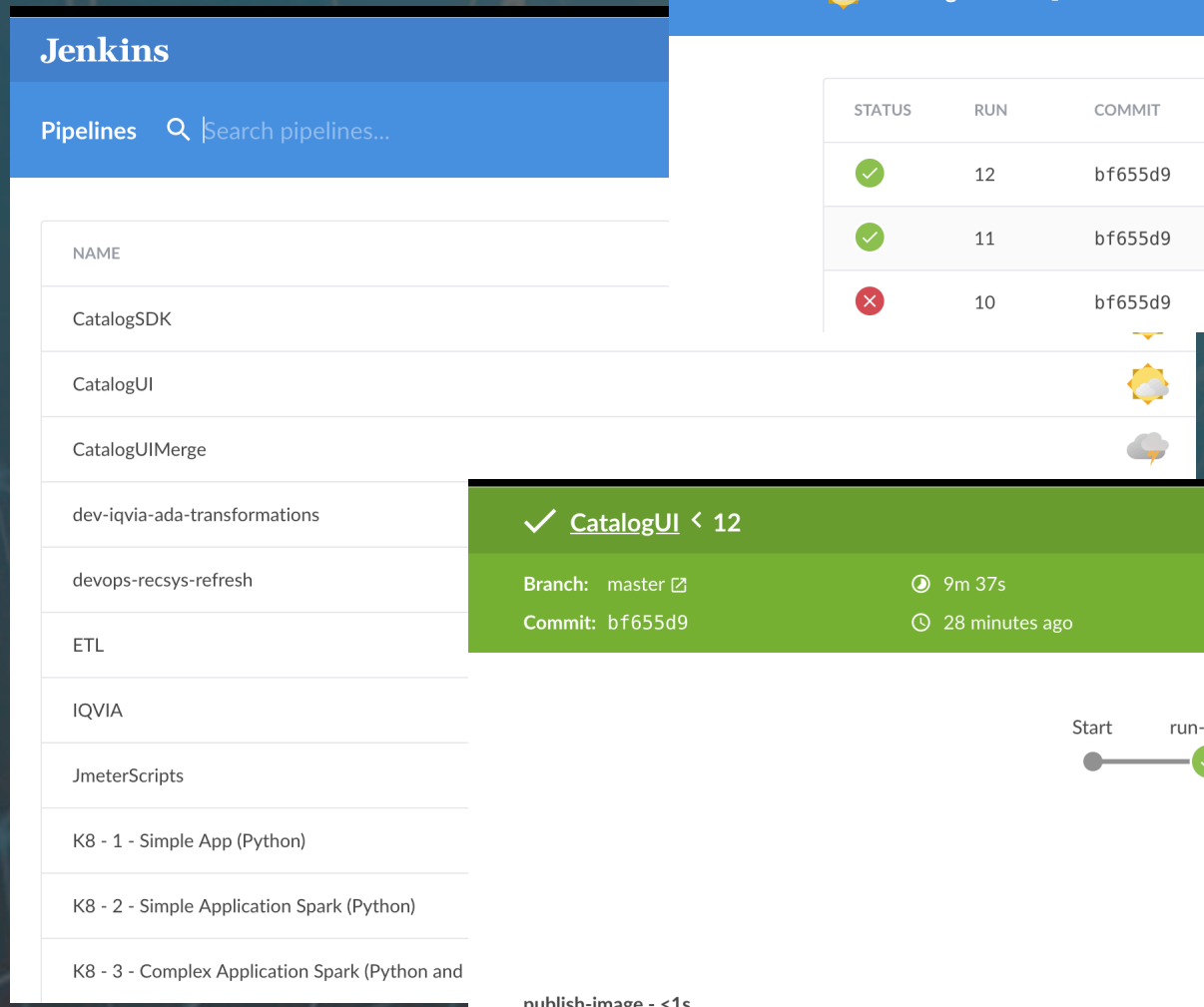
Jenkins

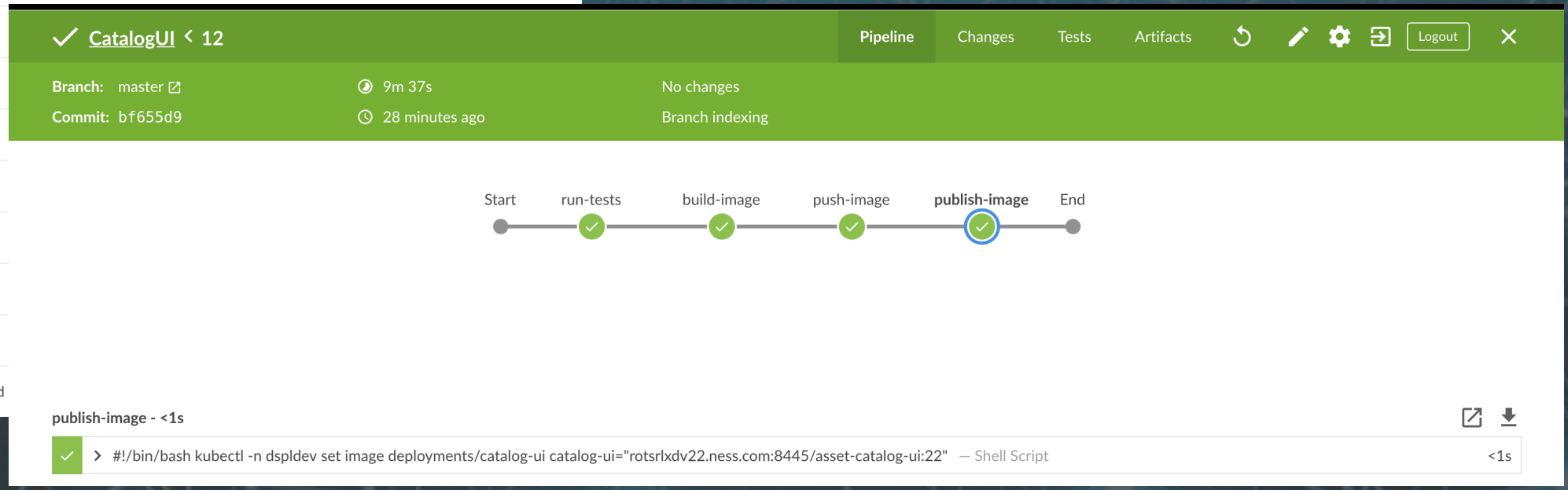
Open Blue Ocean

Ness Digital Engineering

NextGenML Platform



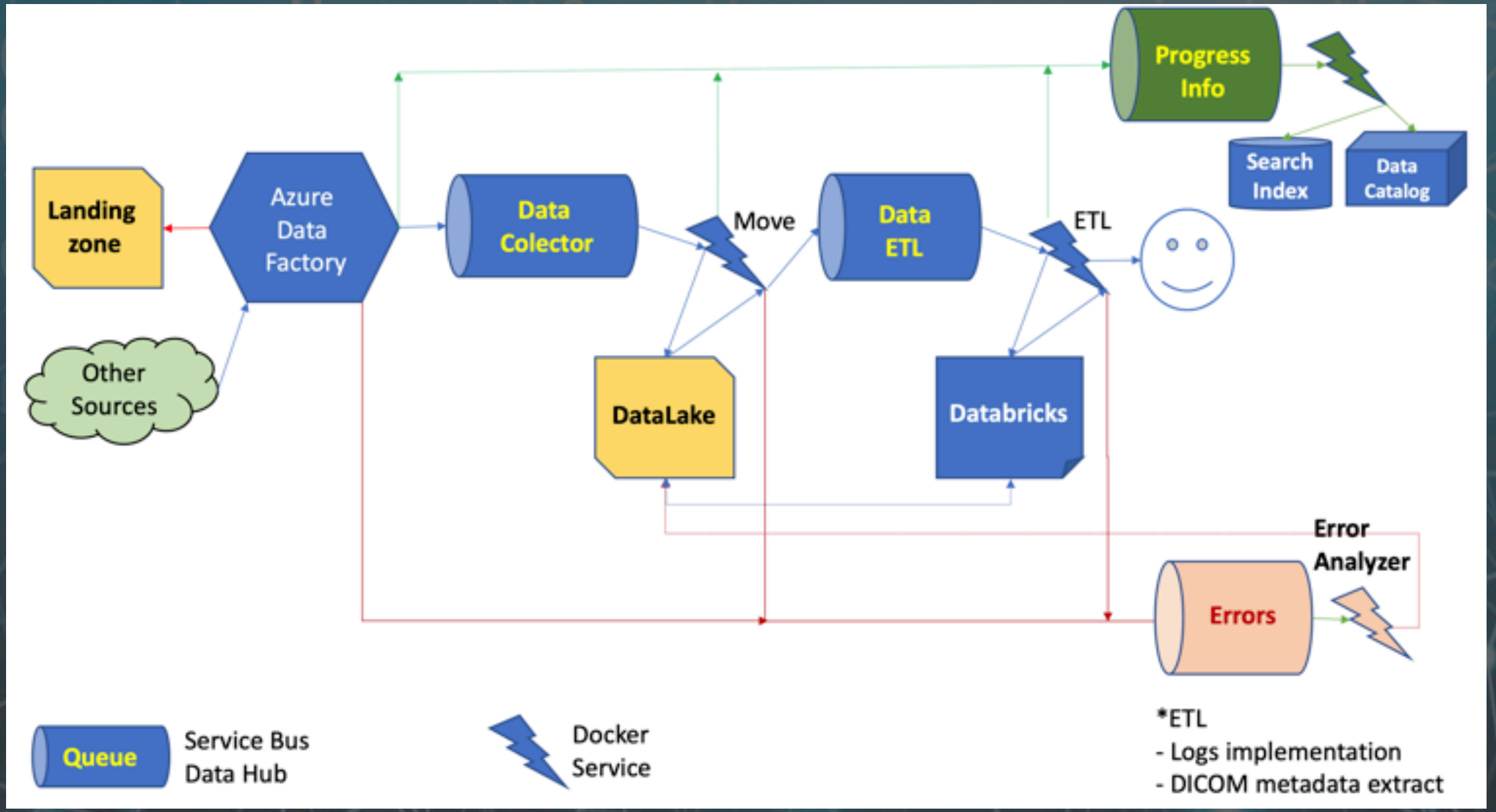




Hybrid Cloud – ETL solution

Ness Digital Engineering

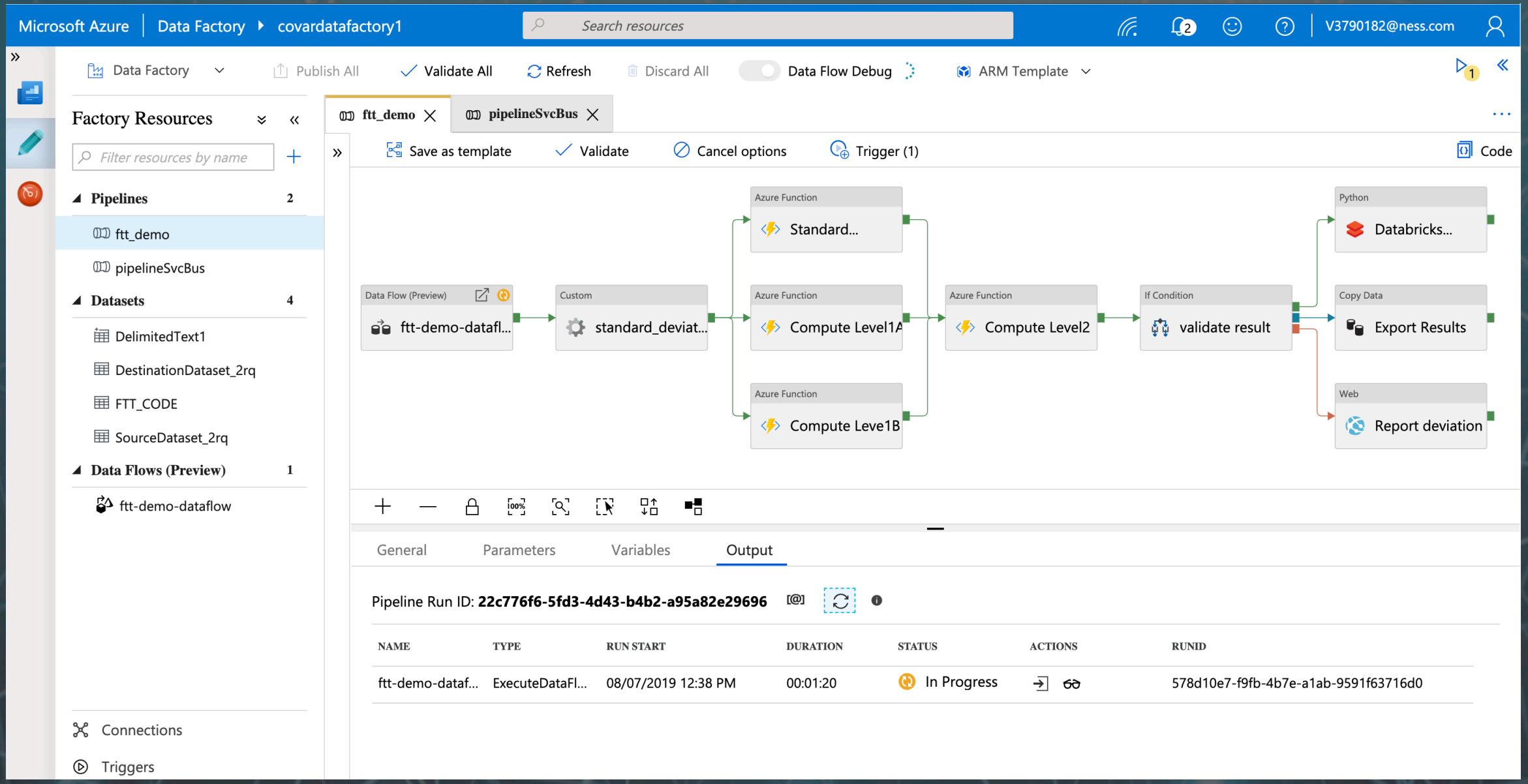
NextGenML Platform



Azure Data Factory

Ness Digital Engineering

NextGenML Platform



ABDOMEN – markers

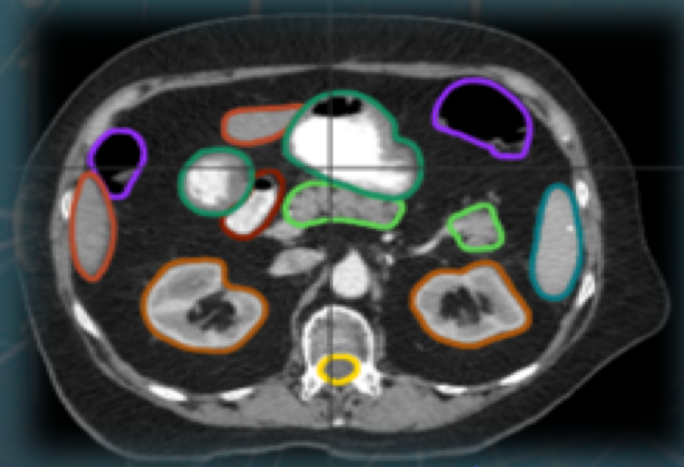
"Small Intestine", [200, 200, 100],
"Large Intestine", [255, 255, 0],

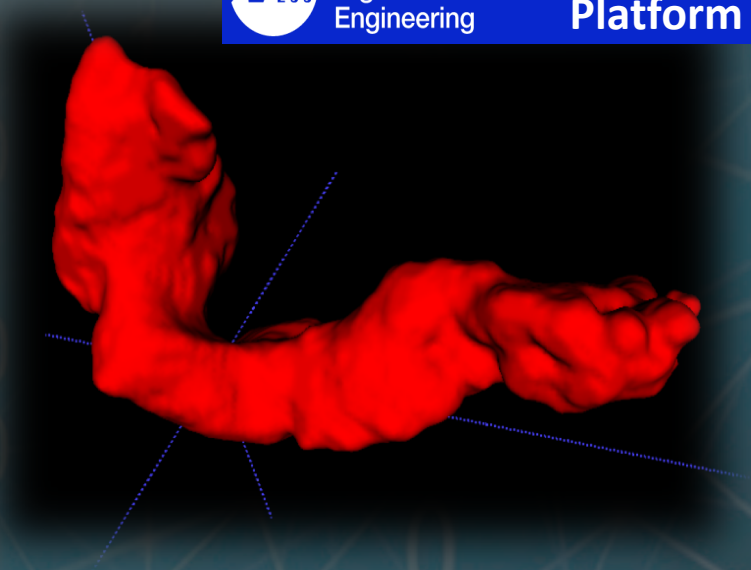
"Spinal Canal", [150, 100, 50],
"Gall Bladder", [40, 100, 240]

"Spleen", [0, 200, 120],

"Liver", [30, 200, 50],
"Stomach", [120, 80, 10],

"Pancreas", [200, 70, 30],
"Duodenum", [20, 100, 50],.





Ness Digital Engineering

NextGenML Platform

• Stack Technology

Platform

- Kubernetes (onPrem) + Docker
- Azure Kubernetes Cluster (AKS)
- Nexus
- Azure Container Registry(ACR)
- GlusterFS

Workflow

- Argo->Kubeflow

DevOps

- Helm
- kSonnet
- Kustomize
- Azure DevOps

Code Management & CI/CD

- Git
- TeamCity
- SonarQube
- Jenkins

Security

- MS Active Directory
- Azure VPN
- Dex (K8s) integrated with GitLab

Machine Learning

- TensorFlow (model training, boarding, serving)
- Keras
- Seldon

Storage (Azure)

- Storage Gen1 & Gen2
- Data Lake
- File Storage

ETL (Azure)

- Databricks
- Data Factory (ADF)
- HDInsight (Kafka and Spark)
- Service Bus (ASB)
- Lambda functions & VMs
- Spark on K8
- Cache for Redis

Monitoring and Logging

- Graphana
- Prometeus
- GrayLog

Thank you!

Radu.Moldovan@ness.com
Skype: r.moldovan
Linkedin: linkedin.com/in/raduadrianmoldovan

Ness Timisoara