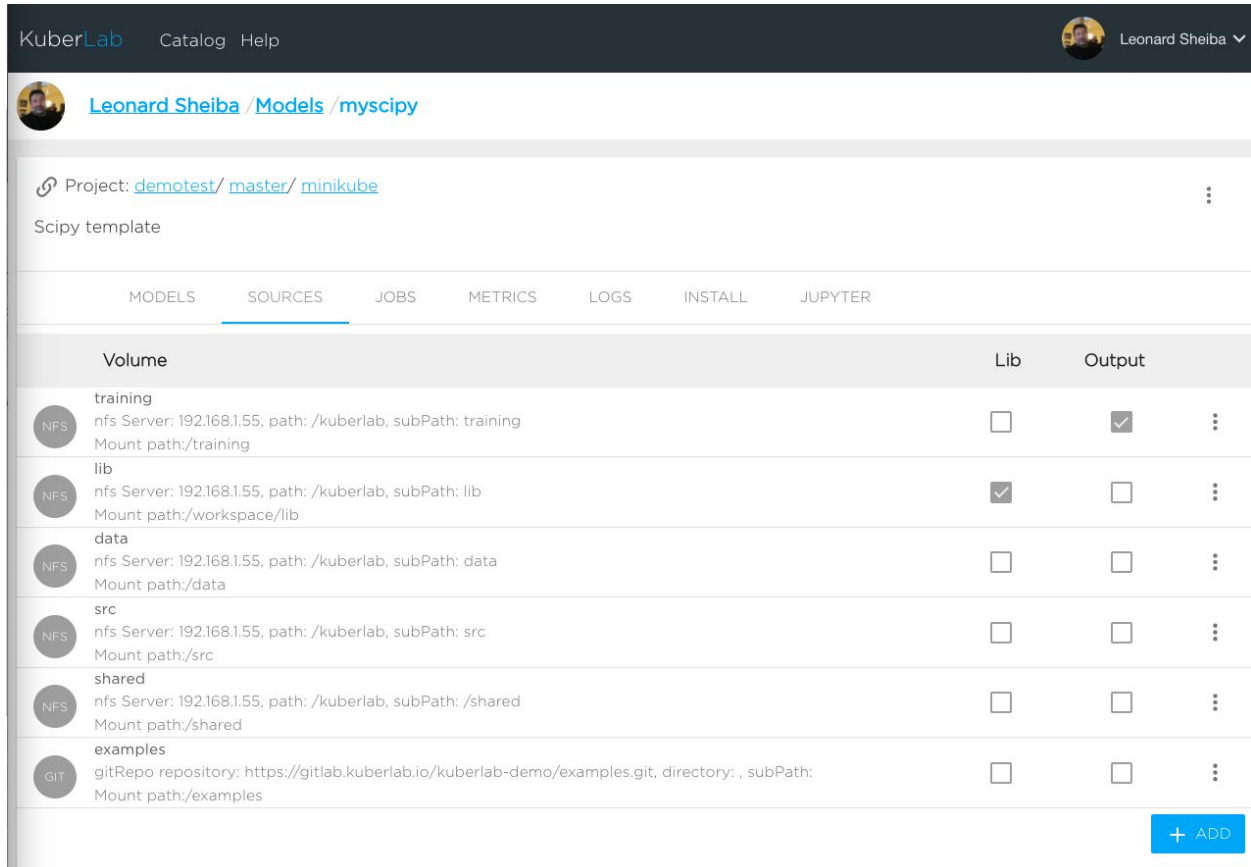


## Storage and Data Sources Tutorial

When model created from generic template it will be configured with default set of data sources. Let's go to the SOURCES tab and examine them in details.



The screenshot shows the KuberLab interface with the user 'Leonard Sheiba' logged in. The breadcrumb path is 'Leonard Sheiba / Models / myscipy'. The project is 'demotest / master / minikube' and it's a 'Scipy template'. The 'SOURCES' tab is active, showing a table of volumes:

Volume	Lib	Output
<b>training</b> nfs Server: 192.168.1.55, path: /kuberlab, subPath: training Mount path:/training	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<b>lib</b> nfs Server: 192.168.1.55, path: /kuberlab, subPath: lib Mount path:/workspace/lib	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>data</b> nfs Server: 192.168.1.55, path: /kuberlab, subPath: data Mount path:/data	<input type="checkbox"/>	<input type="checkbox"/>
<b>src</b> nfs Server: 192.168.1.55, path: /kuberlab, subPath: src Mount path:/src	<input type="checkbox"/>	<input type="checkbox"/>
<b>shared</b> nfs Server: 192.168.1.55, path: /kuberlab, subPath: /shared Mount path:/shared	<input type="checkbox"/>	<input type="checkbox"/>
<b>examples</b> gitRepo repository: https://gitlab.kuberlab.io/kuberlab-demo/examples.git, directory: , subPath: Mount path:/examples	<input type="checkbox"/>	<input type="checkbox"/>

A '+ ADD' button is located at the bottom right of the table.

On the SOURCES page we can see the list of Volumes. Each volume represent a datasource configured to be accessed as folder in file system accessible by the model code or by the tools like Jupyter or Tensorboard integrated into environment.

When model is created from the template, those volumes, preconfigured in the template, will be initialised. Some volumes are mandatory, some optional and specific to the template and user can add and configure custom volumes based on the available datasource types.

**Mandatory Volumes** - every volume has corresponding environment variable exposed throughout the model environment

1. **TRAINING** - \$TRAINING\_DIR

TRAINING volume (or training folder) is used to save training logs and checkpoints.

TensorBoard is looking at the folder of the training volume to look for data needed to

display analysis of the model training. Serving module will also use training folder as a source for the model to serve

## 2. LIB - \$LIB\_DIR

LIB volume (or lib folder) is used to install third party libraries and component during runtime. This functionality is exposed on INSTALL tab.

## 3. DATA - \$DATA\_DIR

DATA volume (or data folder) is where model will look for data by default. There are several ways to save the data to the data folder which we will describe later.

## 4. SRC - \$SRC\_DATA

SRC volume (or src folder) is the mapping of the model template git repository into the model storage space. The volume type is GIT and it is not persistent. In the future it will be possible to perform standard git operations on the data located in the GIT volume.

## 5. SHARED - TBD

### Custom Volumes





#### 1. EXAMPLES

EXAMPLES volume (or examples folder) is specific to the model template. In this case It is referencing git repository with example project. This volume has type GIT and is not persistent.

#### 2. USER volume is created by model developer. It can have any type of available storage resource.

### Supported Volume Types:

1. GIT - not persistent, map GIT repository into model environment
2. NFS - persistent, mount NFS shares into model environment
3. S3 - persistent, map S3 buckets into model environment
4. CLUSTER - persistent storage predefined as part of the cluster, require minimum or no configuration

Volume Type		Output folder	
Volume	Lib	Output	
 training nfs Server: 192.168.1.55, path: /kuberlab, subPath: training Mount path:/training	<input type="checkbox"/>	<input checked="" type="checkbox"/>	⋮
 src gitRepo repository: https://github.com/kuberlab-catalog/tensorflow, directory: , subPath: /tensorflow/src Mount path:/src	<input type="checkbox"/>	<input type="checkbox"/>	⋮
 lib nfs Server: 192.168.1.55, path: /kuberlab, subPath: lib Mount path:/workspace/lib	<input checked="" type="checkbox"/>	<input type="checkbox"/>	⋮
 data nfs Server: 192.168.1.55, path: /kuberlab, subPath: data Mount path:/data	<input type="checkbox"/>	<input type="checkbox"/>	⋮

### Volume Definition

Volume defined by the following attributes:

1. Name of the volume
2. SubPath defines the folder position in the filesystem layout of the storage device. If SubPath is started with leading "/" it is shared and will be defined relative to the shared folder of the storage device. Without leading "/" folder is private to the model and will be defined relative to the private model folder
3. MountPath defines the folder as it is seen by the model code and other components.
4. Volume type can be:

#### Cluster Storage volume:

This is the volume defined by the storage configured as part of the cluster. The configuration parameters are used from cluster configuration automatically.

Name *	Sub Path	MountPath *
data	data	/data
Type *		
Cluster storage		
Storage		
default	Server	path
<input type="checkbox"/> Train Log Dir <input type="checkbox"/> Library Dir		
		CANCEL SAVE

#### NFS volume

Name *	Sub Path	MountPath *
tutorials	/tutorials	/tutorials
Type *		
NFS		
server		
192.168.1.55	path	
	/kuberlab	
<input type="checkbox"/> Train Log Dir <input type="checkbox"/> Library Dir		
		CANCEL ADD

## S3 volume

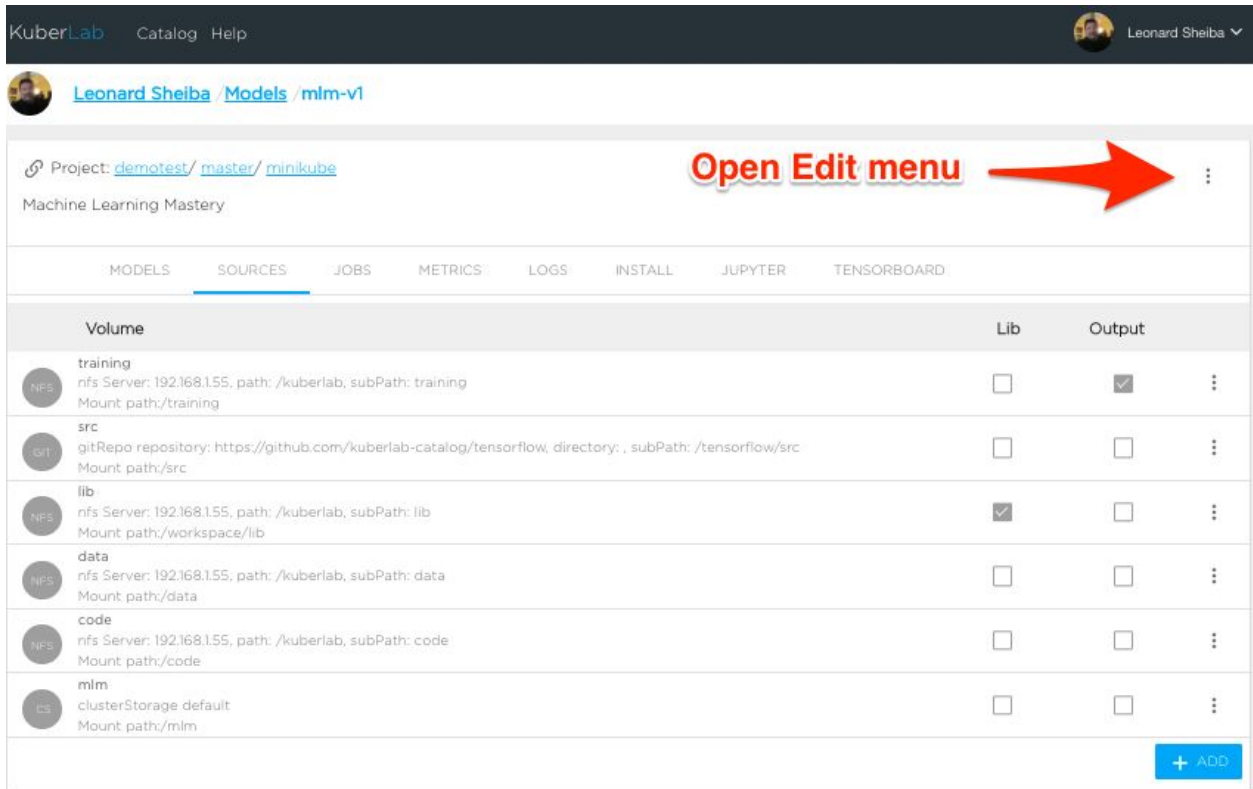
Name *	Sub Path	MountPath *
tutorials	/tutorials	/tutorials
Type *		
s3 bucket	<b>S3 bucket</b>	<b>S3 server to use</b>
<b>Cloud Account</b>		
Bucket *	Server	Account
<input type="checkbox"/> Train Log Dir	<input type="checkbox"/> Library Dir	
		CANCEL <b>ADD</b>

## GIT Volume

Name *	Sub Path	MountPath *
tensorflow-programs-and-tutorials	/Tensorflow-Programs-and-Tutorials	/tensorflow-programs-and-tutorials
Type *		
GIT		
Select repository	<b>Repository URL</b>	
Repository url	https://github.com/Isheiba/Tensorflow-Programs-and-Tutorials	
<input type="checkbox"/> Train Log Dir	<input type="checkbox"/> Library Dir	
		CANCEL <b>SAVE</b>

Volume mapping for Jupyter.

Jupyter require additional configuration which will map Volumes into Jupyter folder. To do that several steps need to be done.









KuberLab Catalog Help Leonard Sheiba

Leonard Sheiba / Models / mim-v1

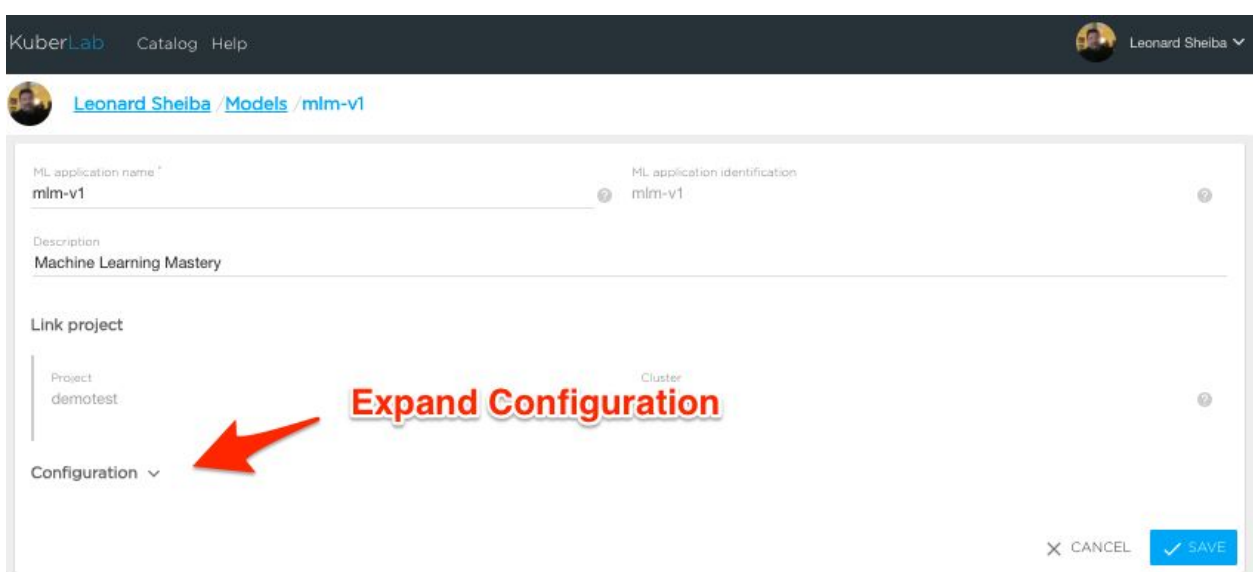
Project: demotest / master / minikube

Machine Learning Mastery

MODELS SOURCES JOBS METRICS LOGS INSTALL JUPYTER TENSORBOARD

Volume	Lib	Output
 training nfs Server: 192.168.1.55, path: /kuberlab, subPath: training Mount path:/training	<input type="checkbox"/>	<input checked="" type="checkbox"/>
 src gitRepo repository: https://github.com/kuberlab-catalog/tensorflow, directory: , subPath: /tensorflow/src Mount path:/src	<input type="checkbox"/>	<input type="checkbox"/>
 lib nfs Server: 192.168.1.55, path: /kuberlab, subPath: lib Mount path:/workspace/lib	<input checked="" type="checkbox"/>	<input type="checkbox"/>
 data nfs Server: 192.168.1.55, path: /kuberlab, subPath: data Mount path:/data	<input type="checkbox"/>	<input type="checkbox"/>
 code nfs Server: 192.168.1.55, path: /kuberlab, subPath: code Mount path:/code	<input type="checkbox"/>	<input type="checkbox"/>
 mim clusterStorage default Mount path:/mim	<input type="checkbox"/>	<input type="checkbox"/>

+ ADD



KuberLab Catalog Help Leonard Sheiba

Leonard Sheiba / Models / mim-v1

ML application name \* mim-v1

ML application identification mim-v1

Description Machine Learning Mastery

Link project

Project demotest

Cluster

Configuration

X CANCEL SAVE

Configuration ^

JUPYTER

TENSORBOARD

Resource jupyter

Execution directory

Timeout minutes

Execution command

Node allocator

Execution Arguments

Resources

CPU (m)	CPU Limit (m)
100m	1000m
Memory (Mi/Gi)	Memory Limit (Mi/Gi)
256Mi	8Gi
GPU	Replicas
1	0

Environment Arguments

name	value
------	-------

Advanced ^

Images

cpu	gpu
kuberlab/mlboard-jupyter:latest	kuberlab/mlboard-jupyter-gpu:latest

Ports

name	protocol	port	TargetPort
http	TCP	8888	8888

Volumes

Name	Sub path	Mount path	
lib		/notebooks/lib	
Name	Sub path	Mount path	
src		/notebooks/src	
Name	Sub path	Mount path	
training		/notebooks/training	
Name	Sub path	Mount path	
data		/notebooks/data	
Name	Sub path	Mount path	
code		/notebooks	
Name	Sub path	Mount path	
mlm		/notebooks/mlm	

APPLY

Save when finished

CANCEL

SAVE

In the volumes section of Jupyter configuration form select ADD volume

Volumes

Name	Sub path	Mount path	
lib		/notebooks/lib	
src		/notebooks/src	
training		/notebooks/training	
data		/notebooks/data	
code		/notebooks	

**Select ADD volume**

**+ ADD**

Select volume Name to use

Volumes

Name	Sub path	Mount path	
lib		/notebooks/lib	
src		/notebooks/src	
training		/notebooks/training	
data		/notebooks/data	
code		/notebooks	

**Select name**

Name Sub path Mount path **APPLY**

When you select Name, drop down menu will show available volume names to attach to

src

lib

data







code

**Select Volume**

mlm

Type the MountPath where to map new volume relative to the root folder “/notebooks”. If you type path “/notebooks/mlm”, when you open Jupyter notebook you will see “/mlm”

#### Volumes

Name	Sub path	Mount path	
lib		/notebooks/lib	
src		/notebooks/src	
training		/notebooks/training	
data		/notebooks/data	
code		/notebooks	
mlm	▼ Sub path	/notebooks/mlm	

**Selected volume** (arrow pointing to 'code' row)

**Path in Jupyter Notebook** (arrow pointing to '/notebooks' in the 'code' row)

**APPLY**

Afterwards select APPLY and SAVE configuration. Environment will be rebuilt and reloaded, which may take a little extra time.