

## Post Training Quantization

In [16]:

```
import tensorflow as tf
from tensorflow import lite
from tensorflow.keras.models import load_model
```

In [17]:

```
#ref : https://stackoverflow.com/questions/70044693/unable-to-convert-convlstm-to-tflite-model
import absl.logging
absl.logging.set_verbosity(absl.logging.ERROR)
filename = "model.h5"
model = load_model(filename)
tf.saved_model.save(model, "./saved_model")
```

INFO:tensorflow:Assets written to: ./saved\_model/assets

INFO:tensorflow:Assets written to: ./saved\_model/assets

In [18]:

```
from tensorflow import lite
converter = lite.TFLiteConverter.from_saved_model("./saved_model")
converter.target_spec.supported_ops = [tf.lite.OpsSet.TFLITE_BUILTINS, tf.lite.OpsSet.SELECT_TF_OPS]
converter._experimental_lower_tensor_list_ops = False
tflite_model = converter.convert()
```

In [19]:

```
print("Size of model in Bytes before quantization")
print(len(tflite_model))
```

Size of model in Bytes before quantization  
30545388

In [20]:

```
converter = tf.lite.TFLiteConverter.from_saved_model("./saved_model1/")
converter.optimizations = [tf.lite.Optimize.DEFAULT]
converter.target_spec.supported_ops = [tf.lite.OpsSet.TFLITE_BUILTINS, tf.lite.OpsSet.SELECT_TF_OPS]
converter._experimental_lower_tensor_list_ops = False
tflite_quant_model = converter.convert()
```

In [21]:

```
print("Size of Model after Quantization")
print(len(tflite_quant_model))
```

Size of Model after Quantization  
7647648