**Machine Learning Engineer : Task Challenge**

**Problem Statement**

You are provided with a dataset of ~5k 512x512 images, your program should accept an

512x512 input image and return N images from the provided dataset similar to the input image.

**Solution**:

## For this task, I have used a pre-trained ResNet50 network which is trained on the ImageNet database and transfer learn it to find the feature vector for the image in the database using Pytorch 1.0 and FastAI library.

The whole process is done in following steps:

1. Load the data using dataset loaders of Pytorch using FastAI library
2. Take a pre-trained network, in this case, a ResNet50 and remove it’s last fully connected layers
3. Add new fully connected layers at the end of the network and train only those layers using the image in dataset, while keeping all the other layers frozen
4. Train the entire network by unfreezing all the layers.

# Extracting image embeddings using Pytorch Hooks. This is done using The SaveFeatures class .It will save the intermediate computation in a numpy array which can be retrieved using SaveFeatures

1. Then LSH is used to create a hash for the given image and then compare the distance from image embedding of the pictures in the dataset which shares the same hash value.
2. Then we take the input from the user, get its embeddings using the trained network and finally query it with the LSH table to find N(user-defined) similar images.
3. KMeans and t-SNE from sklearn library are used to find the cluster of the similar image and visualize them on a plot.

**Note: The brief of the code is explained as comments in the notebook attached.**