

| | | | |
|---|---|---|---|
| 1. Data is downloaded and unzipped in cwd | 2. Clean Data Reduce to 28790 images. 8434, 8069,578, 3903, 7806 | 3. Data Preprocessing: Resize Images All images are resized to 256 by 256 pixels | 4. Data Preprocessing: Data Augmentation Augment Data for Class 2 and 3 to balance the dataset |
|---|---|---|---|

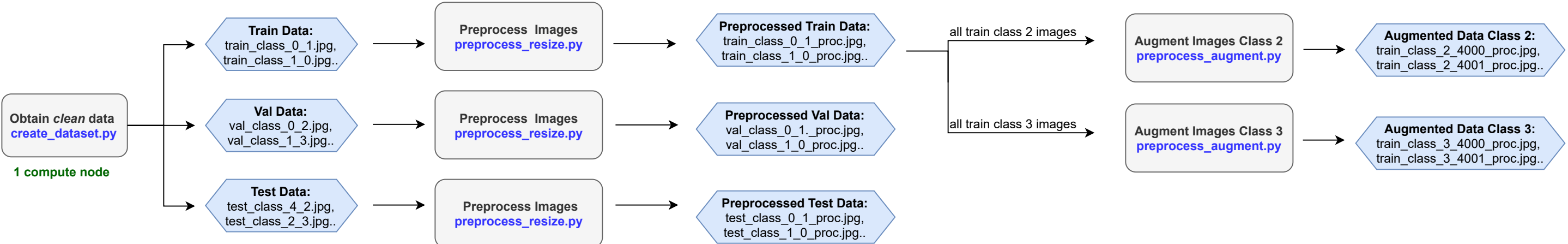
Download **galaxy-zoo-the-galaxy-challenge.zip**
File size: 1.9 GB



Pegasus
Downloads the data from an online repository (1.9 GB)

The images are assigned to classes based on their properties. Subset of the images is used as a dataset going forward. Total 28,793 images each between 8 kB to 16 kB. **Total: 386.0 MB**

Pegasus
launches multiple data preprocessing jobs in parallel.



N compute nodes in parallel

The data preprocessing jobs can be parallized based on number of available nodes. **Total Input: 386.0 MB**

| | | |
|---|---|---|
| 5. Hyperparameter Optimization Tune model to find the best setting (Optimizer, Activation) | 6. Train Model Use the best optimization parameters to train model | 7. Test Model classification metrics, confusion matrix, visualize what network learned |
|---|---|---|

Pegasus

