In 20B.yml

```
14      # parallelism settings ( you will want to change these based on your cluster setup, ideally scheduling pipeline stages
15      # across the node boundaries )
16      "pipe_parallel_size": 4,
17      "model_parallel_size": 2,
```

- Define parallel settings,in simple english

In train.py

```
from megatron.neox_arguments import NeoXArgs
from megatron.training import pretrain


def main(input_args=None, overwrite_values=None):
    neox_args = NeoXArgs.consume_neox_args(
        input_args=input_args, overwrite_values=overwrite_values
    )
    neox_args.configure_distributed_args()
    neox_args.build_tokenizer()  # tokenizer needs to be build in training in order to set the padding vocab
    neox_args.initialize_tensorboard_writer()  # is initialized if tensorboard directory is defined
    neox_args.initialize_comet()  # is initialized if comet directory is defined
    pretrain(neox_args=neox_args)
```

- Import pretrain from megatron.training and call pretrain in the main.

In megatron.training.py

```
219     def pretrain(neox_args):
240
241         # Initialize and get arguments, timers, and Tensorboard writer.
242         initialize_megatron(neox_args=neox_args)
243
244         # Create data loaders
245         timers("train/valid/test data loaders").start()
246         data_loaders = build_train_valid_test_data_loaders(neox_args=neox_args)
247         update_iterations(neox_args=neox_args, data_loaders=data_loaders)
248         timers("train/valid/test data loaders").stop()
249
250         # Model, optimizer, and learning rate.
251         timers("model and optimizer").start()
252         model, optimizer, lr_scheduler, reference_model = setup_model_and_optimizer(
253             neox_args=neox_args, use_cache=False, iteration=neox_args.iteration
254         )
255         timers("model and optimizer").stop()
256
257         if neox_args.serve_model_weights:
258             start_server(model)
259             # sync...
260             torch.distributed.barrier()
```

- Call setup_model_and_optimizer

In setup_model_and_optimizer

```
1233    def setup_model_and_optimizer(neox_args, use_cache=False, iteration=None):
1276            _lr_scheduler = lr_scheduler
1277
1278            model, optimizer, _, lr_scheduler = deepspeed.initialize(
1279                model=model,
1280                optimizer=optimizer,
1281                args=neox_args,
1282                lr_scheduler=_lr_scheduler,
1283                dist_init_required=False,
1284                model_parameters=_model_params,
1285                # Need to remove the below so that it doesn't conflict with --deepspeed_config required by autotuning
1286                # config_params=neox_args.deepspeed_config,
1287                mpu=mpu if not neox_args.is_pipe_parallel else None,
1288            )
1289            if needs_reference_model:
1290                reference_model, _, _, _ = deepspeed.initialize(
1291                    model=reference_model,
1292                    optimizer=ref_optimizer,
1293                    args=neox_args,
1294                    lr_scheduler=ref_lr_scheduler,
1295                    dist_init_required=False,
1296                    model_parameters=ref_param_groups,
1297                    mpu=mpu if not neox_args.is_pipe_parallel else None,
1298                )
```

- deepspeed.initialize is called and given the parameters defined in whatever config is being used (in our case 20B.yml).