1. Lệnh thực thi:

spark-submit \

--master spark://172.20.201.152:7077 \

--archives environment.tar.gz \

--driver-cores 1 \

--driver-memory 4g \

--num-executors 1 \ #6 với 6 node, 12 với 12 node

--executor-cores 2 \

--executor-memory 4g \

--jars \

/home/ubuntu/anaconda3/envs/py37/lib/python3.7/site-packages/bigdl/share/core/lib/all-2.4.0-20230420.050641-1.jar,\

/home/ubuntu/anaconda3/envs/py37/lib/python3.7/site-packages/bigdl/share/dllib/lib/bigdl-dllib-spark\_2.4.6-2.4.0-jar-with-dependencies.jar,\

/home/ubuntu/anaconda3/envs/py37/lib/python3.7/site-packages/bigdl/share/orca/lib/bigdl-orca-spark\_2.4.6-2.4.0-jar-with-dependencies.jar \

--conf spark.executorEnv.PYSPARK\_PYTHON=environment/bin/python \

--conf spark.executorEnv.PYTHONPATH=environment \

--conf spark.scheduler.minRegisteredResourcesRatio=1.0 \

--conf spark.executor.extraClassPath=environment/lib/python3.7/site-packages/bigdl/share/core/lib/all-2.4.0-20230420.050641-1.jar:\

environment/lib/python3.7/site-packages/bigdl/share/dllib/lib/bigdl-dllib-spark\_2.4.6-2.4.0-jar-with-dependencies.jar:\

environment/lib/python3.7/site-packages/bigdl/share/orca/lib/bigdl-orca-spark\_2.4.6-2.4.0-jar-with-dependencies.jar \

--conf spark.driver.extraClassPath=environment/lib/python3.7/site-packages/bigdl/share/core/lib/all-2.4.0-20230420.050641-1.jar:\

environment/lib/python3.7/site-packages/bigdl/share/dllib/lib/bigdl-dllib-spark\_2.4.6-2.4.0-jar-with-dependencies.jar:\

environment/lib/python3.7/site-packages/bigdl/share/orca/lib/bigdl-orca-spark\_2.4.6-2.4.0-jar-with-dependencies.jar \

--conf spark.driverEnv.PYSPARK\_PYTHON=environment/bin/python \

--conf spark.driverEnv.PYTHONPATH=environment \

--conf spark.executorEnv.RAY\_START\_WAIT\_TIMEOUT=600 \

main.py

2. Tập dữ liệu Private, sử dụng file main.py

3. Tập dữ liệu VinDr, sử dụng file main\_mvindr.py

4. Trong hàm main:

def main():

    dataset = "hdfs://172.20.201.155:9000/user/ubuntu/dataset/PRIVATE-BREAST-MAMMO-VN"

    model\_name = "xceptionfuzzyFC"  # Tên mô hình (Tương ứng là tên file trong folder /src)

    app\_name = "SparkBreastCancerClassifier"

Trong biến model\_name, đổi tên tương ứng với mô hình sử dụng (cũng chính là tên file cấu hình trong folder /src)

5. Cấu trúc file /src/model\_name.py

class Fuzzy: *(nếu cần)*

….

def build\_{model\_name}\_model (…):

….

def train\_{model\_name} (…):

model = build\_{model\_name}\_model( )

….