1. Every cloud-side program produces a log file of the same name as itself.

CEPII\_GeoDist.csv,EIA.csv.bz2

HK\_#Y\_rx.csv.bz2, Swiss\_mx\_71-72.csv

%-comtrade/ #Y.csv.bz2

%-work/ cifob\_model.rds.bz2

%-mirror-analysis/@-#Y-M\_matched.csv.bz2

%-comtrade/ mfob.csv

%-work/ @-Y#-gaps.csv.bz2

/efs/work/ mm|o|l\_m|x\_k2T|F.csv.bz2

/efs/work/ tm\_k2T|F.csv.bz2

/efs/work/ trf.csv

./data/ rl#cty.csv

/efs/unct/ @-#Y-input.csv.bz2

bulk\_download.R

Bulk download annual trade data from Comtrade. Remove redundant data. Update download info.

mirror\_match.R

Preprocessing, pair import and export declarations; separate matched, orphaned, lost

input\_adj.R

Subset matched data and add attributes for building and applying the model.

gaps.R

Applying the model to country data and calculating weights to calculate TM. Input: years=all

tm\_cty\_agg.R

Identify flows, aggregate commodity. input: cty, all\_trade=T&F, years=years.

%-work/ summ\_prd.csv, gaps.log

bridge.csv

cifob\_model.R

Using global data to build the model that estimates the gap between CIF and FOB.

revenue\_loss.R

add tariff, aggregate commodities. no change of k.

%-comtrade/ download.log

bulk\_download.log

Updated Years

> 2

mirror\_match-node.R

<= 2

%-mirror-analysis/ @-#Y-M|X-junked-HK-adj.csv.bz2

%-mirror-analysis/ mirror\_match.csv, mirror\_match.log

prepd.sh

pub\_note.R

%-comtrade/ exr.csv

# of Countries

./data/ trains.csv, wto.csv

else: cty=null

Clean tariff rate data from WITS.

trf\_effictive.R

%-mirror-analysis/ @-#Y-M\_orphaned|lost.csv.bz2

mol\_cty\_agg.R

Identify import & export, aggregate commodities. input: k=2, all\_trade=T&F, cty, years=years.

gaps\_tm.R

gaps+tm. Input: cty=cty, k=6

Download and clean Comtrade publication note; extract exchange rates and fob info.

cty <55

/efs/work/ tm\_k6T|F.csv.bz2

k=2

k=6

2. To launch multiple computing units, run ‘spark-master.sh’ in the master node and ‘spark-worker.sh’ in the worker nodes. 3. For single node program, attach swap memory and run ‘spacious.sh’.

#: a number, #Y: year

%: gfi, @: Comtrade, a|b: a or b

./: folder of the program

files in ~/UN-Comtrade/ , files in ~/WB-WITS/

Data (Stored on S3)

Program (multiple computing units)

Program (single computing unit)

Data (Stored locally or on EFS)

By-product Outputs

Another Process

IFFs worldwide reports

IFFs national reports

implicit inputs