

Instructions & Guidelines

- You are free to choose any mainstream language (Java/Python/Scala)
- We expect you to attempt this assignment with honesty and integrity
- **Avoid** using any 3rd party library for the main logic
- Pre-inform us if you are delaying your submission
- You can zip your solution and email it back. Also explain in a small write up
 - 1) Your approach, code & output
 - 2) Further scope of improvements which you can think of for now
 - 3) Details about 3rd party libraries or dictionaries used

Assignment

Question1:

Given file (india_phoneno) contains 100,000 records. Each record contains **PhoneNo, City, State, Country**. Task is to find city codes for cities given in the file using the phone numbers present in file.

For example

City	State	Country	City code	#Records	
Pune	Maharashtra	India	020	10	
Mumbai	Maharashtra	India	022	20	... Example 1

Question2:

Given file (as_tweets.txt) contains 100,000 tweets about the movie “**American Sniper**”. These are unprocessed real tweets. As you may know, the movie was directed by **Clint Eastwood**, starring **Bradley Cooper** and was based on the life story of **Chris Kyle**.

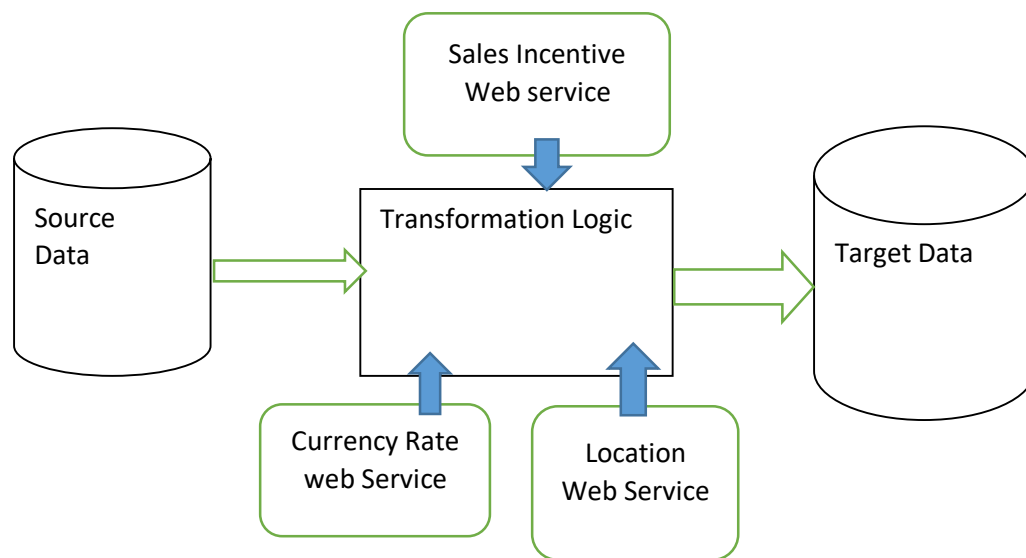
First task is to break the twitter hashtags into proper words, for example:

#goldenglobes -> golden globes,

#siennamiller -> sienna miller ... Example 2

PS: you may need to download any public English dictionary of words

Question3: Please refer following architecture diagram & sample tables with dummy data for a sales application



Sample source data: (Monthly Sales data):

Sales ID- Region+ Country Code+ City Code

Salesperson ID	Region	Country	City	Sales Person Level Region wise	Sales Amt-Base Country	Month & Year
APACINDMUM0001	APAC	INDIA	MUMBAI	APAC2	65000INR	May-17
NACANTOR0001	NA	CANADA	TORONTO	NA2	500CAD	May-17
APACINDPUN0001	APAC	IND	PUNE	APAC1	130000 INR	May-17
NACANTOR0001	NA	CANADA	TOR	NA1	1000CAD	May-17

Currency Conversion Service provides exchange rates against USD

Currency Pair	Exchange Rate
USDEUR	1.3
USDINR	64.18

Location Web service:

City Name	City Abbreviations	Country Name	Country Abbreviations	REGION
MUMBAI	MUM	INDIA	IND	APAC
BOMBAY	MUM	INDIA	IND	APAC

POONA	PUN	INDIA	IND	APAC
PUNE	PUN	INDIA	IND	APAC
TORONTO	TOR	CANADA	CAN	NA
PARIS	PAR	FRANCE	FRN	EMEA

Sales Incentive Web service provides region wise incentives calculation formula

REGION	Sales Person Level Region wise	Sales Incentive formula
APAC	APAC1	(20% of Sales Amt Base Currency or 20USD whichever is less)
APAC	APAC2	(30% of Sales Amt Base Currency or 30 USD whichever is less)
NA	NA1	(2% of Sales Amt USD or 20USD whichever is less)
NA	NA2	(3% of Sales Amt USD or 30USD whichever is less)

Target data : Salesperson Table Example 3

Salespers on ID	Regi on	Coun try	City	Sales Amt- Base Curren cy	Sales Amt- USD	Glob al Ran k	Region Rank	Countr y Rank	Sales Incentive	Month
APACIND MUM000 1	APA C	INDIA	MUM BAI	65000I NR	1000U SD	G10 0	APAC2 0	IND5	3250INR	May-17
NACANT OR0001	NA	CANA DA	TORO NTO	500CA D	365US D	G12 5	NA45	CAN20		May-17
APACIND PUN0001	APA C	INDIA	PUNE	13000 0 INR	2000 USD	G50	APAC2 5	IND3		May-17
NACANT OR0001	NA	CANA DA	TORO NTO	1000C AD	730 USD	G60	NA30	CAN10		May-17

Transformation Logic performs following:

- Converts Sales Amt in base country currency to Sales Amt USD.
- Calculates Global Rank based on USD Sales Amt.
- Location web service provides Lookup of city abbreviations where Transformation logic looks for the city abbreviations and matches CITY Name.

- d. Sales Incentive web service provides incentive formula based on sales person region wise, Transformation logic calculates incentive amount based on the REGION, Sales Person Level region wise.
- e. Also Transformation logic calculates region wise Ranks based on the Total sales happened for that Month.

Task 1) Create Dummy datasets for each table and save them to different files. Write spark(java/scala) program to do mentioned transformations and generate the final output for dummy datasets you created

Task2) Write unit test cases for application

Question 4: What is one thing you did in last year but may have done in better ways