Datastage Question Set 2

Q.1 Consider a Data with two Columns, say CustomerName(Varchar) and Deadline(Date or DateTime). Now the requirement is to check the input record, if the input date is less than currentdate, then increase the Deadline by 3 months(or 90 Days), until it is more than current date.

Output all this records with all steps of deadline.

	CustomerNumber	DeadLine
1	19	1954-04-19 00:00:00.000
2	37	1950-01-20 00:00:00.000
3	47	1950-11-01 00:00:00.000
4	49	1955-01-19 00:00:00.000
5	64	1958-06-10 00:00:00.000
6	80	1961-01-01 00:00:00.000
7	84	1960-01-01 00:00:00.000
8	88	1953-01-01 00:00:00.000
9	1664	1953-04-06 00:00:00.000
10	1701	1956-06-01 00:00:00.000
11	1702	1953-12-14 00:00:00.000
12	1708	1958-03-06 00:00:00.000
13	1712	1958-04-06 00:00:00.000

(Image:Q.1 Input)

	CustomerNumber	DeadLine	NewDeadLine
237	19	1954-04-19 00:00:00.0000000	2012-09-11 00:00:00.0000000
238	19	1954-04-19 00:00:00.0000000	2012-12-10 00:00:00.0000000
239	19	1954-04-19 00:00:00.0000000	2013-03-10 00:00:00.0000000
240	19	1954-04-19 00:00:00.0000000	2013-06-08 00:00:00.0000000
241	19	1954-04-19 00:00:00.0000000	2013-09-06 00:00:00.0000000
242	19	1954-04-19 00:00:00.0000000	2013-12-05 00:00:00.0000000
243	19	1954-04-19 00:00:00.0000000	2014-03-05 00:00:00.0000000
244	19	1954-04-19 00:00:00.0000000	2014-06-03 00:00:00.0000000
245	37	1950-01-20 00:00:00.0000000	2014-02-13 00:00:00.0000000
246	37	1950-01-20 00:00:00.0000000	2014-05-14 00:00:00.0000000
247	37	1950-01-20 00:00:00.0000000	1988-09-27 00:00:00.0000000
248	37	1950-01-20 00:00:00.0000000	2007-12-17 00:00:00.0000000
249	37	1950-01-20 00:00:00.0000000	2001-07-21 00:00:00.0000000

(Image:Q.1 Output)

Q.2 There are two Input Files(InputFile1 and InputFile2), each with 1 column(say '*A*' and '*B*')[both int] which may contain any number without sorting. The task is to separate these records into three files. First file will contain all the records which exists in both the files, second file will contain all the records which are unique and present in InputFile1 and third file will contain all the records which are unique and present in InputFile2.

Sample:

InputFile1(*A*)	InputFile2(*B*)
1	**5**
3	**6**
5	**7**
2	**4**
4	**8**

OutputFile1(*A*)	OutputFile2(*A*)	OutputFile3(*A*)
4	**1**	**6**
5	**2**	**7**
	3	**8**

Q.3 Candidates are filling forms in some xyz institute, they are submitting photocopy of their documents, but the documents they submit are not all the same, some may submit 1 document some may submit 3 and so on. You have the list of all the documents that have possibility of submission by the candidate. Now the job is to find out which documents the candidate has not submitted based on only two inputs, List of candidates and their submitted documents and list of all documents.

Sample:

List of Documents(*Docs*)	List of Candidates and their Documents (*CandidateID, Document*)
a	**1,c**
b	**1,e**
c	**2,a**
d	**2,d**
e	**2,e**

Q.4 There are two columns *A* and *B*, the requirement is to reverse the second column with first column intact, the first column is not sorted and the its sequence should not change.(Please refer to sample given below)

InputFile(*A, B*)	OutputFile(*A, B*)
4, a	**4, e**
2, b	**2, d**
5, c	**5, c**
1, d	**1, b**
3, e	**3, a**

Q.5 You have list of company and their offices in in two columns(*CompanyName, City*). The requirement is to group with *CompanyName* and get the city count and concatenate the city names in 1 string.

Sample:

InputFile(*CompanyName, City*)	OutputFile(*CompanyName,AggregatedData*)
IBM, Pune	**IBM, "Pune, Hyd, 2"**
IBM, Hyd	**Infosys, "Bngl, Hyd, Mysore, Pune, 4"**
Infosys, Bngl	**GGK, "JH, Uppal, Madhapur, 3"**
Infosys, Hyd	
Infosys, Mysore	
Infosys, Pune	
GGK, JH	
GGK, Uppal	
GGK, Madhapur	

Q.6 There are two columns(*ID, Flag*) the *ID* can have any flag, the *ID* may repeat many times, the requirement is that if the Flag for some *ID*s change, then those records should be dropped (Check the sample for explanation).

Sample:

InputFile(*ID, Flag*)	OutputFile(*ID, Flag*)
101,1	**101,1**
101,1	**101,1**
102,0	**103,2**
102,0	**104,2**
102,1	**104,2**
102,0	
103,2	
104,2	
104,2	
105,1	
105,2	

Q.7 There are three column(*ID, Flag1, Flag2*) Flag1 and Flag2 can have values 'Y' or 'N'

The *ID*'s with both flag as 'N' should be dropped.

InputFile(*ID, Flag1, Flag2*)	OutputFile(*ID, Flag1, Flag2*)
100,Y,Y	**100,Y,Y**
101,N,Y	**101,N,Y**
102,Y,N	**102,Y,N**
103,N,N	**104,Y,N**
104,Y,N	**105,N,Y**
105,N,Y	

Q.8 Consider two columns(*Dept and EmployeeID*), Design a job which gives out first and last record in each Dept. If there is 1 single ID in Dept., then i.) it should give out 2 rows for input record, 1 for each first and last. ii.) it should give out 1 row for input record, 1 for both.

Q.9 Design a mapping to give out following output data from input as shown in sample.

InputFile(*Col*)	OutputFile1(*Col*)	OutputFile2(*Col*)	OutputFile3(*Col*)
1	**1**	**4**	**7**
2	**2**	**5**	**8**
3	**3**	**6**	**9**
4	**10**	**13**	**16**
5	**11**	**14**	
6	**12**	**15**	
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

…and so on…

Q.10 Input column (*ColumnA*) is a string type, consisting an ID(*int*) column and LoginID(*Varchar*) concatenated, the job is to separate the two columns, but problem is the length both the column is not fixed (refer the sample). Design the job to achieve this requirement with input as shown in sample.

Sample:

InputFile(*ColumnA*)	OutputFile(*ID, LoginID*)
1kei62	**1,kei62**
23John	**23,John**
2345234Scott99	**2345234,Scott99**
0568kumar0	**0568,kumar0**

Q.11 A retail store has records of purchases of item by the customer, A customer can purchase more than 1 quantity of any product, in this case a column *QTY* will store the quantity of product. Now due to some reason they want to store details of all items separately, i.e if an item is purchase with *QTY* 5, the record should be repeated 5 times with *QTY* 1. Refer to sample for clarification.

InputFile(*CustID,SeqID,ItemID,QTY*)	OutputFile(*CustID,SeqID,ItemID,Qty,ProdSeq*)
101,1,214,3	**101,1,214,1,1**
101,2,235,3	**101,1,214,1,2**
102,1,607,2	**101,1,214,1,3**
	101,2,235,1,1
	101,2,235,1,2
	101,2,235,1,3
	102,1,607,1,1
	102,1,607,1,2

Q.12 Design a job to get the output as shown below with the given input. The number of alphabet rows and numeric rows are not fixed.

InputFile(*Col*)	OutputFile(*Col*)
A	**A,B,C,D,E,1**
B	**A,B,C,D,E,2**
C	**A,B,C,D,E,3**
D	
E	
1	
2	
3	

Q.13 There is a column having both numbers and alphabet, the requirement is to separate all the numbers in one column and alphabet in other column.

InputFile(*Col*)	OutputFile(*Alpha*,*Number*)
R2h87373r7	**Rhr,2873737**
E128h98e3	**Ehe,128983**
2de9083eyh	**deeyh,29083**
Ue283ur092	**Ueur,283092**

Q.14 Design a job for Input and Output as shown below

InputFile(*ColA*)	OutputFile(*ColA*)
aabbbcccc	**a1a2b1b2b3c1c2c3c4c5**
eeeeqiii	**e1e2e3e4q1i1i2i3**
jjjqoeee	**i1i2i3q1o1e1e2e3**

Q.15 There are two columns(*SrNo, Record*) the Record column has a data in random fashion but the *SrNo* has the serial number in a proper sequence, and we need to maintain this *SrNo* with the *Record*, but we also need the know what is the occurrence of the *Record* if it is or it is not repeated (Refer the sample for clarification).

InputFile(*SrNo,Record*)	OutputFile(*SrNo, Record, Occurance*)
1,b	**2,a,1**
2,a	**3,a,2**
3,a	**1,b,1**
4,b	**4,b,2**
5,b	**5,b,3**
6,b	**6,b,4**
7,s	**9,f,1**
8,w	**7,s,1**
9,f	**8,w,1**

Q.16 Consider the input as shown in Sample and design a job to give out output again as shown in Sample.

InputFile(*ColA*)	OutputFile(*ColA*)
a	**a**
a	**1**
b	**a**
b	**2**
b	**b**
c	**1**
c	**b**
c	**2**
c	**b**
	3
	c
	1
	c
	2
	c
	3
	c
	4