Q1 a)
Consider the example from the question:

src	tgt	weight
117	51	1
194	51	1
299	51	3
230	151	51
194	151	79
51	130	10

This is the input to the MAP function. To illustrate, in the first iteration, we have 117 51 1. This is converted to a string and split at the tab space. We map the target node to the weight such that output of the MAP function will be:

KEY	VALUE
51	1
51	1
51	3
151	51
151	79
130	10

This is the input to the Reducer function. The reducer function combines all values from the same KEY. Our reducer function specifically reduces entries with same KEY to one entry and maps it to the maximum VALUE for that KEY. The output of the reducer function will be,

KEY	VALUE
51	3
151	79
130	10

Q1 b)
Step 1: In java, divide the single dataset(inputFile) into separate datasets for Student records (<Department_ID, Name>) and Department records (<Department_ID, Department_Name>). This is the input to the Map phase.

STUDENT	DEPARTMENT
[Department_ID, Name]	[Department_ID, Department_Name]
1234, Alice 1234, Bob 1123, Joe	1123, CSE 1234, CS

Step 2: Create a custom writable (TaggedKey class in the article). The map output key will be a composite key that will belong to this class.

Step 3: The composite key is made up of the join key (in this case Department_ID) and an attribute 'tag' which tags the identity of the data (1=name, 2=department name).

STUDENT	DEPARTMENT
Composite key: [Department_ID, 1], Value: [Name]	Composite key: [Department_ID, 2] Value: [Department_Name]
[1234, 1], Alice [1234, 1], Bob [1123, 1], Joe	[1123, 2], CSE [1234, 2], CS

Step 3: Partition the data on key Department_ID by creating a custom partitioner class.

Step 4: Create a custom sorting comparator class to sort data first on Department_ID and then on tag.

Step 5: Group data based on natural key using custom grouping comparator class. Input to reducer class:

KEY	VALUE
[Department_ID]	{[Name],[Department_Name]}
1123 1234 1234	[Joe, CSE] [Alice, CS] [Bob, CSE]

Step 6: Iterate through the values for a key(Department_ID) and complete the join for student Name and Department_Name. Output of reducer is of the form:

1123, Joe, CSE 1234, Alice, CS 1234, Bob, CS