
TASK 2

In order to join the two files on a common key, 3 rounds of map-reduce will be required.

The two files employee_names.csv and employee_pay.csv before giving as input to map-reduce are as follows:

employee_names			employee_pay		
id	first_name	last_name	id	salary	bonus
59ea7840fc13ae1f6d000096	Benny	Ventom	59ea7840fc13ae1f6d000096	\$84217.20	
59ea7840fc13ae1f6d000097	Cara	Motherwell	59ea7840fc13ae1f6d000097	\$54737.84	
59ea7840fc13ae1f6d000098	Willem	Haresign	59ea7840fc13ae1f6d000098	\$92092.21	
59ea7840fc13ae1f6d000099	Trish	Farlane	59ea7840fc13ae1f6d000099	\$86117.63	
59ea7840fc13ae1f6d00009a	Camilla	Limpkin	59ea7840fc13ae1f6d00009a	\$92821.29	
59ea7840fc13ae1f6d00009b	Godwin	Caffrey	59ea7840fc13ae1f6d00009b	\$60633.55	\$1779.07
59ea7840fc13ae1f6d00009c	Elvera	Custed	59ea7840fc13ae1f6d00009c	\$94483.80	\$4328.59
59ea7840fc13ae1f6d00009d	Vinson	Farres	59ea7840fc13ae1f6d00009d	\$93127.77	
59ea7840fc13ae1f6d00009e	Kliment	Pitchford	59ea7840fc13ae1f6d00009e	\$90581.06	
59ea7840fc13ae1f6d00009f	Matty	Heater	59ea7840fc13ae1f6d00009f	\$86947.54	
59ea7840fc13ae1f6d0000a0	Juana	Begg	59ea7840fc13ae1f6d0000a0	\$80646.00	\$2924.69
59ea7840fc13ae1f6d0000a1	Alix	Layus	59ea7840fc13ae1f6d0000a1	\$50423.64	\$1495.33
59ea7840fc13ae1f6d0000a2	Tessa	Brandes	59ea7840fc13ae1f6d0000a2	\$60695.47	
59ea7840fc13ae1f6d0000a3	Hedwig	Fishley	59ea7840fc13ae1f6d0000a3	\$85763.22	
59ea7840fc13ae1f6d0000a4	Cordelia	Aubray	59ea7840fc13ae1f6d0000a4	\$60600.37	

Round 1	Output of round 1:
Mapper: Inserts "A" into employee_names.csv records	
Reducer: Identity reducer is used.	

	<table><tr><td>59ea7840fc13ae1f6d000096</td><td>A</td><td>Benny</td><td>Ventom</td></tr><tr><td>59ea7840fc13ae1f6d000097</td><td>A</td><td>Cara</td><td>Motherwell</td></tr><tr><td>59ea7840fc13ae1f6d000098</td><td>A</td><td>Willem</td><td>Haresign</td></tr><tr><td>59ea7840fc13ae1f6d000099</td><td>A</td><td>Trish</td><td>Farlane</td></tr><tr><td>59ea7840fc13ae1f6d00009a</td><td>A</td><td>Camilla</td><td>Limpkin</td></tr><tr><td>59ea7840fc13ae1f6d00009b</td><td>A</td><td>Godwin</td><td>Caffrey</td></tr><tr><td>59ea7840fc13ae1f6d00009c</td><td>A</td><td>Elvera</td><td>Custed</td></tr><tr><td>59ea7840fc13ae1f6d00009d</td><td>A</td><td>Vinson</td><td>Farres</td></tr><tr><td>59ea7840fc13ae1f6d00009e</td><td>A</td><td>Kliment</td><td>Pitchford</td></tr><tr><td>59ea7840fc13ae1f6d00009f</td><td>A</td><td>Matty</td><td>Heater</td></tr><tr><td>59ea7840fc13ae1f6d0000a0</td><td>A</td><td>Juana</td><td>Begg</td></tr><tr><td>59ea7840fc13ae1f6d0000a1</td><td>A</td><td>Alix</td><td>Layus</td></tr><tr><td>59ea7840fc13ae1f6d0000a2</td><td>A</td><td>Tessa</td><td>Brandes</td></tr><tr><td>59ea7840fc13ae1f6d0000a3</td><td>A</td><td>Hedwig</td><td>Fishley</td></tr><tr><td>59ea7840fc13ae1f6d0000a4</td><td>A</td><td>Cordelia</td><td>Aubray</td></tr></table>	59ea7840fc13ae1f6d000096	A	Benny	Ventom	59ea7840fc13ae1f6d000097	A	Cara	Motherwell	59ea7840fc13ae1f6d000098	A	Willem	Haresign	59ea7840fc13ae1f6d000099	A	Trish	Farlane	59ea7840fc13ae1f6d00009a	A	Camilla	Limpkin	59ea7840fc13ae1f6d00009b	A	Godwin	Caffrey	59ea7840fc13ae1f6d00009c	A	Elvera	Custed	59ea7840fc13ae1f6d00009d	A	Vinson	Farres	59ea7840fc13ae1f6d00009e	A	Kliment	Pitchford	59ea7840fc13ae1f6d00009f	A	Matty	Heater	59ea7840fc13ae1f6d0000a0	A	Juana	Begg	59ea7840fc13ae1f6d0000a1	A	Alix	Layus	59ea7840fc13ae1f6d0000a2	A	Tessa	Brandes	59ea7840fc13ae1f6d0000a3	A	Hedwig	Fishley	59ea7840fc13ae1f6d0000a4	A	Cordelia	Aubray
59ea7840fc13ae1f6d000096	A	Benny	Ventom																																																										
59ea7840fc13ae1f6d000097	A	Cara	Motherwell																																																										
59ea7840fc13ae1f6d000098	A	Willem	Haresign																																																										
59ea7840fc13ae1f6d000099	A	Trish	Farlane																																																										
59ea7840fc13ae1f6d00009a	A	Camilla	Limpkin																																																										
59ea7840fc13ae1f6d00009b	A	Godwin	Caffrey																																																										
59ea7840fc13ae1f6d00009c	A	Elvera	Custed																																																										
59ea7840fc13ae1f6d00009d	A	Vinson	Farres																																																										
59ea7840fc13ae1f6d00009e	A	Kliment	Pitchford																																																										
59ea7840fc13ae1f6d00009f	A	Matty	Heater																																																										
59ea7840fc13ae1f6d0000a0	A	Juana	Begg																																																										
59ea7840fc13ae1f6d0000a1	A	Alix	Layus																																																										
59ea7840fc13ae1f6d0000a2	A	Tessa	Brandes																																																										
59ea7840fc13ae1f6d0000a3	A	Hedwig	Fishley																																																										
59ea7840fc13ae1f6d0000a4	A	Cordelia	Aubray																																																										
<p>Round 2</p> <p>Mapper: Inserts "B" into employee_pay.csv records</p> <p>Reducer: Identity reducer is used.</p>	<p>Output of round 2:</p> <table><tr><td>59ea7840fc13ae1f6d000096</td><td>B</td><td>\$84217.20</td><td></td></tr><tr><td>59ea7840fc13ae1f6d000097</td><td>B</td><td>\$54737.84</td><td></td></tr><tr><td>59ea7840fc13ae1f6d000098</td><td>B</td><td>\$92092.21</td><td></td></tr><tr><td>59ea7840fc13ae1f6d000099</td><td>B</td><td>\$86117.63</td><td></td></tr><tr><td>59ea7840fc13ae1f6d00009a</td><td>B</td><td>\$92821.29</td><td></td></tr><tr><td>59ea7840fc13ae1f6d00009b</td><td>B</td><td>\$60633.55</td><td>\$1779.07</td></tr><tr><td>59ea7840fc13ae1f6d00009c</td><td>B</td><td>\$94483.80</td><td>\$4328.59</td></tr><tr><td>59ea7840fc13ae1f6d00009d</td><td>B</td><td>\$93127.77</td><td></td></tr><tr><td>59ea7840fc13ae1f6d00009e</td><td>B</td><td>\$90581.06</td><td></td></tr><tr><td>59ea7840fc13ae1f6d00009f</td><td>B</td><td>\$86947.54</td><td></td></tr><tr><td>59ea7840fc13ae1f6d0000a0</td><td>B</td><td>\$80646.00</td><td>\$2924.69</td></tr><tr><td>59ea7840fc13ae1f6d0000a1</td><td>B</td><td>\$50423.64</td><td>\$1495.33</td></tr><tr><td>59ea7840fc13ae1f6d0000a2</td><td>B</td><td>\$60695.47</td><td></td></tr><tr><td>59ea7840fc13ae1f6d0000a3</td><td>B</td><td>\$85763.22</td><td></td></tr><tr><td>59ea7840fc13ae1f6d0000a4</td><td>B</td><td>\$60600.37</td><td></td></tr></table>	59ea7840fc13ae1f6d000096	B	\$84217.20		59ea7840fc13ae1f6d000097	B	\$54737.84		59ea7840fc13ae1f6d000098	B	\$92092.21		59ea7840fc13ae1f6d000099	B	\$86117.63		59ea7840fc13ae1f6d00009a	B	\$92821.29		59ea7840fc13ae1f6d00009b	B	\$60633.55	\$1779.07	59ea7840fc13ae1f6d00009c	B	\$94483.80	\$4328.59	59ea7840fc13ae1f6d00009d	B	\$93127.77		59ea7840fc13ae1f6d00009e	B	\$90581.06		59ea7840fc13ae1f6d00009f	B	\$86947.54		59ea7840fc13ae1f6d0000a0	B	\$80646.00	\$2924.69	59ea7840fc13ae1f6d0000a1	B	\$50423.64	\$1495.33	59ea7840fc13ae1f6d0000a2	B	\$60695.47		59ea7840fc13ae1f6d0000a3	B	\$85763.22		59ea7840fc13ae1f6d0000a4	B	\$60600.37	
59ea7840fc13ae1f6d000096	B	\$84217.20																																																											
59ea7840fc13ae1f6d000097	B	\$54737.84																																																											
59ea7840fc13ae1f6d000098	B	\$92092.21																																																											
59ea7840fc13ae1f6d000099	B	\$86117.63																																																											
59ea7840fc13ae1f6d00009a	B	\$92821.29																																																											
59ea7840fc13ae1f6d00009b	B	\$60633.55	\$1779.07																																																										
59ea7840fc13ae1f6d00009c	B	\$94483.80	\$4328.59																																																										
59ea7840fc13ae1f6d00009d	B	\$93127.77																																																											
59ea7840fc13ae1f6d00009e	B	\$90581.06																																																											
59ea7840fc13ae1f6d00009f	B	\$86947.54																																																											
59ea7840fc13ae1f6d0000a0	B	\$80646.00	\$2924.69																																																										
59ea7840fc13ae1f6d0000a1	B	\$50423.64	\$1495.33																																																										
59ea7840fc13ae1f6d0000a2	B	\$60695.47																																																											
59ea7840fc13ae1f6d0000a3	B	\$85763.22																																																											
59ea7840fc13ae1f6d0000a4	B	\$60600.37																																																											
<p>Round 3</p> <p>Mapper: Identity mapper is used thus combining the outputs of both the previous steps. The identity mapper will combine and sort the results by key where key is the first 2 columns (id and A/B).</p>	<p>Output of mapper phase after combining and sorting (input to the reducer step):</p>																																																												

Reducer:

String firstname = null

String lastname = null

For each line read from the input file:

- Look at the second column, if it's A, update values of firstname and lastname with 3rd and 4th column.
- If it's B, check if variables firstname and lastname are null.
 - If either of them is not null, then append to output, the 1st column, firstname, lastname 3rd column, and 4th column of the current line and set firstname and lastname to null.
 - If both firstname and lastname are null, do nothing.

59ea7840fc13ae1f6d000096	A	Benny	Ventom
59ea7840fc13ae1f6d000096	B	\$84217.20	
59ea7840fc13ae1f6d000097	A	Cara	Motherwell
59ea7840fc13ae1f6d000097	B	\$54737.84	
59ea7840fc13ae1f6d000098	A	Willem	Haresign
59ea7840fc13ae1f6d000098	B	\$92092.21	
59ea7840fc13ae1f6d000099	A	Trish	Farlane
59ea7840fc13ae1f6d000099	B	\$86117.63	
59ea7840fc13ae1f6d00009a	A	Camilla	Limpkin
59ea7840fc13ae1f6d00009a	B	\$92821.29	
59ea7840fc13ae1f6d00009b	A	Godwin	Caffrey
59ea7840fc13ae1f6d00009b	B	\$60633.55	\$1779.07
59ea7840fc13ae1f6d00009c	A	Elvera	Custed
59ea7840fc13ae1f6d00009c	B	\$94483.80	\$4328.59
59ea7840fc13ae1f6d00009d	A	Vinson	Farres
59ea7840fc13ae1f6d00009d	B	\$93127.77	
59ea7840fc13ae1f6d00009e	A	Kliment	Pitchford
59ea7840fc13ae1f6d00009e	B	\$90581.06	
59ea7840fc13ae1f6d00009f	A	Matty	Heater
59ea7840fc13ae1f6d00009f	B	\$86947.54	
59ea7840fc13ae1f6d0000a0	A	Juana	Begg
59ea7840fc13ae1f6d0000a0	B	\$80646.00	\$2924.69
59ea7840fc13ae1f6d0000a1	A	Alix	Layus
59ea7840fc13ae1f6d0000a1	B	\$50423.64	\$1495.33
59ea7840fc13ae1f6d0000a2	A	Tessa	Brandes
59ea7840fc13ae1f6d0000a2	B	\$60695.47	
59ea7840fc13ae1f6d0000a3	A	Hedwig	Fishley
59ea7840fc13ae1f6d0000a3	B	\$85763.22	
59ea7840fc13ae1f6d0000a4	A	Cordelia	Aubray
59ea7840fc13ae1f6d0000a4	B	\$60600.37	

Output of the reducer phase: (Same as the expected output)

	59ea7840fc13ae1f6d000096	Benny	Ventom	\$84217.20	
	59ea7840fc13ae1f6d000097	Cara	Motherwell	\$54737.84	
	59ea7840fc13ae1f6d000098	Willem	Haresign	\$92092.21	
	59ea7840fc13ae1f6d000099	Trish	Farlane	\$86117.63	
	59ea7840fc13ae1f6d00009a	Camilla	Limpkin	\$92821.29	
	59ea7840fc13ae1f6d00009b	Godwin	Caffrey	\$60633.55	\$1779.07
	59ea7840fc13ae1f6d00009c	Elvera	Custed	\$94483.80	\$4328.59
	59ea7840fc13ae1f6d00009d	Vinson	Farres	\$93127.77	
	59ea7840fc13ae1f6d00009e	Kliment	Pitchford	\$90581.06	
	59ea7840fc13ae1f6d00009f	Matty	Heater	\$86947.54	
	59ea7840fc13ae1f6d0000a0	Juana	Begg	\$80646.00	\$2924.69
	59ea7840fc13ae1f6d0000a1	Alix	Layus	\$50423.64	\$1495.33
	59ea7840fc13ae1f6d0000a2	Tessa	Brandes	\$60695.47	
	59ea7840fc13ae1f6d0000a3	Hedwig	Fishley	\$85763.22	
	59ea7840fc13ae1f6d0000a4	Cordelia	Aubray	\$60600.37	