**Student Name: Liang Leran Jenny**

**Student UOW ID: 7063167**

**Solution 3**

customers.txt

|  |  |
| --- | --- |
| Customer Number | Customer Name |
| 00001 | James |
| 00002 | Harry |
| 00003 | Peter |
| 00004 | Jane |

orders.txt

|  |  |  |
| --- | --- | --- |
| Order Number | Customer Number | Total Order Value |
| 0000001 | 00001 | 34.5 |
| 0000002 | 00001 | 23.0 |
| 0000003 | 00002 | 123.0 |
| 0000004 | 00003 | 12.3 |

Map Phase:

I will implement one mapper for customer details and another mapper for order details.

For customer details, I will take in one tuple at a time and tokenize and fetch the customer number. The customer number will be my key of the key-value pair, making the key-value pair: [customer number, 1].

For order details, I will follow the same steps as customer details. Customer number will be the key to the key-value pair, making the key-value pair: [customer number, 1]

Reduce Phase:

In the reduce phase, I will add the count of each customer number value.

If the sum is 1, I will print the customer number as this indicates that a customer has made zero orders as shown using the sample above.

Before reduction:

[00001, 1], [00001, 1], [00001, 1], [00002, 1], [00002, 1], [00003, 1], [00003, 1], [00004, 1]

After reduction:

[00001, 3], [00002, 2], [00003, 2], [00004, 1]

Customer 00004, Jane, will be listed.