

IT 563: Data Mining

Lab 6

ID:201811024

Name- Rajat Kumar

Assignment- To perform Natural Join Using Map Reduce

Mapper:

```
import sys
for line in sys.stdin:
    # Setting some defaults
    user_id = ""
    product_id = "-"
    location = "-"

    line = line.strip()
    splits = line.split("\t")
    if len(splits) == 5: # Transactions have more columns than users
        user_id = splits[2]
        product_id = splits[1]
    else:
        user_id = splits[0]
        location = splits[3]
    print '%s\t%s\t%s' % (user_id,product_id,location)
```

Reducer:

```
#!/usr/bin/env python
import sys
import string

last_user_id = None
cur_location = "-"

for line in sys.stdin:
    line = line.strip()
    user_id,product_id,location = line.split("\t")
```

```
if not last_user_id or last_user_id != user_id:
    last_user_id = user_id
    cur_location = location
elif user_id == last_user_id:
    location = cur_location
    print '%s\t%s' % (product_id,location)
```