Task :

Part 1:

```
# Required libraries in R;
library(stringr)
library(reshape)
library(dplyr)

#create the tables
tableA <- data.frame("product" = c(100, 101, 102),
                     "tags" = c("chocolate, sprinkles", "chocolate, sprinkles", "glazed"))

out <- strsplit(as.character(tableA$tags),', ')
tags = do.call(rbind, out)
product=tableA$product

tableA=data.frame(product,tags)
tableA= melt(tableA, id=(c("product")))
tableA = tableA[,-2]
tableA

  product     value
1     100 chocolate
2     101 chocolate
3     102    glazed
4     100 sprinkles
5     101 sprinkles
6     102    glazed


tableB <- data.frame("customer" = c('A','A', 'B', 'C', 'C', 'B', 'A', 'C'),
                     "product" = c(100, 101, 101, 100, 102, 101, 100, 102))

a=merge(x = tableA, y = tableB, by = "product", all.x = TRUE)
a
a=a[,-1]



a1=a %>%
  group_by(customer, value) %>%
```

```
  summarise(n = n())


new=dcast(a1, customer ~ value , value.var="n", fill=0)
new

  customer  chocolate  glazed  sprinkles
1        A          3        0          3
2        B          2        0          2
3        C          1        4          1
```

Bonus Question: **If the two starting tables were in a relational database or Hadoop cluster and each had 100 million rows, how might your approach change?**

Ans:
We have to integrate our R engine with hadoop. Also we can create our user defined function to do same task and then integrate R with Mysql or hadoop after that we can easily do the same task on 100 millions rows.

I have already done task like that using Mysql and R engine.
Here we have to create data table in r and then connect our r engine to mysql database and finally pass this table into mysql database.