Download sample data:
https://s3-eu-west-1.amazonaws.com/insider-sample-data/rec/sample.csv

Example Row:

2016-12-28 06:39:32
http://www.shoppbagg.com/de-DE/DE/product/SHOPP-BAGG/Damen/Hose/2613648/512771
http://www.shoppbagg.com/de-DE/DE/product/SHOPP-BAGG/Damen/Hose/2613648/513058 productDetail
["Damen","Hosen"] Hose
http://akstatic.shoppbagg.com/ProductImages/20162/2/2613648/M_20162-6K9807Z8-2AE_A.jpg
http://www.shoppbagg.com/de-DE/DE/product/SHOPP-BAGG/Damen/Hose/2613648/513058 \N \N TL
0 \N 1482896122020e8376dcc13.a90752a1
7c6f44d0-03c5-ac52-5c8e-36069f95334f_1482899722 2016 12 28 6 513058

Fields:

1) Date
2) Referrer
3) Current URL
4) Page Type (main, category, productDetail, cart, success, other)
5) Product Categories
6) Product Name
7) Image URL
8) –
9) –
10) –
11) Currency
12) Cart Amount
13) –
14) **User ID**
15) **Session ID**
16) Year
17) Month
18) Day
19) Hour
20) **Product ID**

1) Load this data into your local Spark environment. Write a query which results like below. It means, how many times these two products are viewed together.

P_ID1	P_ID2	frequency
a	b	10
a	c	7
a	d	1
b	c	1
c	d	1

2) Set up an Elasticsearch service in your local environment, prepare a data model and store your results in it. Demonstrate the case blow:
- Send a request to a productid and show the productids that you recommend to that product as a response. They should be sorted by frequency.

3) Draw an architecture by showing how the system will work in a production environment in the Cloud.
- Specify the technologies that you will use in all the layers. (e.g. Lambda Architecture, SMACK architecture)
- Explain why you use the technologies that you will design.