Question 1

Describe a Data Engineer role in an organisation and its main responsibilities.

Question 2

From the three tables in Appendix A

    a. Show schema generation query
    b. Show the SQL query for number of customers purchasing more than 5 books
    c. Show the SQL query for a list of customers who never purchased anything
    d. Show the SQL query for list of book purchased with the users

You can attempt this question in any SQL based database (SQLSRV, MYSQL, OracleDB).

Question 3

Based on Question 2, implement all queries and the ingestion/extraction process of Appendix A in Python.

You can attempt this question in your own development workspace and share GitHub repository or gist URL.

Question 4 - *This is an optional question*

Megah Holdings Berhad is a diversified holding company in a few industries. Each industry has different best practices and different ERP implementation methods.

The revenue optimization team requires a dashboard which displays and analyses daily sales at the end of business day from three business units. Describe a data pipeline with the following input sources:

    a. Retail company ERP - Real time API in XML
    b. Chicken Broiler/Farm ERP - Hourly batch file generation in FTP server
    c. Trading company ERP - Manual Excel files download

Appendix A - *Bookstore POS System*

customers.csv

id	name	email	tel	created_at	updated_at
1	Irfan Bakti	irfan88@example.com	+60123456789	2019-08-07 08:13:21	2019-08-07 08:13:21
2	Jack Smmith	jack.smmith@acme.io	+60132456781	2019-08-07 08:13:21	2019-08-07 08:13:21
3	Nazir		+601185434012	2019-08-07 08:13:21	2019-08-07 08:13:21
4	Faiz Ma	faiz.ma@jholow.cn	+6019772002	2019-08-07 08:13:21	2019-08-07 08:13:21
5	Isham Rais	isham@pmo.gov.my	+60135482020	2019-08-07 08:13:21	2019-08-07 08:13:21
6	Shanon Teoh	shahnon.teoh@st.com.sg		2019-08-07 08:13:21	2019-08-07 08:13:21

invoices.csv

id	number	sub_total	tax_total	total	customer_id	created_at	updated_at
1	20190001	30.00	0.00	30.00	1	2019-08-07 08:13:21	2019-08-07 08:13:21
2	20190002	150.00	0.00	150.00	2	2019-08-07 08:13:21	2019-08-07 08:13:21
3	20190003	30.00	0.00	30.00	2	2019-09-15 08:13:21	2019-09-15 08:13:21
4	20190004	55.00	0.00	55.00	3	2019-09-15 08:13:21	2019-09-15 08:13:21
5	20190005	450.00	0.00	0.00	6	2019-09-15 08:13:21	2019-09-15 08:13:21

invoice_lines.csv

id	description	unit_price	quantity	sub_total	tax_total	total	tax_id	sku_id	invoice_id
1	Book #1	30.00	1	30.00	0.00	30.00	NULL	1	1
2	Book #2	25.00	4	100.00	0.00	100.00	NULL	2	2
3	Book #3	50.00	1	50.00	0.00	30.00	NULL	3	2
4	Book #1	30.00	1	30.00	0.00	30.00	NULL	1	3
5	Book #1	30.00	1	30.00	0.00	30.00	NULL	1	4
6	Book #2	25.00	1	25.00	0.00	25.00	NULL	2	4
7	Book #1	30.00	5	150.00	0.00	150.00	NULL	1	5
8	Book #3	50.00	6	300.00	0.00	300.00	NULL	3	5