**Question 1**

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

A Data Engineer in a company is like a data organizer. They create safe places for lots of information and build paths for data to move smoothly. They tidy up and take care of the data, handle technical things like computers, and work with teams to keep information safe. If there's a problem, they fix it and make guides so everyone can use the data easily. Simply put, Data Engineers help companies use their data wisely and easily.

**Question 2**

From the three tables in Appendix A

1. Show schema generation query

CREATE TABLE `bookstore\_pos\_system`.`customer` (

`id` INT NOT NULL,

`name` VARCHAR(45) NOT NULL,

`email` VARCHAR(255) NULL,

`tel` VARCHAR(45) NULL,

`created\_at` TIMESTAMP NOT NULL,

`updated\_at` TIMESTAMP NOT NULL,

PRIMARY KEY (`id`),

UNIQUE INDEX `email\_UNIQUE` (`email` ASC) VISIBLE,

UNIQUE INDEX `tel\_UNIQUE` (`tel` ASC) VISIBLE

);

CREATE TABLE `bookstore\_pos\_system`.`invoices` (

`id` INT NOT NULL,

`number` INT NOT NULL,

`sub\_total` DECIMAL(10,2) NOT NULL,

`tax\_total` DECIMAL(10,2) NOT NULL,

`total` DECIMAL(10,2) NOT NULL,

`customer\_id` INT NOT NULL,

`created\_at` TIMESTAMP NOT NULL,

`updated\_at` TIMESTAMP NOT NULL,

PRIMARY KEY (`id`)

);

CREATE TABLE `bookstore\_pos\_system`.`invoices\_lines` (

`id` INT NOT NULL,

`description` VARCHAR(45) NOT NULL,

`unit\_price` DECIMAL(10,2) NOT NULL,

`quantity` INT NOT NULL,

`sub\_total` DECIMAL(10,2) NOT NULL,

`tax\_total` DECIMAL(10,2) NOT NULL,

`total` DECIMAL(10,2) NOT NULL,

`tax\_id` VARCHAR(45) NOT NULL,

`sku\_id` INT NOT NULL,

`invoice\_id` INT NOT NULL,

PRIMARY KEY (`id`)

);

1. Show the SQL query for number of customers purchasing more than 5 books

SELECT

customers.id,

customers.name,

SUM(invoices\_lines.quantity) AS total\_books\_purchased

FROM

customers

JOIN invoices ON customers.id = invoices.customer\_id

JOIN invoices\_lines ON invoices.id = invoices\_lines.invoice\_id

GROUP BY

customers.id

HAVING

total\_books\_purchased > 5;

1. Show the SQL query for a list of customers who never purchased anything

SELECT

customers.id,

customers.name

FROM

customers

LEFT JOIN

invoices ON customers.id = invoices.customer\_id

WHERE

invoices.id IS NULL;

1. Show the SQL query for list of book purchased with the users

SELECT

customers.id AS customer\_id,

customers.name AS customer\_name,

invoices\_lines.description AS book\_description,

invoices\_lines.quantity AS book\_quantity

FROM

customers

JOIN

invoices ON customers.id = invoices.customer\_id

JOIN

invoices\_lines ON invoices.id = invoices\_lines.invoice\_id;

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.

This 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:

1. Retail company ERP - Real time API in XML
2. Chicken Broiler/Farm ERP - Hourly batch file generation in FTP server
3. 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 Smith | 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 |

invoices\_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 | 50.00 | 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 |