1. Consider the migration of customer and transactional data from an external legacy system to a new bespoke platform. Describe what you think the key challenges and principles would be in the preparation and execution of the migration.

Some of the main reasons for cloud migration are mentioned below:

Availability: Availability of critical data or applications.

Flexibility and Scalability: Readily to changing demand or to scale with the business.

Reduced Cost: Clouds can be cheaper than a company can in-house.

Simplified Management: Reduce responsibility for configuring, managing, and securing its IT infrastructure.

Cloud Migration Challenges: Cloud migration can bring significant benefits, but there are quite a few challenges along the way. Top five challenges that organizations encounter when moving to the cloud:

1. **A cloud migration strategy:** Based on the goals, cloud services can be selected, and a strategy can be developed to guide the migration.
2. **Complex existing architecture**: This requires experienced engineers to create infrastructures.
3. **Long migration process**: Often, migrations are performed in multiple stages with in-depth testing and validation between each stage of the process.
4. **High cloud costs**: An in-depth understanding of a business's cloud needs and providers' offerings can help to optimize cloud infrastructure and drive down the costs of the cloud.
5. **Data security and compliance risks**: A crucial part is that data and applications will be secure throughout the entire process.
6. **Data conversions and Data reductions.**

2. Consider the following data objects: Customer, CustomerPhoneNumber, CustomerCitizenship, CustomerAddress, CustomerEmployment.
   a) List the key attributes commonly associated with these objects
   b) Draw an Entity Relationship Diagram that shows the relationships between these objects (one-to-one, one-to-many, many-to-many, zero-to-one, zero-to-many)
   c) If we were to create a database representation for this model what primary keys and foreign keys would you suggest

Please look at the files RE_Cus_Emp.pdf (ER diagram) and SQLCodes.txt files to see attributes and database representation.

3. Consider the migration of data from a live production system that must stay online during the event. How would you minimise any disruption?

Considering details of security protocols that are used for transferring data.

A log of the steps should be saved to follow up on the signs of progress.

Alarming of the failure should be planned.

A rollback strategy should be planned.

Cleansing activities need to be performed across all source systems.

Data can come in several formats and use different technologies for storage, which means it may require conversion before being used by a new platform.

Map the application and database schema correctly.

Remediate the poor data quality before migrating the data.

A strong data governance plan, including issues and data management, must be put in place to ensure that proper follow-up activities such as QC & compliance checks are carried out throughout the project life cycle.

Managing the latencies should be planned.

Test migration of data with smaller sizes to evaluate the process.

Split data and schedule data ingestions.

Manage the scalability of the data migration considering the speed and size of the data.

4. Consider the following database tables and contents:

CustomerType	
CustomerTypeId	CustomerTypeDisplayName
1	Individual
2	Organisation

Customer		
CustomerId	CustomerName	CustomerTypeId
1	John Smith	1
2	Peterson Ltd	2
3	Paul Taylor	1
4	Janet Peters	1

Order		
OrderId	CustomerId	OrderDate
1	1	2010-01-01
2	4	2010-01-02
3	2	2010-01-02
4	2	2010-02-10
5	2	2010-03-05
6	1	2011-10-01
7	4	2011-11-01
8	2	2012-05-01

Given the database above please write a SQL query that would return the data in the following format:

CustomerName	CustomerTypeDisplayName	NumberOfOrders	EarliestOrderDate
John Smith	Individual	2	2010-01-01
...	...	...	...

Please look at the corresponding answer that put in SQLCodes.txt file.

5. Consider a live production trading system. There is an incoming daily CSV file containing new and updated client restrictions.
    a) Describe how you would design a process for transforming and loading the data into the target Restrictions table
    b) Does the time and frequency of file delivery have any impact on the process?
    c) What are the limitations and risks given the provided file format and database schema and how could you improve upon them?

Answer 5-b:

Yes, If we have a high amount of data in quick time frequencies we missed or down the ingestions because it needs some time for transformation and some time for ingestion.

Answer 5-c:

There is a risk of type conversion of the IsEnabled column because it should be boolean. Moreover, the definitions of the key columns for redundant data should be considered. If there are no keys then we get an inconsistent result that contains only one redundant row. For example, if we have two TypeA with identical name values then an inconsistency will happen.

CSV file

```
RestrictionId, Name, RestrictionType, IsEnabled
1, John, TypeA, 1
2, John, TypeB, 1
3, John, TypeC, 0
4, Jane, TypeA, 0
5, Jane, TypeB, 1
6, Jane, TypeC, 1
```
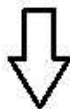
Restrictions

Name	IsRestrictionTypeAEnabled	IsRestrictionTypeBEnabled	IsRestrictionTypeCEnabled

6. Consider a data migration task of moving Customer details from a legacy system (Legacy.Customer table) into a new environment that's not live yet (Production.Customer and Production.Employment tables).
   a) What challenges with data quality and potential data cleansing would you identify and how could those issues be resolved?
   b) What reconciliation methods might you use to prove a successful data migration?
   c) If the target environment was live would it change your approach?

```sql
-- Source

CREATE TABLE [Legacy].[Customer](
    [Name] [varchar] (500) NOT NULL,
    [Surname] [varchar] (500) NOT NULL,
    [DOB] [date] NULL,
    [Gender] [varchar](10) NULL,
    [EmploymentStatus] [varchar](10) NOT NULL,
    CONSTRAINT PK_CustomerLegacy PRIMARY KEY ([Name],[Surname])
)

-- Targets

CREATE TABLE [Production].[Customer](
    [CustomerId] [int] NOT NULL PRIMARY KEY,
    [Name] [varchar] (150) NOT NULL,
    [Surname] [varchar] (150) NOT NULL,
    [DateOfBirth] [smalldatetime] NOT NULL,
    [GenderId] [int] NOT NULL REFERENCES Genders(Id)
)
CREATE TABLE [Production].[Employment](
    [CustomerId] [int] NOT NULL PRIMARY KEY REFERENCES [Customer]([CustomerId]),
    [EmploymentStatus] [int] NOT NULL REFERENCES Statuses(Id),
    [EmployerName] [varchar] (250) NOT NULL,
    [Occupation] [varchar] (50) NULL
)
```

Answer 6.1, 6.2) We should be considered if the length of data is more than 150 characters then we ignore the longer data for Name and Surname. We should set the primary key as an automatic increment case. In the conversion of data to smalldatetime we miss the seconds and the seconds are rounded to the nearest minutes value, so if we need the second, we should not convert the date to smalldatetime.

If we try to insert seconds, values up to 29.998 it is rounded down to the nearest minute. Values above 29.999 seconds are rounded up. In the conversion of the varchar type to the int type we should be careful that if there is a character on data then an error presents also if the data has decimal values, then the data are rounded to an int value.

Answer 6.3) Speedness of the data transformation depends on the data storage capabilities, sometimes it can automatically buffer data and then storage sometimes is out of its possibility and misses data.