

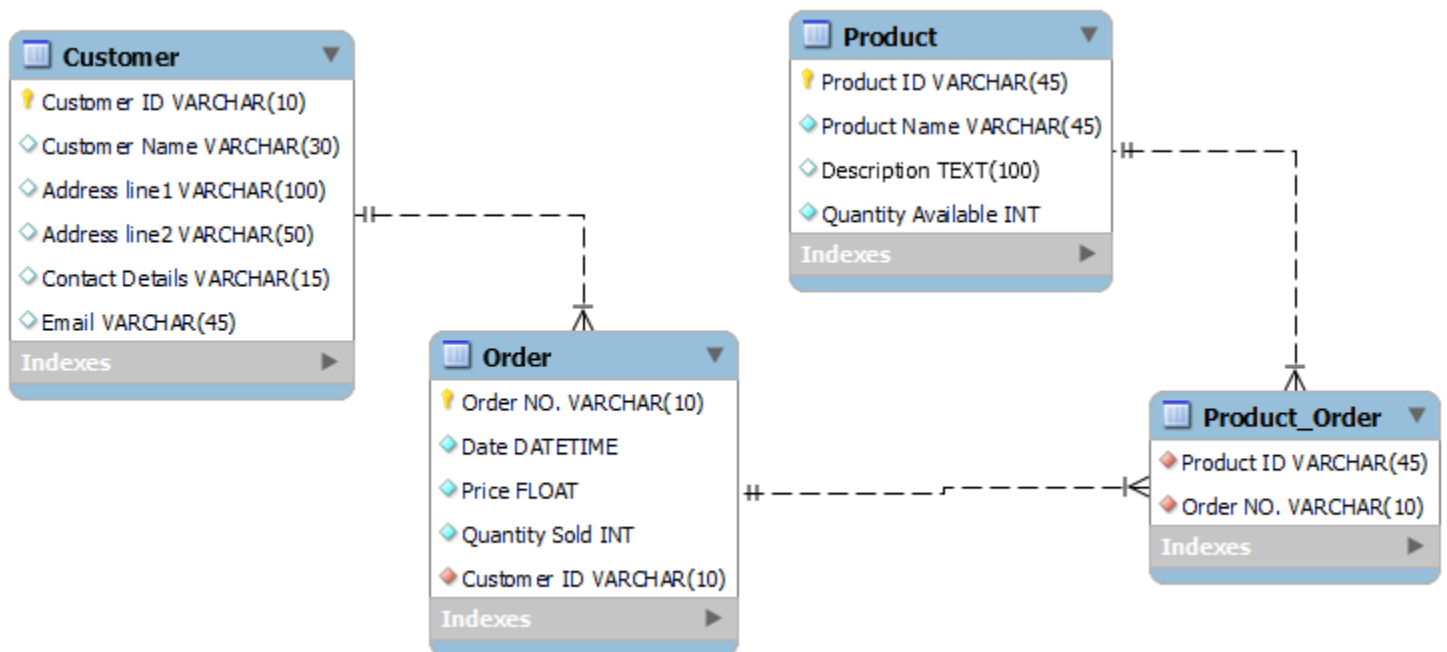
# Data Management Fundamentals – UFCFLR-15-M

Student : Parashkumar Shah

Student ID : 21051422

---

**Answer 1:** A logical Entity-Relationship (ER) model showing all entities, relationships, and Keys generated based on given invoice sample taken from ZZ Trading of Bristol is presented below.



## **Answer 2:** Describing and Explaining the three aspects of the relational model.

The relational model of databases has three related aspects known as the **Structural** aspect, the **Integrity** aspect, and the **Manipulative** aspect.

**The Structural aspect:** The structure of the Relational model is purely based on tables at the core. Tables and nothing else than the tables. And a table includes rows, columns, and records. While the terminologies and interpretation in Relational model is little different, below are the components and terminologies used in the relational model of data.

- **Relation**

The main structure or a body which comprises all the columns and rows. There are several relations in the database which record a unique set of data in it.

In the below example the whole structure is a Relation.

Player	Matches	Goals	Assists	Minutes	Saves	Rating
John	20	1	0	1800	92	7
Jim	18	0	0	1620	108	8
Jacob	38	1	2	3420	0	4
Jerome	38	3	1	3420	0	5
Jack	38	2	1	3420	0	6
Jonah	38	9	15	3420	0	7
Jason	38	8	12	3000	0	6
Jen	38	10	3	3000	0	5
Jackie	38	2	44	3420	0	10
Jill	38	3	5	1800	0	6
Janet	38	18	2	2800	0	8
Jade	38	39	20	3420	0	10

- **Tuple**

The number of records or rows in the relation is known as tuple. Each tuple tends to have a unique identifier which is known as unique ID. There can be multiple tuples in one Relation, but they must be derived from a same type of transaction.

In the above example each record like this is a tuple.

Player	Matches	Goals	Assists	Minutes	Saves	Rating
John	20	1	0	1800	92	7

- **Attribute**

The internal separation of a Relation by different columns is known as Attributes. Each Relation can have many attributes which carry unique fields to be recorded. There can only be one data type permitted to be recorded in the Attribute.

In the above example each field like this is an Attribute

Matches
20
18
38
38
38
38
38
38
38
38
38
38

- Domain**

It specifies the kind of data represented by the Attribute. A domain is a set of each possible value that are validly contained in an Attribute. It is not to be mistaken with datatypes like be integer, variable-character, datetime, string, etc. As two Attributes can have same data type also.

In the above example “Goals” and “Assists” are both carrying same data type but the domain is different.

Goals	Assists
1	0
0	0
1	2
3	1
2	1
9	15
8	12
10	3

- Degree of Relation**

The number of Attributes in the relation is known as Degree of Relation.

In the above example there is 7 Degree of Relation.

Player	Matches	Goals	Assists	Minutes	Saves	Rating
--------	---------	-------	---------	---------	-------	--------

- Cardinality of Relation**

The number of records or numbers of Tuples is known as Cardinality of relation.

In the above example it is 12.

John
Jim
Jacob
Jerome
Jack
Jonah
Jason
Jen
Jackie
Jill
Janet
Jade

**The Integrity aspect:** There are three constraints of Integrity aspects in any relational data model. These help validate the structure of the data and help avoid errors.

- **Entity Integrity**

It defines the concept of the Primary key. It establishes a rule that every relation must have a unique column/Attribute to be chosen as Primary key and that must be unique.

It can not be Null as well.

Example:

In answer 1 of this exam, Customer ID is a primary key of Customer Table. Similarly Order NO. and Product ID are primary keys for their respective tables.

- **Referential Integrity**

It defines the concept of foreign key. According to the rule, foreign key can either be referred to a primary key of some table in the database or it can be Null saying that there is no relation between represented objects or the relation is not known.

Example:

In the answer 1, Customer ID in Order table is foreign key which is referred to Customer table.

- **Domain Integrity**

It specifies that all the columns must have defined domain. It allows to record the data of same data type only. The data types can be Integer, varchar, datetime, text, etc. but they have to be the same in a particular column.

Example:

Description under Product table is restricted with Text() and Date in Order table is restricted with Datetime in answer 1.

**The Manipulative aspect:** It allows to manipulate the data recorded in the tables in database through restricted format established according to the rules of algebra. The four basic manipulative aspects of relational model of data are as follow.

- **Insert**

This allows to insert the data into tables of the database. Insert operations can be performed without having any anomaly in all cases.

Example: We want to insert a table in database to record details of invoice and another table for recording customer details even before making any sale. Then we can fetch the data whenever any transaction occurs through following code.

```
INSERT INTO table VALUES (value1, [value2, ... ])
```

- **Update**

If any changes happen in transactions, it needs to be changed in the database as well. Update operations can be performed without any anomalies in Relational data bases.

Example: Suppose we need to change the contact details of a particular customer then.

```
UPDATE table_name SET column_name = value [, column_name = value ...]
```

- **Delete**

For unwanted records or columns or even whole table, delete query will execute and delete all unnecessary data from the database.

Example: we want to delete records of all years before 2000.

```
DELETE FROM table_name [WHERE condition];
```

- **Record Retrieval**

For fetching a desired set of data, we have a set of commands. These are some of the very basic queries of relational databases.

Exmples: We want to find all the purchases made by a particular client

```
SELECT column_name FROM table_name [WHERE condition];
```

---

**Answer 3:** One or more relations(tables) are taken as input by a relational operator and a new solution is produced as output (Closure). Relational algebra consists of 8 high-level operators. The Three Major Manipulative Operators in relational algebra are described and explained as follows.

- **Restrict:**

A unary operator or monadic. It helps filter out the row/ records according to condition we give. Here all the columns will remain intact, but rows will get filtered as per the condition. Multiple restrict conditions can be applied in a single query.

Let's presume Z is a relation and G, H are attributes

Then, restriction: Z WHERE X theta Y

Possibility for theta: =, <>, >, <, >=, <=, etc

Let's understand it by an example:

In the following example, lets us find out the records where country is UK.

So, our SQL query in Relational database will be: **RESTRICT table WHERE country='UK';**



Here we restricted the outcome of all data by putting a condition to select countries which matches UK.

- **Project:**

It is unary operator. Needs one or more than one table/relation. It is used to filter out columns or attributes from a table in a database.

Let's presume Z is a relation and G, H, I are attributes

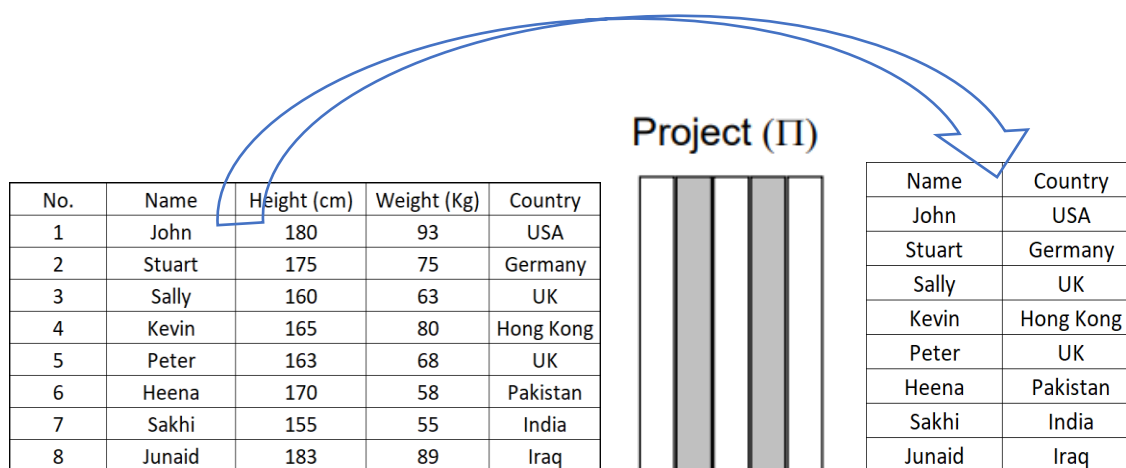
Then,  $A[G,H,I]$ :

Identify Projection:  $A = A$

Nullity Projection:  $A[] = \emptyset$

Let's use the same example to fetch desired columns out.

So, our SQL query in Relational database will be: **PROJECT table [Name, Country];**



Here we project out desired column – Name and Country

- **Product:**

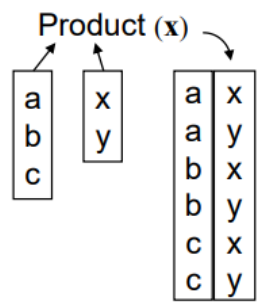
It is a binary operator. Needs minimum two or more tables/relations to product. It Builds a relation from two specified relations consisting of all possible combinations of rows, one from each of the two relations. It is also known as Cross product or Cartesian product.

Let's understand it by example.

Here, Two tables with completely different records and attributes are being combines with the help of Product.

So, our SQL query in Relational database will be: **PRODUCT table\_a WITH table\_b;**

No.	Name	Height (cm)	Weight (Kg)	Country
1	John	180	93	USA
2	Stuart	175	75	Germany
3	Sally	160	63	UK
4	Kevin	165	80	Hong Kong
5	Peter	163	68	UK
6	Heena	170	58	Pakistan
7	Sakhi	155	55	India
8	Junaid	183	89	Iraq



Qualification	Salary (US\$)
Engineering	120000
IT	150000

No.	Name	Height (cm)	Weight (Kg)	Country	Qualification	Salary (US\$)
1	John	180	93	USA	Engineering	120000
2	Stuart	175	75	Germany	Engineering	120000
3	Sally	160	63	UK	Engineering	120000
4	Kevin	165	80	Hong Kong	Engineering	120000
5	Peter	163	68	UK	Engineering	120000
6	Heena	170	58	Pakistan	Engineering	120000
7	Sakhi	155	55	India	Engineering	120000
8	Junaid	183	89	Iraq	Engineering	120000
1	John	180	93	USA	IT	150000
2	Stuart	175	75	Germany	IT	150000
3	Sally	160	63	UK	IT	150000
4	Kevin	165	80	Hong Kong	IT	150000
5	Peter	163	68	UK	IT	150000
6	Heena	170	58	Pakistan	IT	150000
7	Sakhi	155	55	India	IT	150000
8	Junaid	183	89	Iraq	IT	150000

In the above example a new table has been created with all the rows included in table a as well s table b. Each row of first table is multiplied by each row of second table.

-----



## **Answer 6(a):** Explaining the intrinsic value of data and how organizations can make use of that value.

### **Data**

Facts and Figures collected for the purpose of reference or analysis is called data. Proofs have been collected that Data has been stored from ancient civilizations. Data is everywhere and is generated ever since from the birth of this solar system. We create a huge amount of data every second knowingly or unknowingly. With the means and mediums getting more advanced, the storage and processing of data has been evolved to a totally different level. Of course, Data do not have any value on its own but when knowledge obtained from collection of data by analyzing it and understanding the patterns it becomes as valuable as a rare gem.

### **Intrinsic Value if Data**

Today world is running on data, many organizations have made their business out of it and Data today is no less than commodity as it is being generated, mined, collected, processed, and traded. Each commodity has a buyer who knows its value and is ready to pay the unimaginable price for as he is one of those few who knows its true worth. Similarly, data possesses its own intrinsic value and a fact to believe that it is rising with each passing day.

From human DNA to cosmic radiations generated from exoplanets, from every minute spent on social media to every online order made, Devices are all designed to collect and store data into cloud. Data collected today might be more valuable at or after certain point of time and even lose its worth after a certain long period of time. A true value of data is determined by assessing the data collected and where it can be made most useful by understanding its potential, then make informed investments in data to make the most of it.

In the modern era, Data is the new oil of 21<sup>st</sup> century it is giving rise to information age and driving economy towards knowledge economy. More the data – more the value. Data can never die; it is never affected by depreciation or needs any maintenance cost and data can be used for unlimited times for different purposes.

### **Organizations can benefit from that value**

Internet has become the giant source of data generation and it is impossible to figure true worth of internet, With the advancement in technology and advanced gadgets it will become even more powerful in future. Tech companies know this, and they have started their race in collection rather generation of data and make most out of it. Big Data collected from customers empowers companies to get into their head and plan strategies accordingly.

Examples:

- 1) Alphabet Inc (parent company of google) with over one Billion monthly active users reported revenue of \$65.1 billions in third quarter of 2021 and Facebook booked net profit of \$9.2 billion in the same period.
- 2) TikTok had just been launched in 2016 and was valued more than \$50 Billion in 2020.
- 3) Many other companies in health care, real estate and banking are also largely benefited from knowing the true value of data.

They provide free service, and when something is free you are the commodity. This is how they get benefit of Intrinsic value of data as people giving it do not recognize what they are feeding.

# **Answer 6(b)**: Outlining and discussing the concept of data modelling. Explaining features characterizing a good data model.

## **Data Modelling**

Data Modelling is a process that defines and analyze data requirements needed to support the organizations. It refers to structure of data store and design of data content by documenting inter relationship between data and presenting it with diagram and definitions.

User specifies its requirements, and, on that basis, formal model of domain is achieved through data modelling. Three kinds of instances are described by ANSI – Conceptual, Logical and Physical.

## **Method to create data model**

### **1. Entity-Relationship model**

- It is high level relational model
- Provides better view
- It is presented through Diagram consisting of relation, Entities and Attributes

### **2. Network Technique**

- Represents data in form of Graph
- It maintains multiple parent-child records in generalized manner

### **3. Hierarchical technique**

- It is a tree like structure
- One root node and other child nodes attached
- Very rarely used now

### **4. Object-Oriented model**

- It creates the object that contains stored values
- It communicates while supporting data abstraction, inheritance, and encapsulation

### **5. Relational Technique**

- Describes different relations between entities

## **Benefits of Data Modelling**

- Higher Quality
- Reduced Cost
- Clearer Scope
- Better Documentation
- Faster Performance
- Fewer Errors

## **Feature of Good Data model:**

### **Multiple usage of data**

The data stored in the data base should be independent of any specific application and It should be eligible to be used for multiple times.

### **Non-redundancy**

Duplication of records required more space and requires extra processing, also it creates mess in the data models. Data models should not include Redundant records.

### **Simple and Elegant**

There is more reliability and consistency in simple data models. It provides clean and neat visualization, makes recording easy and provides wonderful outcomes.

### **Flexibility**

Data models should be flexible and should be able to adjust in the times of changes in requirement of business. Whether the change is small or big it should remain stable in every situation possible.

### **Integration**

A well-designed Data model should be able to cope up with current and future databases of an organization. The coding schemas and definitions should be consistent, it should assemble different versions in step.

### **Reflection to real world problem**

It must record all the data available in the organization requirement and present that data in its true form.

### **Secure**

It must be secure with username and password protections, as company's confidential information is being stored in that. It must limit the access to the users.

---

## **Answer 7(a)**: Reasons for organizations are taking up NoSQL based solution with reference to “Polyglot Persistence”.

Polyglot Persistence is a term coined by Neal Ford in 2006 that provides architectural design based on collection of different databases at a part of platforms core design. With polyglot persistence, a company always chooses the right tool for the right job instead on relying on one piece of technology.

### **Reasons:**

#### **Faster response time**

All the features of all Databases are leverages in one application, which makes its response time very fast.

#### **Increases Scalability**

The final app scales well with data, When the databases are modeled properly All NoSQL databases tend to scale well.

#### **Simplifies operations**

It helps select the best component for the situation and helps eliminate fragmentation.

#### **Efficiency**

A polyglot Persistence application would automatically assign relevance-oriented processing, MongoDB cannot return result using relevance as a priority in the listings, Which the ElasticSearch app can.

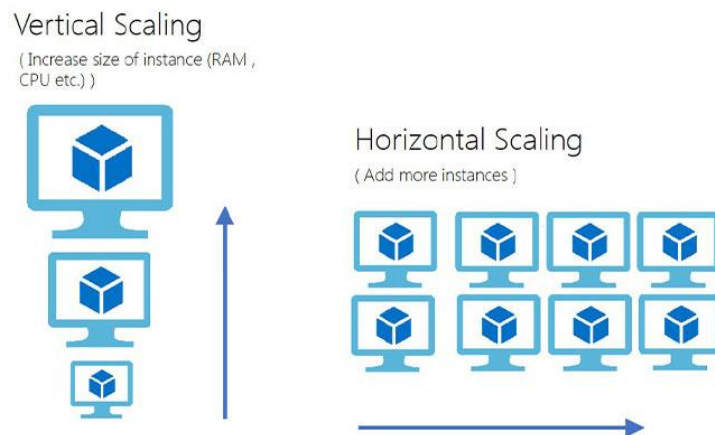
#### **Benefit of multiple databases**

Unique features from different can be combined to make a smooth experience in the app, from better handling of data calling the queries or large-scale analytics, all is possible.

---

## **Answer 7(b):** Difference between horizontal and vertical scaling NoSQL and SQL databases.

Scalability of an application can be measured by the number of requests it can effectively support at one time. The point at which it is no more possible for application to handle any additional request, it is the scalability limit of that application.



### **Vertical Scaling:**

- It means to improve the performance of the data base server by increasing the power. Example by upgrading its CPU or upgrading RAM.
- This is also known as Scaling up.
- More suitable for Relational Data base, where the data is stored at one place. Example MySQL, Amazon RDS
- There is no need to change the logic of the, just the code can be run on higher machine.
- It can help to handle huge task smoothly.
- Also describes as distributed programming as it involves distributing jobs.

### **Horizontal Scaling:**

- It means to add more servers and distribute the database across them instead of upgrading the database server.
  - This allows you work on one database hosted by multiple servers.
  - It is often referred as Scaling out.
  - More suitable for NoSQL data bases which operates on multiple servers. Examples MongoDB, Cassandra
  - It requires to break logic into small segments so that they can be run across multiple servers.
  - This can help do the task quickly.
  - Actor model: concurrent programming on multi-core machines is often performed via multi-threading and in-process message passing.
-

## **Answer 7(c):**

The first thing for the client is to record the generated invoice.

First Persistence layer:

This would be of Key-Value Pair where CustomerID from customer table and purchaseID from purchase table will be recorded and assigned key-Value pair.

Second Persistence layer:

In this layer, Document model will be used to store all the invoices, which will be assigned unique ID and all the fields would be given Key.

Third Persistence layer:

In the final layer graph model will be used to connect invoice with Customer id and Purchase ID. Thus, this will make querying and reporting accurate.

---