**Entity relationship diagram to relational schema - Exercise 5**

**Solved Exercise - Reducing Entity Relationship Diagram into Tables, Convert ER diagram to tables, relational schemas, ER model to relational model**

**Entity Relationship Diagram Exercise 5**

**Question:**

Convert/reduce the ER Diagram given in figure 1 below;

|  |
| --- |
| [https://4.bp.blogspot.com/-MmgB0J3yHI4/VyTXUNwHNCI/AAAAAAAABaA/V6KdKMRsZo8ymHLQJohB8yC6BA5tYH49gCLcB/s640/Basic_ERD.JPG](https://4.bp.blogspot.com/-MmgB0J3yHI4/VyTXUNwHNCI/AAAAAAAABaA/V6KdKMRsZo8ymHLQJohB8yC6BA5tYH49gCLcB/s1600/Basic_ERD.JPG) |
| Figure 1 - ER diagram with Composite attribute, Multi-valued attribute and Many to Many relationship |

**Solution:**

Given in the figure;

**Entity sets and relationship sets**

|  |  |  |
| --- | --- | --- |
| ***Name*** | ***Entity set / Relationship set*** | ***Type*** |
| ***Scientist*** | Entity set | Strong entity set |
| ***Invention*** | Entity set | Strong entity set |
| ***Invents*** | Relationship set | Many-to-Many relationship |

**Entity set *Scientist***

|  |  |  |
| --- | --- | --- |
| ***Attributes*** | ***Attribute Type*** | ***Description*** |
| SID | Simple and Primary key | Scientist ID |
| SNam | Composite | Scientist Name |
| RArea | Multi-valued | Research Area |
| Country | Simple | Country |

**Entity set *Invention***

|  |  |  |
| --- | --- | --- |
| ***Attributes*** | ***Attribute Type*** | ***Description*** |
| IID | Simple and Primary key | Invention ID |
| IName | Simple | Name of the invention |
| Year | Simple | Year of invention |

**Reduction into relational schema**

**Strong entity sets** – *Entity set that has a primary key to uniquely represent each entity is Strong entity set*.

Strong entity sets can be converted into relational schema by having the entity set name as the relation schema name and the attributes of that entity set as the attributes of relation schema.

Then we have,

Scientist (SID, SName, RArea, Country)

Invention (IID, IName, Year)

|  |
| --- |
| ***1. After converting strong entity sets into relation schema*** |
| Scientist (SID, SName, RArea, Country)  Invention (IID, IName, Year) |

**Composite attributes** – *If an attribute can be further divided into two or more component attributes, that attribute is called composite attribute*.

*While converting into relation schemas*, component attributes can be part of the strong entity sets’ relation schema. No need to retain the composite attribute.

In our case, SNam becomes FName, and LName as follows;

Scientist (SID, FName, LName, RArea, Country)

|  |
| --- |
| ***2. After converting composite attributes into relation schema*** |
| Scientist (SID, FName, LName, RArea, Country)  Invention (IID, IName, Year) |

**Multi-valued attributes –***Attributes that may have multiple values are referred as multi-valued attributes*.

In our ER diagram, RArea is a multi-valued attribute. That means, a scientist may have one or more areas as their research areas.

*To reduce a multi-valued attribute into a relation schema*, we have to create a separate table for each multi-valued attribute. Also, we need to include the primary key of strong entity set (parent entity set where the multi-valued attribute belongs) as a foreign key attribute to establish link.

In our case, the strong entity set Scientist will be further divided as follows;

Scientist (SID, FName, LName, RArea, Country)

Scientist\_Area (SID, RArea)

|  |
| --- |
| ***3. After converting multi-valued attributes into relation schema*** |
| Scientist (SID, FName, LName, RArea, Country)  Scientist\_Area (SID, RArea)  Invention (IID, IName, Year) |

**Relationship set** – *The association between two or more entity sets is termed as relationship set*.

A relationship may be either converted into a separate table or not. That can be decided based on the type of the relationship. Only many-to-many relationship needs to be created as a separate table.

Here, we are given a **many-to-many relationship**. That means,

* **one entity** (record/row) of ***Scientist*** is related to **one or more** entities (records/rows) of ***Invention*** entity set (that is, one scientist may have one or more inventions) and,
* **one entity** (record/row) of ***Invention*** is related to **one or more** entities (records/rows) of ***Scientist*** entity set. (that is, one or more scientists  may have invented one thing collectively).

*To reduce the relationship****Invents****into relational schema*, we need to create a separate table for ***Invents***, because ***Invents*** is a **many-to-many** relationship set. Hence, create a table ***Invents*** with the primary keys of participating entity sets (both, Scientist and Invention) as the attributes.

Then we have,

Invents (SID, IID)

Here, SID and IID are both foreign keys and collectively forms the primary key of ***Invents*** table.

Finally, we have the following relation schemas;

|  |
| --- |
| ***4. After converting relationship sets into relation schema*** |
| Scientist (SID, FName, LName, RArea, Country)  Scientist\_Area (SID, RArea)  Invention (IID, IName, Year)  Invents (SID, IID) |

### What is Big Data?

Big Data is also **data** but with a **huge size**. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

## Examples Of Big Data

Following are some the examples of Big Data-

The **New York Stock Exchange** generates about **one terabyte** of new trade data per day.

**Social Media**

The statistic shows that **500+terabytes** of new data get ingested into the databases of social media site **Facebook**, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

**Types Of Big Data**

BigData' could be found in three forms:

1. **Structured**
2. **Unstructured**
3. **Semi-structured**

### ****Structured****

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data. Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving value out of it. However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the rage of multiple zettabytes.

**Examples Of Structured Data**

An 'Employee' table in a database is an example of Structured Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Employee\_ID** | **Employee\_Name** | **Gender** | **Department** | **Salary\_In\_lacs** |
| 2365 | Rajesh Kulkarni | Male | Finance | 650000 |
| 3398 | Pratibha Joshi | Female | Admin | 650000 |
| 7465 | Shushil Roy | Male | Admin | 500000 |
| 7500 | Shubhojit Das | Male | Finance | 500000 |
| 7699 | Priya Sane | Female | Finance | 550000 |

## ****Unstructured****

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc. Now day organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

**Examples Of Un-structured Data**

The output returned by 'Google Search'

**Semi-structured**

Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data is a data represented in an XML file.

Examples Of Semi-structured Data

Personal data stored in an XML file-

<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>

<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>

<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>

<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>

<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>

**Characteristics Of Big Data**

***(i) Volume –*** The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, **'Volume'** is one characteristic which needs to be considered while dealing with Big Data.

***(ii) Variety –*** The next aspect of Big Data is its **variety**.

Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analyzing data.

***(iii) Velocity –*** The term **'velocity'** refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.

Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors,[Mobile](https://www.guru99.com/mobile-testing.html)devices, etc. The flow of data is massive and continuous.

***(iv) Variability –*** This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

**Benefits of Big Data Processing**

Ability to process Big Data brings in multiple benefits, such as-

* + Businesses can utilize outside intelligence while taking decisions

Access to social data from search engines and sites like facebook, twitter are enabling organizations to fine tune their business strategies.

* + Improved customer service

Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.

* + Early identification of risk to the product/services, if any
  + Better operational efficiency

Big Data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.