|  |
| --- |
| Department of Software Engineering  Mehran University of Engineering and Technology, Jamshoro |

|  |  |  |  |
| --- | --- | --- | --- |
| Course: SWE324 - Data Warehousing and Data Mining | | | |
| Instructor | Rabeea Jaffari | **Practical/Lab No.** | 05 |
| Date | 30 April 2019 | **CLOs** | CLO-4: P3 & P4 |
| Signature |  | **Assessment Score** | 1 Marks |

|  |  |
| --- | --- |
| Topic | To become familiar with DW Design |
| Objectives | * To learn conceptual and logical DW Design |

|  |
| --- |
| Lab Discussion: Theoretical concepts and Procedural steps |

**DW:**  OLTP systems or operational databases were not particularly better suited for answering business queries so in order to cater to this requirement, data warehouses were created. A data warehouse is nothing else but a very large database that contains data which is specifically structured for querying and analysis of one’s business activities. The data warehouse usually contains a copy of transactional data stored over time and data from other sources as well. Such data in data warehouse doesn’t undergo deletion. In light of these facts, Bill Inmon, the father of data warehouse has defined it formally as:  
  
**“Data warehouse is subject-oriented, integrated, non-volatile and time variant collection of data which is used in support of business management decisions.”**

**DW DESIGN:** DW design is specified at the following levels similar to an OLTP system design:

1. **Conceptual Design:** Conceptual design of DW system involves the transformation of requirements into a conceptual model which involves identifying the subjects and dimensions required for the organization. It is done with the aid of Multi-dimensional (M/ER) ER diagrams (to be discussed ahead).
2. **Logical Design:** Logical design basically confirms the subject areas identified in conceptual design. This involves defining all possible facts (also known as cubes) and their appropriate dimensions along with their granularity levels (for hierarchies).

1. **Physical Design:** Physical design is the actual implementation of the logical design into the system memory with the help of any DW product and can be done in two fundamental ways:

* Relational
* Multi-dimensional

This lab would be covering the first two DW design levels.

**Conceptual Design Constructs: M/ER Diagram:**

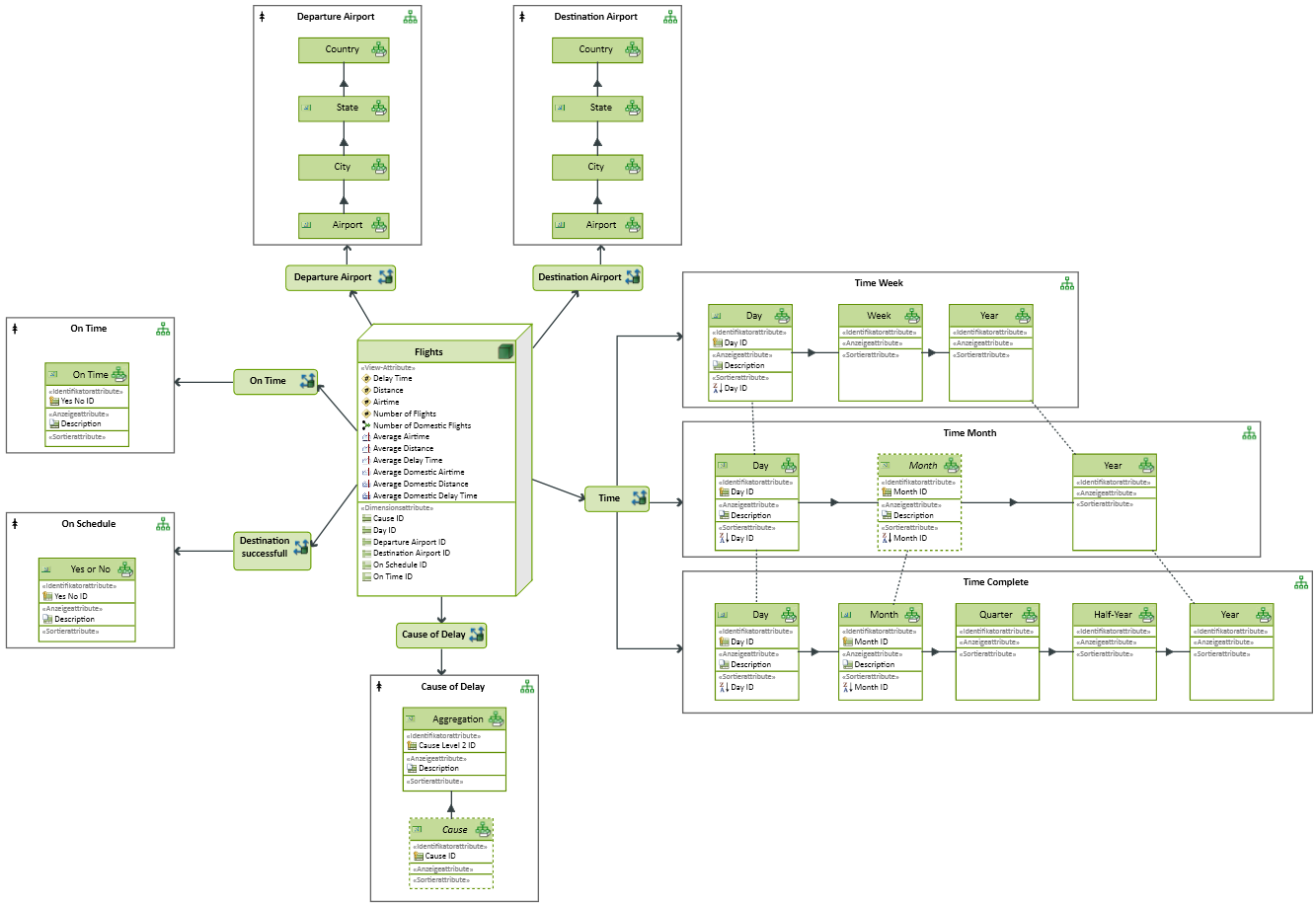
**M/ER Diagram:** It is an extension of the conventional ERD and represents fundamentals to model multidimensional data. The components of an M/ER model are as follows:

1. **Fact node:** It is the focus of interest and includes data that can be measured. It is represented by a cube.
2. **Dimension node:** Dimensions are stored along with facts and identify the parameters of interest for the business, thus they contain data that helps to measure the fact values. It is represented by a box with measures in it describing the dimensions.
3. **Hierarchies:** Showing relationships between various dimension nodes. It is represented by classification arrows:

Sample M/ERD is shown below:



**Logical Design:** Logical design is similar to an M/ERD but with additional information incorporated into it such as granularity level represented by dimension hierarchies.



**Steps:**

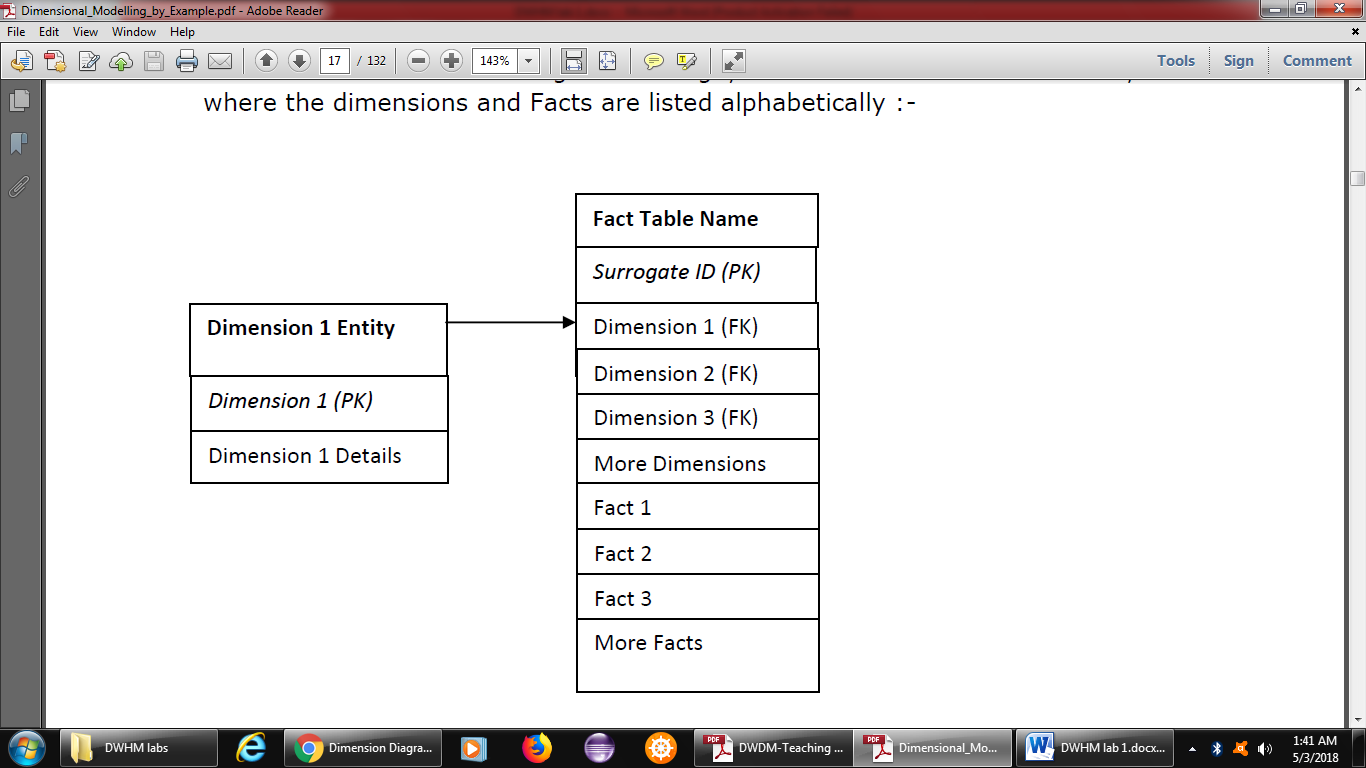
1. **Design Guidelines**: Follow the four step approach:

* Establish the user requirements.
* Determine the grain of the data
* Identify the Dimensions
* Identify the Facts

1. **Always use surrogate keys** they are essential for joining data in Fact Tables and Dimension Tables.
2. **Use conformed dimensions** where applicable: Conformed Dimensions are shared between Tables and must have the same values in order for Data to be retrieved satisfactorily.

For example, date-stamped data in two tables must all be at the same level of granularity – for example, Days, Weeks or Months.

1. **Logical model template:** Follow a generic design, based on Dimensions and Facts, where the dimensions and Facts are listed alphabetically.



**Tools:** M/ERD can be created with a tool downloadable from the following link below:  
  
[www.innovator.de](http://www.innovator.de)   
For help refer to the link below:

<http://help.innovator.de/Enterprise/en-us/Content/I4IA/MultidimensionalModeling.htm>

You have to select the “Measures diagram” option because it contains the cubes as well as the dimensions. For help on how to create a measures diagram in innovator, use the links below:

<http://help.innovator.de/Enterprise/en-us/Content/Ref.MetaM/MeasureDimDia.htm>

<http://help.innovator.de/Enterprise/en-us/Content/I4IA/OLAPMeasureDimDiagram.htm>

|  |
| --- |
| Lab Tasks |
| Submission Date: |

1. Draw M/ERDs for the following scenarios represented by ERDs and business rules to ease your task.
2. Scenario 1: Advertising



**Business Rules**

1. Advertisements are placed by Agencies.

2. Advertisements are placed by Channels, such as TV or Print.

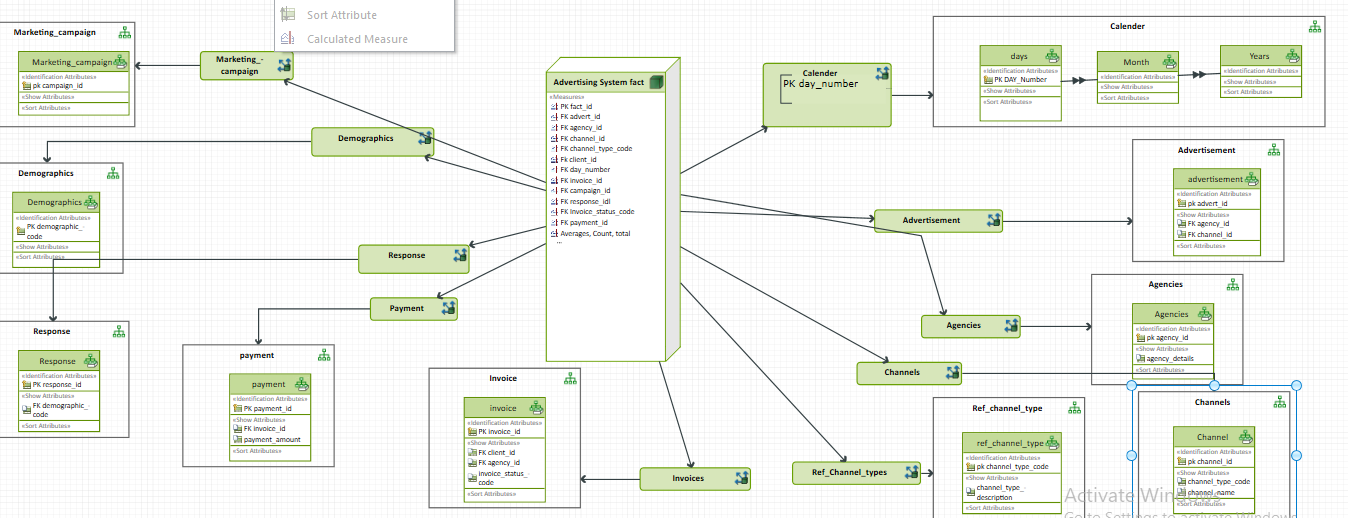
3. Advertisements can be part of a Marketing Campaign.

4. Invoices are produced and result in Payments.

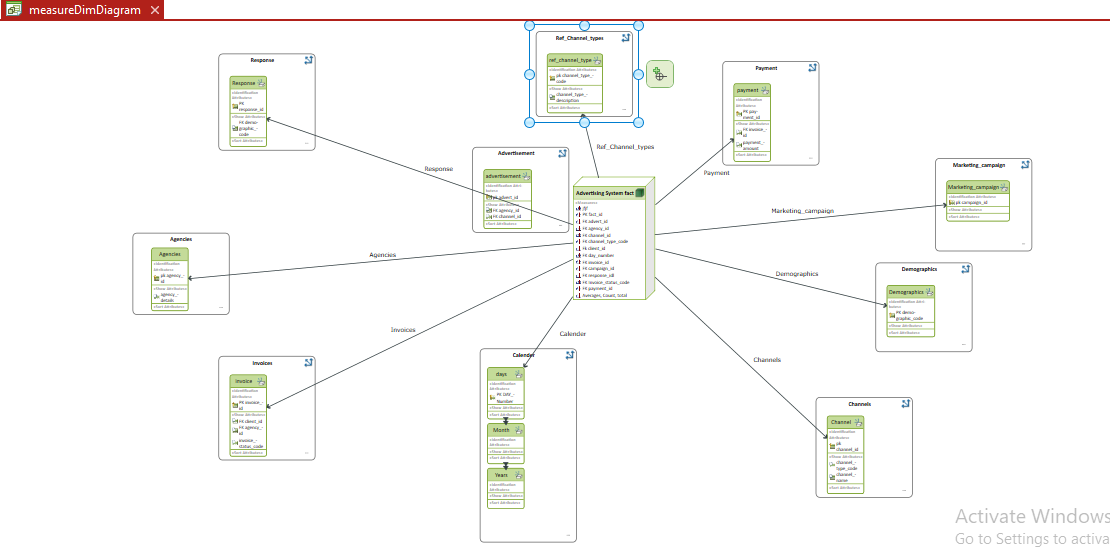
5. Advertisements result in Responses.

6. Responses can be analysed by Demographics.

1. **M/ERD for advertisement scenario(Cube Diagram):**



1. **M/ERD for advertisement scenario(Measure Dimension Diagram):**

****

b)Scenario 2: Amusement Parks



**Business Rules**

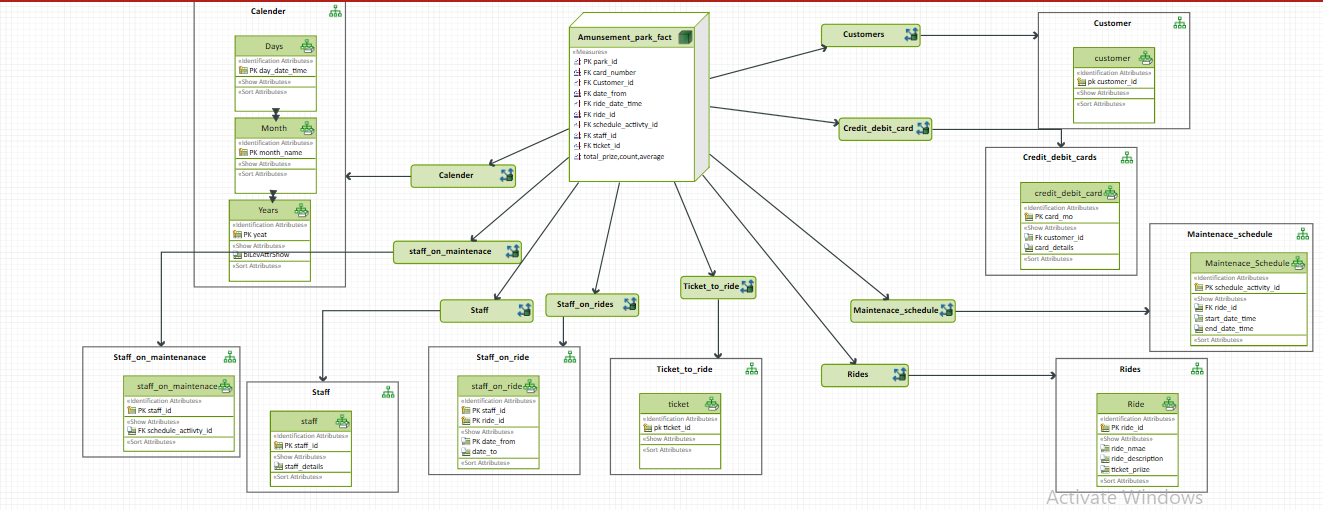
1. Rides are available.

2. Customers use Credit or Debit Cards to buy Tickets to Ride.

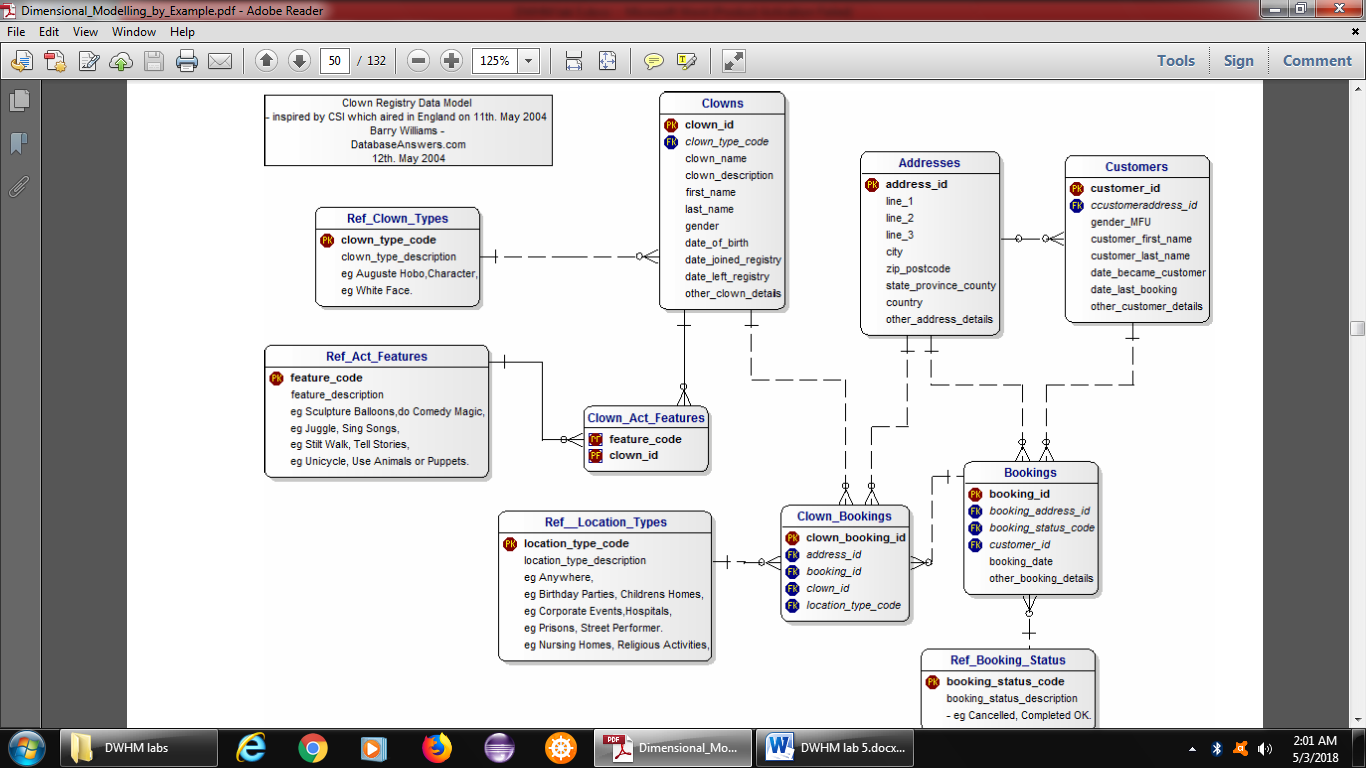
3. Staff operate the Rides.

4. Staff also work on Maintenance Schedules for the Rides.

1. **M/ERD FOR AMUNSEMENT PARK SCENARIO:**



1. Scenario 3: Clown registry



**Business Rules**

1. People registered as Clowns are stored in the Clowns Table.

2. A Clown can have Acts with specific Features.

3. Clowns accept Bookings to perform at specific Locations for specific Customers.

1. **M/ERD for clown registry scenario:**

