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

|  |  |  |  |
| --- | --- | --- | --- |
| Course: SWE324 - Data Warehousing and Data Mining | | | |
| Instructor | Rabeea Jaffari | **Practical/Lab No.** | 02 |
| Date | 09 April 2019 | **CLOs** | CLO-4: P3 & P4 |
| Signature |  | **Assessment Score** | 1 Marks |

|  |  |
| --- | --- |
| Topic | To become Familiar with OLTP System Implementation |
| Objectives | * To learn physical Database Design |

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

**OLTP Physical Design:** Physical design is the actual implementation of the tables created in logical design into the system memory with the help of any database product such as ORACLE, MYSQL, SQLSERVER etc. Thus, this design is DBMS dependent.

This lab would be covering the last OLTP design level.

**OLTP PRODUCT:** Any suitable OLTP product can be used to implement the OLTP system. Available choices are:

* MYSQL
* ORACLE EXPRESS
* SQL SERVER etc.

**GUIDELINES FOR IMPLEMENTATION:**

After choosing an appropriate OLTP product for implementation while implementing any OLTP system’s physical design, following guidelines should be considered:

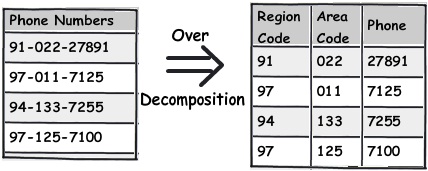
**RULE 1:What is the nature of the application (OLTP or OLAP)?**

**For OLTP, use a normalized table design.**

**RULE 2: Break your data into logical pieces, make life simpler**

This rule is actually the first rule from 1st normal form. One of the signs of violation of this rule is if your queries are using too many string parsing functions like substring, charindex, etc., then probably this rule needs to be applied.

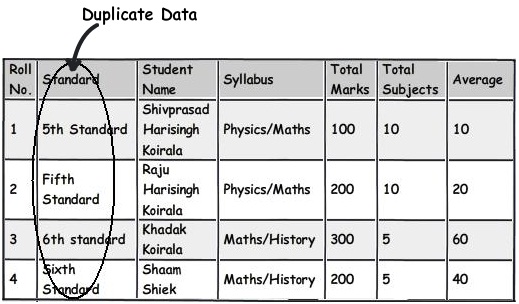
**RULE 3:Do not get overdosed with rule 2**



Over decomposition not required (until your application demands it) since it leads to complexities.

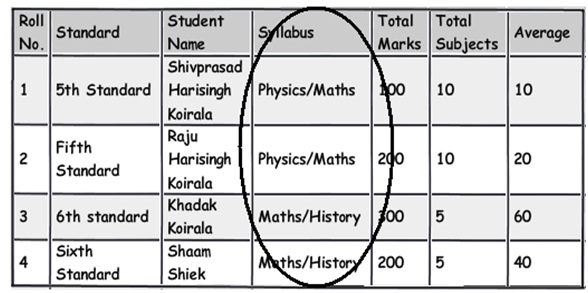
**Rule 4:Treat duplicate non-uniform data as your biggest enemy**

Duplicate data creates confusion.For instance, in the below diagram, you can see “5th Standard” and “Fifth standard” means the same.

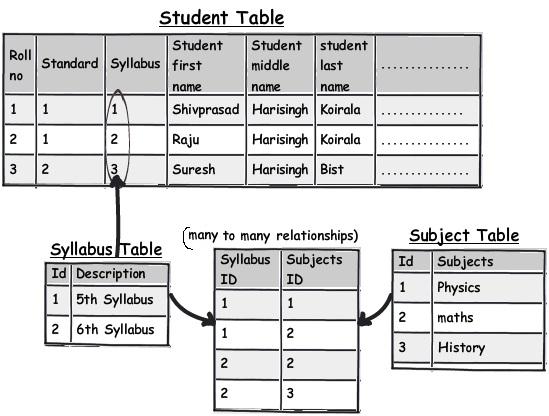


**Rule 5:Watch for data separated by separators**

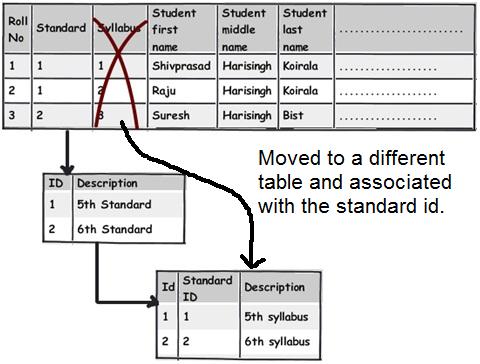
The second rule of 1st normal form says avoid repeating groups.If you see the syllabus field closely, in one field we have too much data stuffed. These kinds of fields are termed as “Repeating groups”. If we have to manipulate this data, the query would be complex and also I doubt about the performance of the queries.



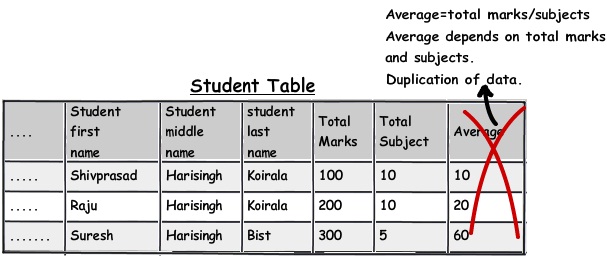
These kinds of columns which have data stuffed with separators need special attention and a better approach would be to move those fields to a different table and link them with keys for better management.



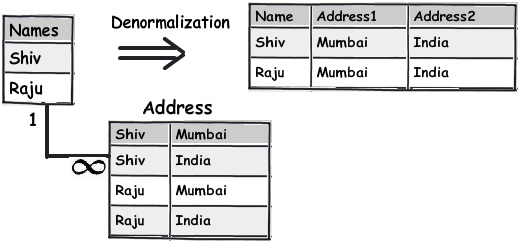
**Rule 6: Watch for partial dependencies**



**Rule 7: Choose derived columns preciously**

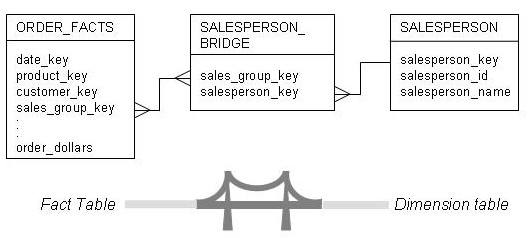


**Rule 8: Do not be hard on avoiding redundancy, if performance is the key**

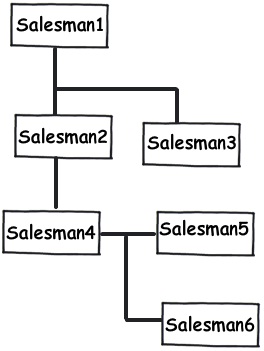


**Rule 9: Bridge table design**

This involves using bridge table for a many-to-many relationship between the tables.



**Rule 10: For unlimited hierarchical data self-reference PK and FK**



**Rule 11: Naming conventions should be appropriate.**

While naming table columns use meaningful names starting with alphabets and try to use lowercase letters as they are referenced without any conflicts in the queries, also try to avoid blank spaces between column names and use and underscore (\_) separator for convenience instead.

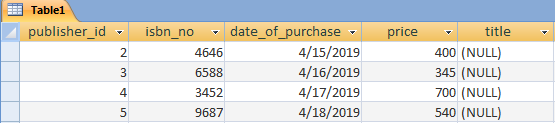
|  |
| --- |
| Lab Tasks |
| Submission Date: 16-04-19 |

1. For each of the descriptions below, perform the following tasks:

i. Physically implement the systems in any suitable OLTP product with proper keys, data types and relationships.

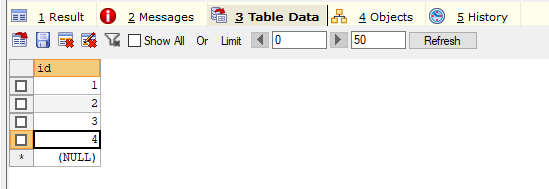
ii. Insert dummy data upto 5 rows.

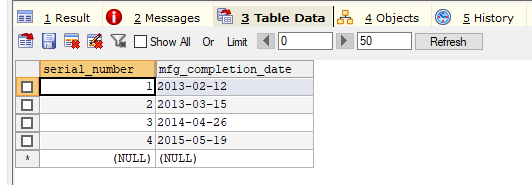
1. A book is identified by its ISBN number, and it has a title, a price, and a date of publication. It is published by a publisher, which has its own ID number and a name. Each book has exactly one publisher, but one publisher typically publishes multiple books over time.

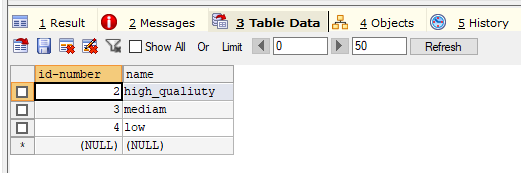
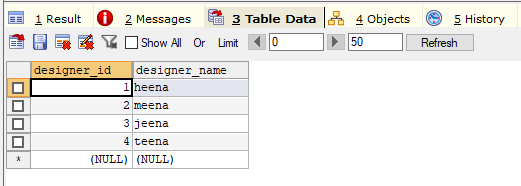




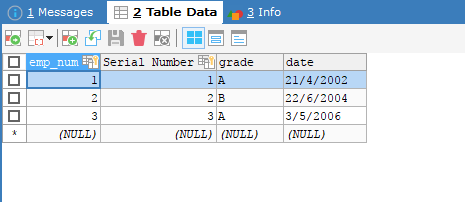
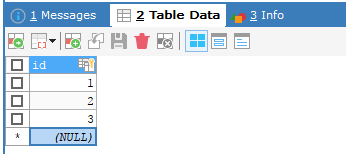
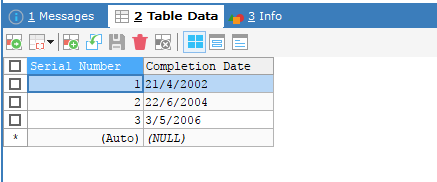
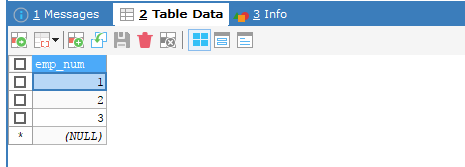
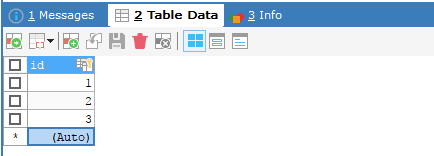
1. A piano manufacturer wants to keep track of all the pianos it makes individually. Each piano has an identifying serial number and a manufacturing completion date. Each instrument represents exactly one piano model, all of which have an identification number and a name. In addition, the company wants to maintain information about the designer of the model. Over time, the company often manufactures thousands of pianos of a certain model, and the model design is specified before any single piano exists.



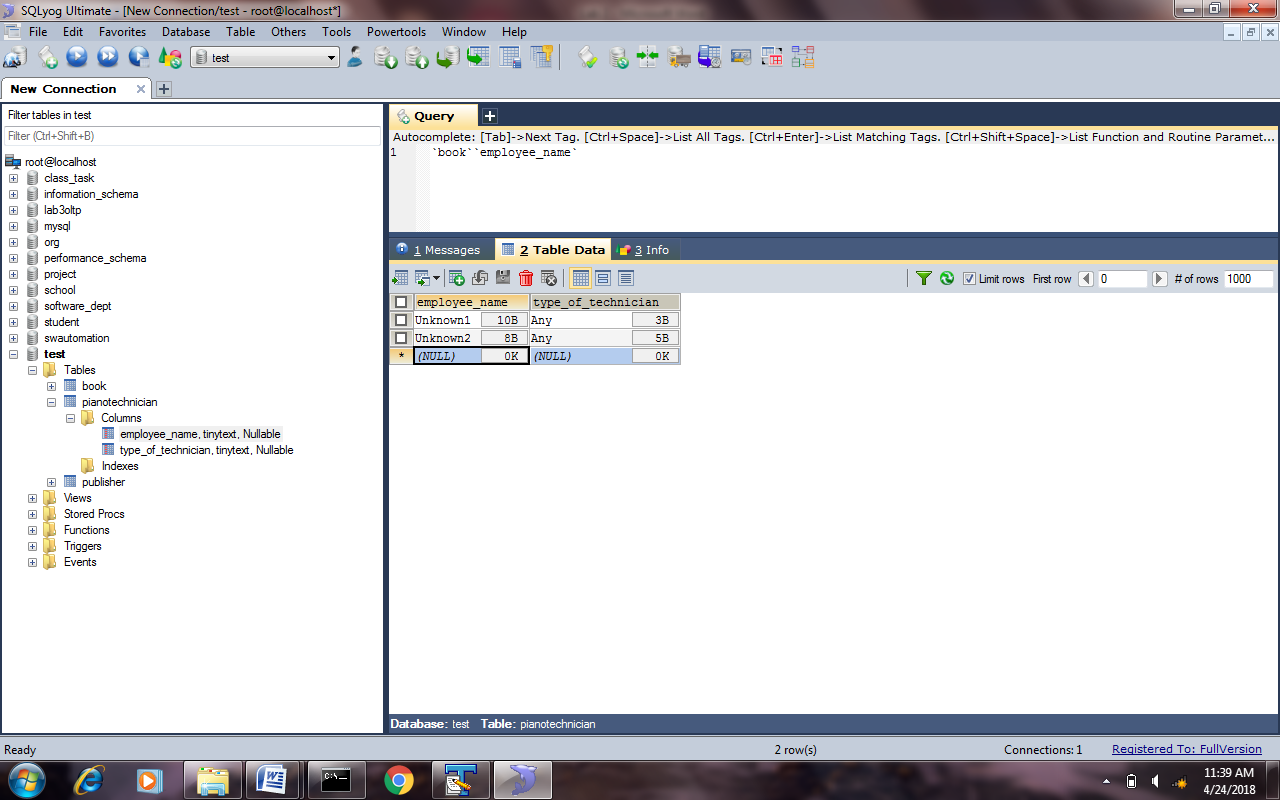




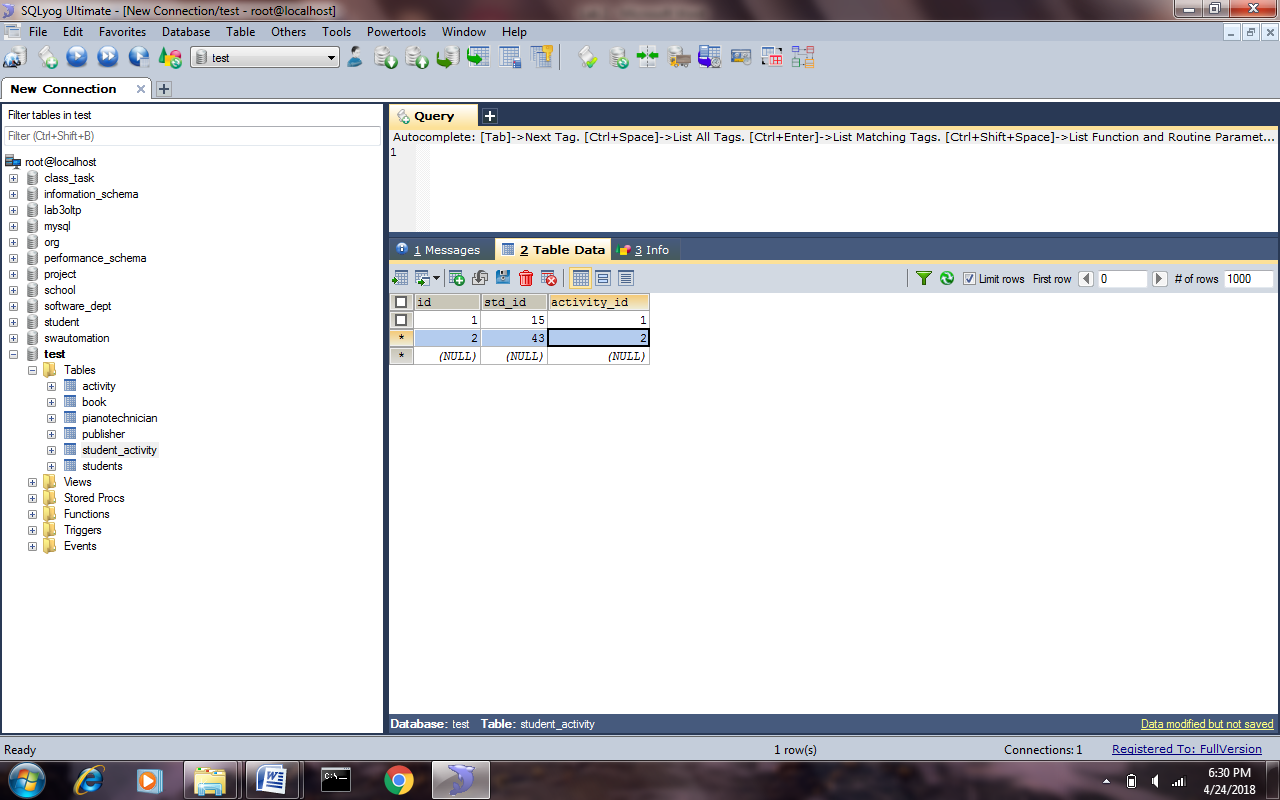
1. A piano manufacturer (see above) employs piano technicians who are responsible for inspecting the instruments before they are shipped to the customers. Each piano is inspected by at least two technicians (identified by their employee number). For each separate inspection, the company needs to record its date and a quality evaluation grade.

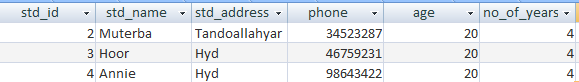
  

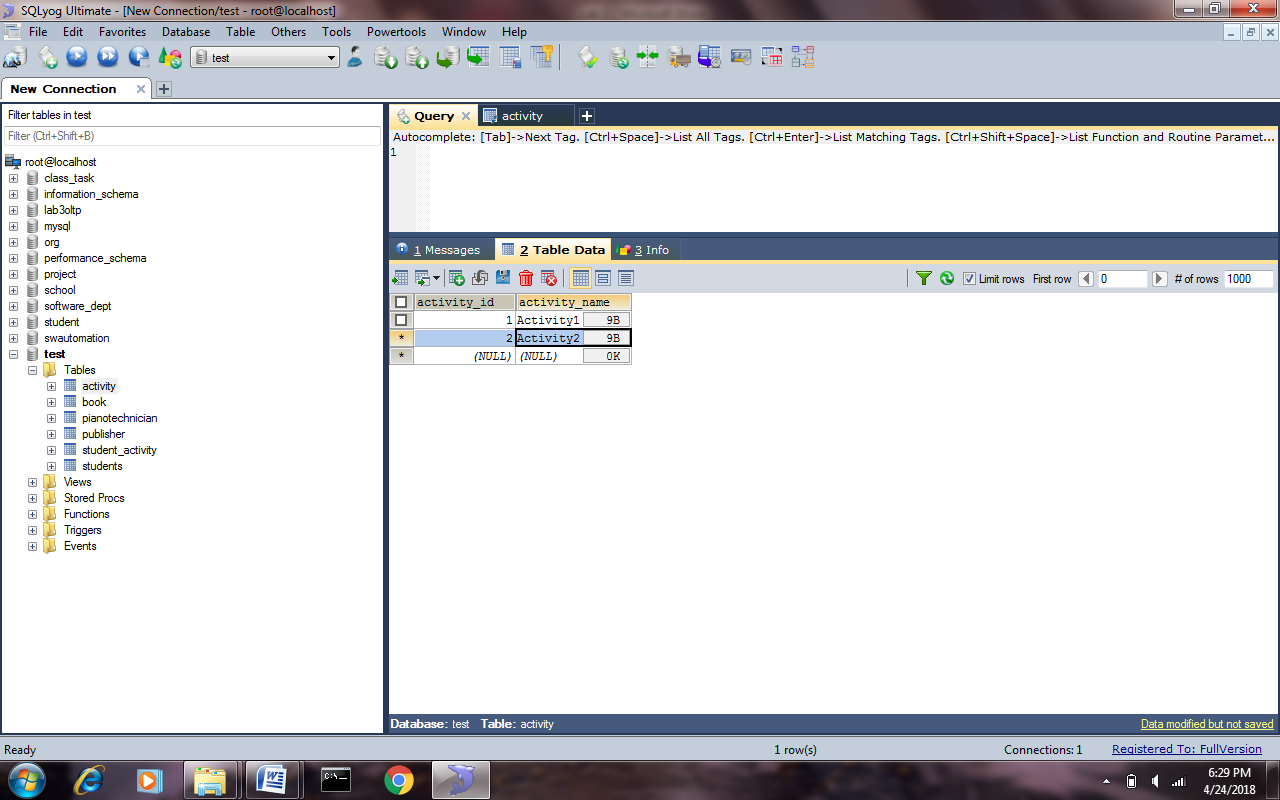
1. The piano technicians (see above) have a hierarchy of reporting relationships: Some of them have supervisory responsibilities in addition to their inspection role and have multiple other technicians report to them. The supervisors themselves report to the chief technician of the company.



1. The entity type STUDENT has the following attributes: Student Name, Address, Phone, Age, Activity, and No of Years. Activity represents some campus-based student activity, and No of Years represents the number of years the student has engaged in this activity. A given student may engage in more than one activity. Draw a logical ERD for this situation. What attribute or attributes did you designate as the identifier for the STUDENT entity? Why?







1. A college course may have one or more scheduled sections or may not have a scheduled section. Attributes of COURSE include Course ID, Course Name, and Units. Attributes of SECTION include Section Number and Semester ID. Semester ID is composed of two parts: Semester and Year. Section Number is an integer (such as 1 or 2) that distinguishes one section from another for the same course but does not uniquely identify a section. How did you model SECTION? Why did you choose this way versus alternative ways to model SECTION?

