Homework 2 RDB Design

**Professor**: Jeremy Harper **TAs**: Bhavana Dhonvan, [Latha Sree Rayala](https://iu.instructure.com/courses/2133191/users/6568879)

**Student Name**: Deepak Rajput **Program**: MS Applied Data Science

Index Table:

|  |  |  |
| --- | --- | --- |
| Sr. No. | Homework Requirement | Page number |
| 1 | Import the data into mysql | 1-2 |
| 2 | Draw the ERD for the database | 3 |
| 3 | Using the database design language to list all the tables, relationships, and attributes, including the PK, FK constraints. | 3-6 |
| 4 | Explain your design rationales. | 7 |

1. **Import the data into mysql**

MySQL code for create relational database tables from provided file.

Load given file via table insert wizard, then created below tables

Code file 

use database\_week1;

create table cases

select distinct case\_id,

submitter\_id,

project\_id,

age\_at\_index,

days\_to\_birth,

days\_to\_death

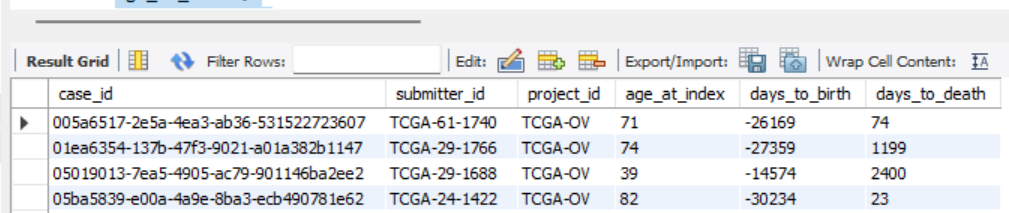
from tcga\_ovariancancerdata;

ALTER TABLE `database\_week1`.`cases`

CHANGE COLUMN `case\_id` `case\_id` VARCHAR(255) NOT NULL ,

ADD PRIMARY KEY (`case\_id`);

select \* from cases;



create table demographic\_details

select distinct demographic\_id,

gender,

race,

vital\_status,

year\_of\_birth,

year\_of\_death

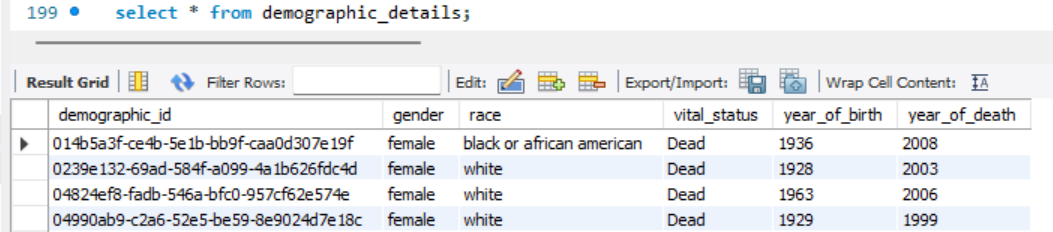
from tcga\_ovariancancerdata;

ALTER TABLE `database\_week1`.`demographic\_details`

CHANGE COLUMN `demographic\_id` `demographic\_id` VARCHAR(255) NOT NULL ,

ADD PRIMARY KEY (`demographic\_id`);

select \* from demographic\_details;



create table diagnosis\_details

select distinct diagnosis\_id,

age\_at\_diagnosis,

days\_to\_last\_follow\_up,

figo\_stage,

icd\_10\_code,

morphology,

primary\_diagnosis,

site\_of\_resection\_or\_biopsy,

synchronous\_malignancy,

tissue\_or\_organ\_of\_origin

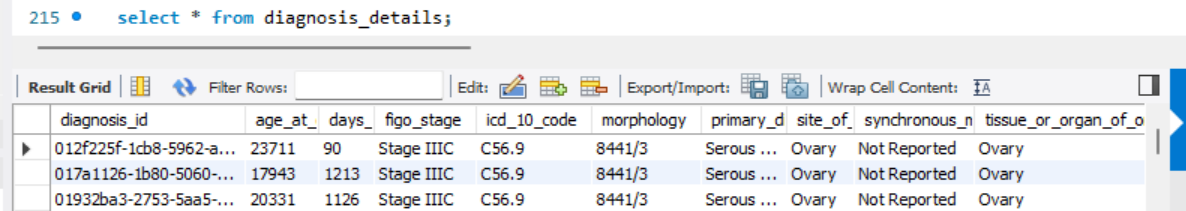
from tcga\_ovariancancerdata;

ALTER TABLE `database\_week1`.`diagnosis\_details`

CHANGE COLUMN `diagnosis\_id` `diagnosis\_id` VARCHAR(255) NOT NULL ,

ADD PRIMARY KEY (`diagnosis\_id`);

select \* from diagnosis\_details;



create table treatment\_details

select distinct treatment\_id,

treatment\_or\_therapy,

treatment\_type,

year\_of\_diagnosis

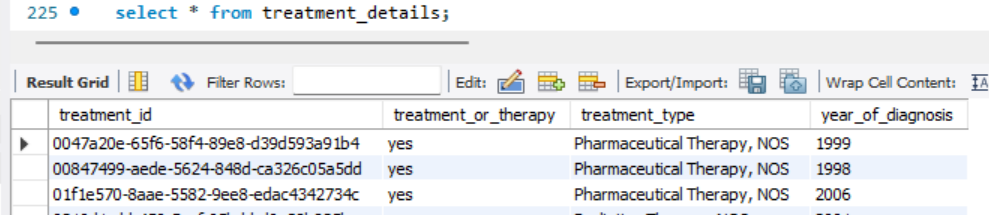
from tcga\_ovariancancerdata;

ALTER TABLE `database\_week1`.`treatment\_details`

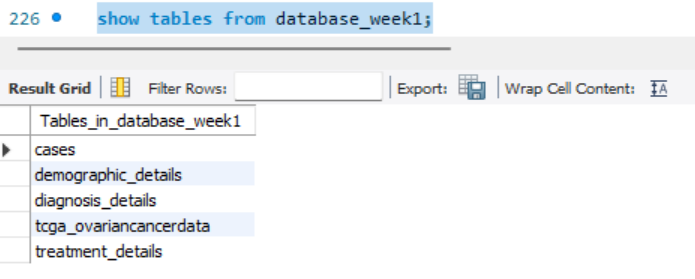
CHANGE COLUMN `treatment\_id` `treatment\_id` VARCHAR(255) NOT NULL ,

ADD PRIMARY KEY (`treatment\_id`);

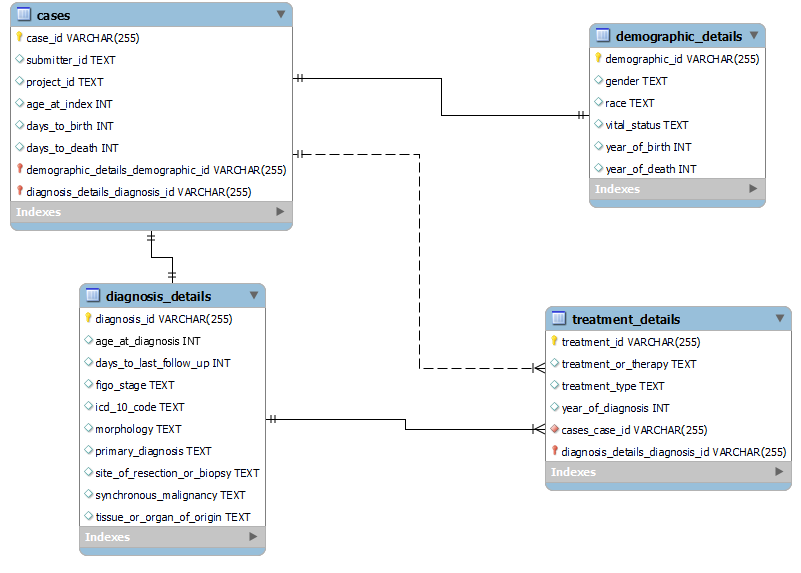
select \* from treatment\_details;



Snaps of tables:



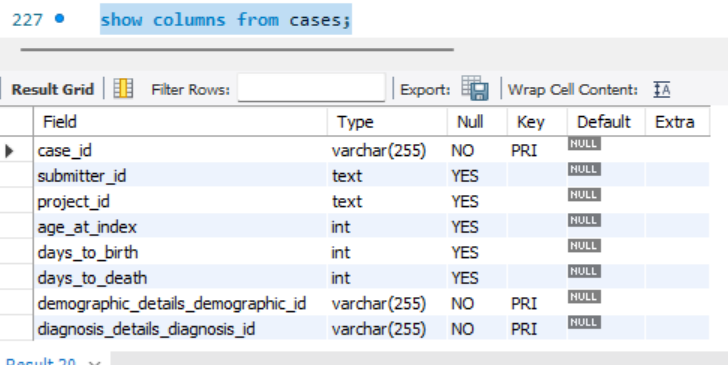
**2. Draw the ERD for the database.**

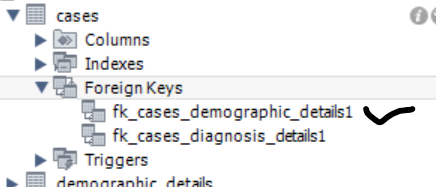


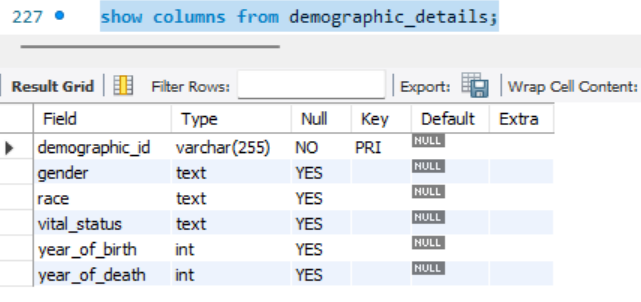
1. **Using the database design language to list all the tables, relationships, and attributes, including the PK, FK constraints.**
2. **Tables:**  Segregated given data in four different tables name as below:

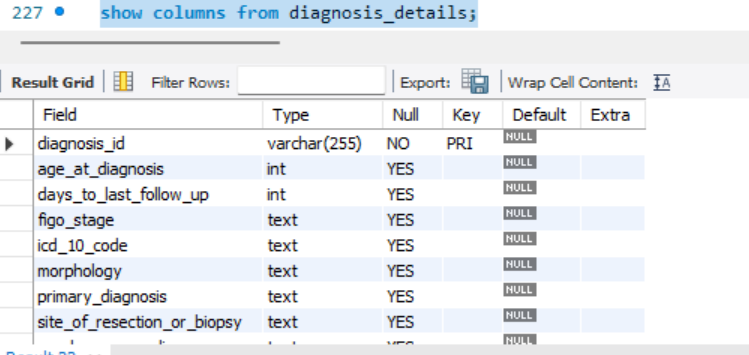
cases, demographic\_details, diagnosis\_details and treatment\_details

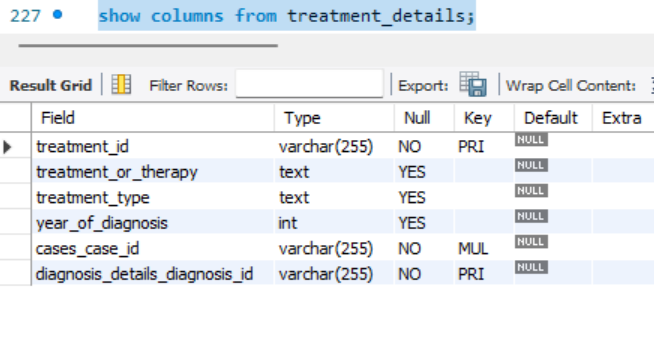
segregation is design according to the uniqueness of attributes which define the entity and its nature. It also saves storage and avoid duplication of entries in the data.

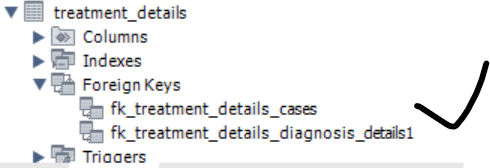












1. **Relationship for the database designs given below**

**One to one identifying**

Between cases and demographic\_details

Between cases and diagnosis\_details

**One to many identifying**

Between diagnosis\_details and treatment\_details

**One to Many non-identifying**

Between cases and treatment\_details

1. **Attributes including PK, FK constraints are given in below excel diagram**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Entity name : cases** | | |  |  | **Entity name : demographic\_details** | | |
| Attributes | case\_id | PK |  |  | Attributes | demographic\_id | PK |
| submitter\_id | |  | | --- | |  | | |  |  | gender |  |
| project\_id | One to One | | race |  |
| age\_at\_index |  |  | vital\_status |  |
| days\_to\_birth |  |  | year\_of\_birth |  |
| days\_to\_death |  |  | year\_of\_death |  |
|  | demographic\_id | FK |  |  |  |  |  |
|  | diagnosis\_id | FK |  |  |  |  |  |
|  | one to one |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  | One to Many |  |  |
| **Entity name : diagnosis\_details** | | |  |  |  |  |  |
| Attributes | diagnosis\_id | PK |  |  | **Entity name : treatment\_details** | | |
| age\_at\_diagnosis |  |  |  | Attributes | treatment\_id | PK |
| days\_to\_last\_follow\_up | one to many | | treatment\_or\_therapy |  |
| figo\_stage |  |  | treatment\_type |
| icd\_10\_code |  |  | year\_of\_diagnosis |
| morphology |  |  |  | case\_id | FK |
| primary\_diagnosis |  |  |  | diagnosis\_id | FK |
| site\_of\_resection\_or\_biopsy |  |  |  |  |  |
| synchronous\_malignancy |  |  |  |  |  |
| tissue\_or\_organ\_of\_origin |  |  |  |  |  |
| treatment\_id |  |  |  |  |  |

1. **Explain your design rationales.**

Student response:

Segregated given database in 4 different Tables and Entities cases, demographic\_details, diagnosis\_details and treatment details to have a clear understanding and relationship between all these entities.

All selected attributes well defined the entity and maintains uniqueness among all entities.

Role of primary key as case\_Id, demographic\_id, diagnosis\_id and treatment\_id are selected for their uniqueness, never changing and not null property.

Foreign elements are defined in table mentioned in ERD diagram and table given above provide required relationship and overlook all require data among tables & database.

By designing the same we can store, analysis, retrieve and represent data efficiently and save storage spare.