Homework 2

1. (5) Identify an application that will allow you to create ER diagrams, download and install it.
   1. The lecture had a couple of suggestions for applications that students have found reasonably easy to learn, and that are free, including: dbdiagram.io, draw.io, SQLDBM and QuickDBD and LUCIDchart. LUCIDchart has a slightly more resource-rich version that starts at $7.95. If you already know how to use Visio or some other system that is fine, just note what it is here and what your experience has been so far.

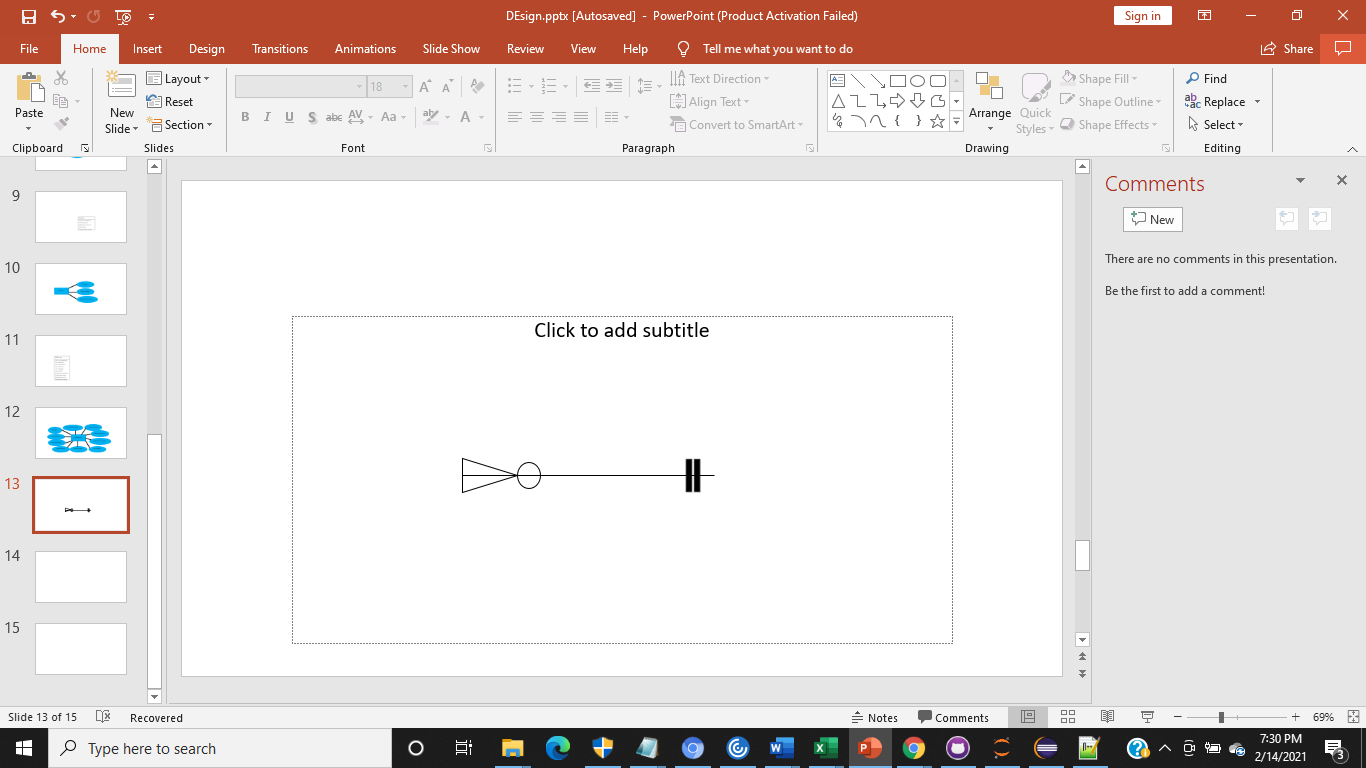
Application to be used:

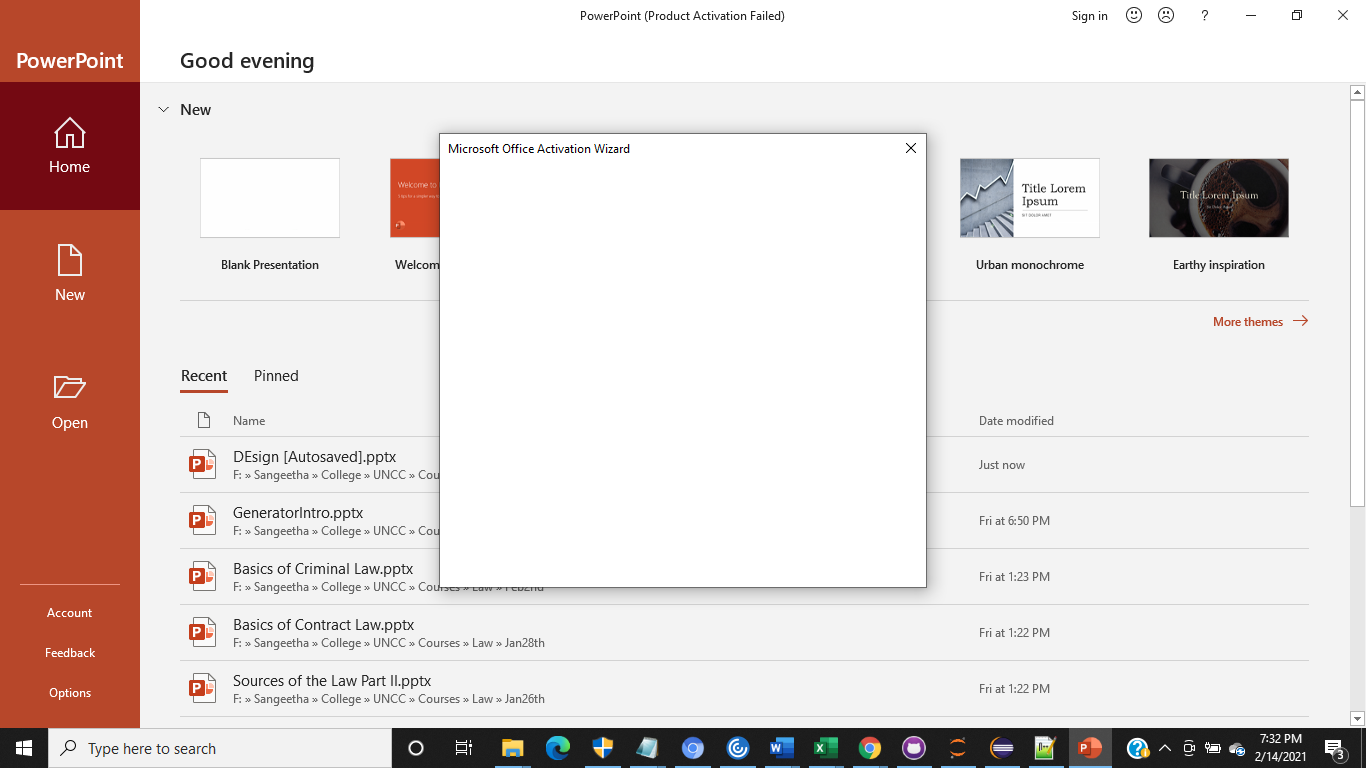
Power Point

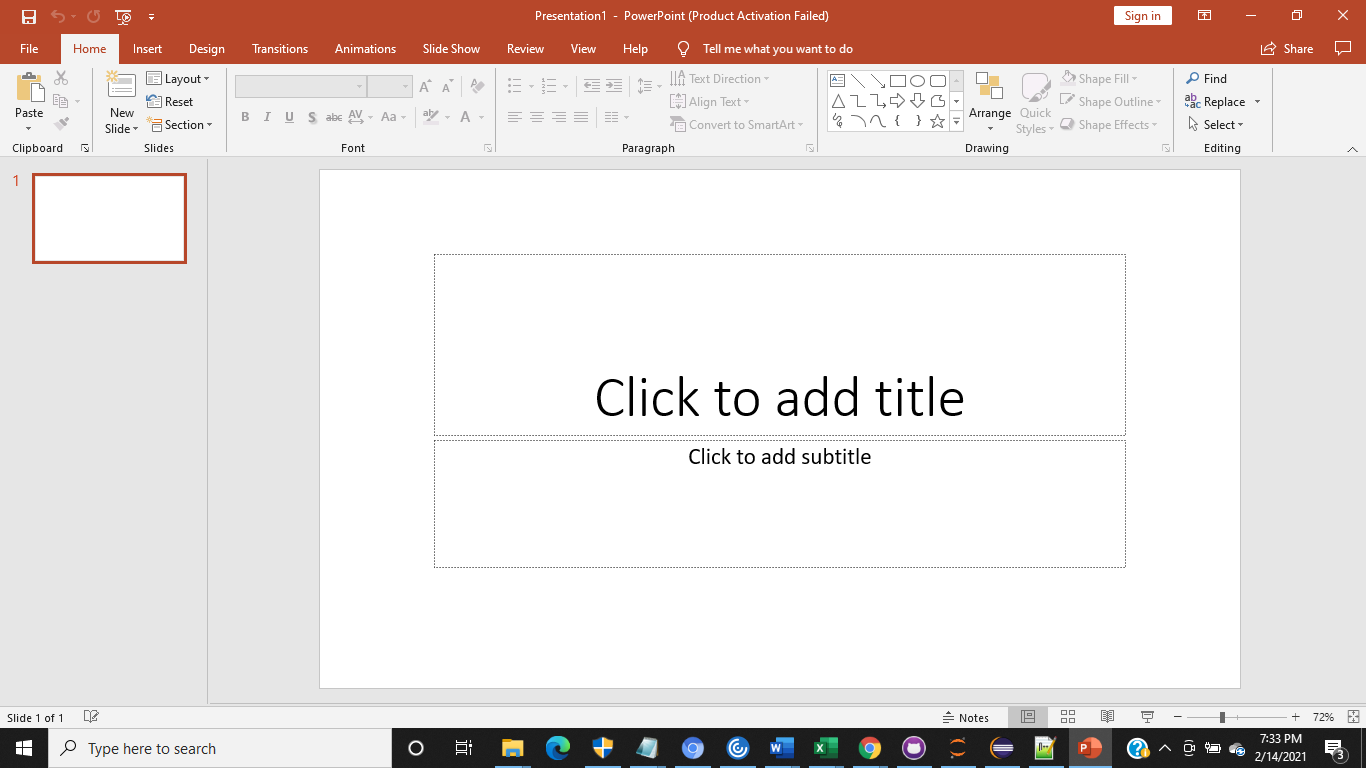
If used before, explain how much experience with it you have had:

I have 14 years of experience in software development. I have prepared my project presentations in power point for management reporting. I have used it every once in 3 months.

Image of installed app, open at first entry page:





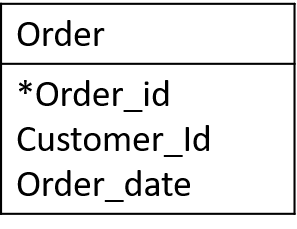
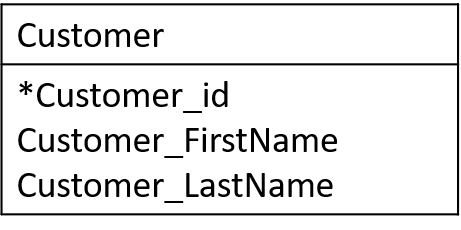


* 1. If you choose to simply use Powerpoint, create a set of the standard representations (ovals, boxes, double-edged versions, connectors, etc.) so that you can cut and paste them when you create models. The lecture provided a link to a cheat-sheet for each style (Chen and IE). Take a picture of that page(s) and paste below.

IE Style:

|  |
| --- |
| Probe |
| \*Probe\_label  \*location  \*nt\_sequence  Attachement\_type |

|  |
| --- |
| Probe\_ID |
| Gene\_name\*  Probe\_nt\_sequence |



Chen Style

Probe\_ID

Probe

Order

Customer

1

places

M

Customer

Places/is placed

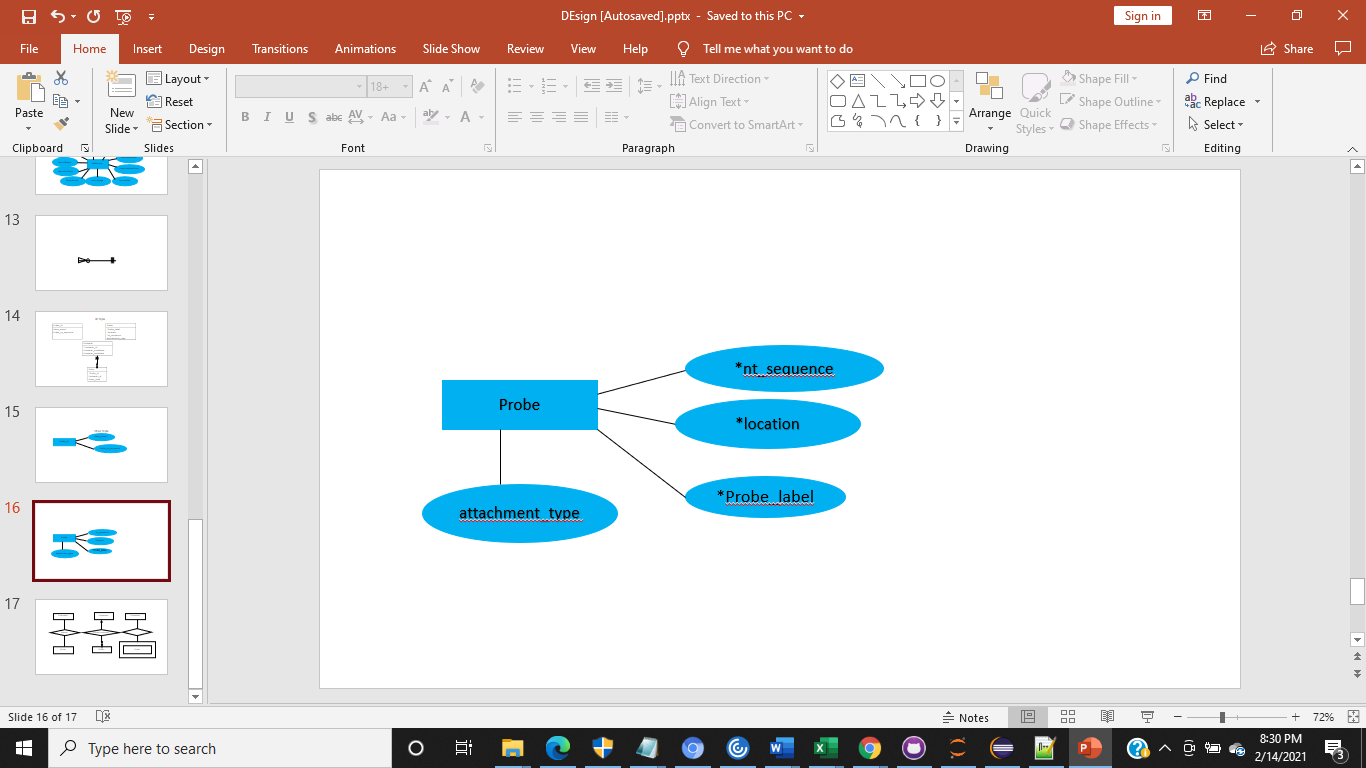
Order

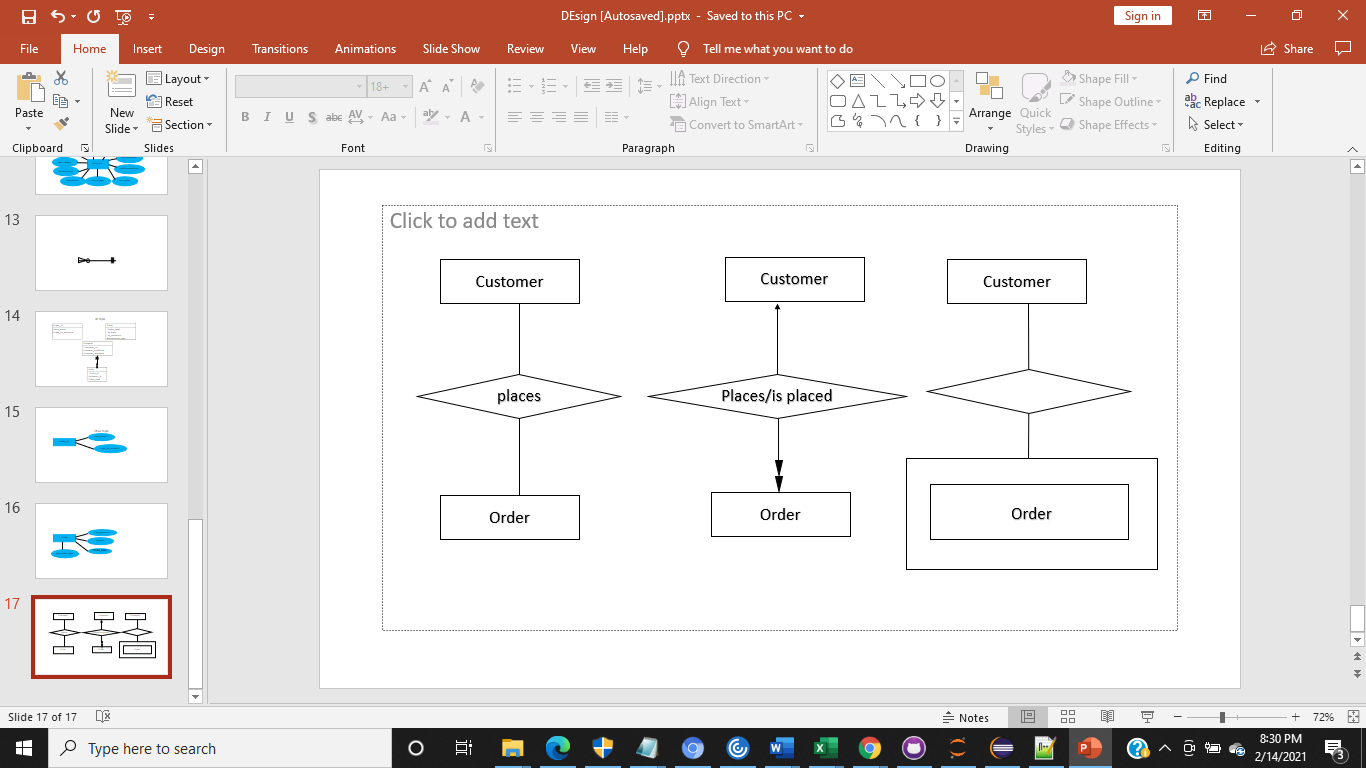
Customer

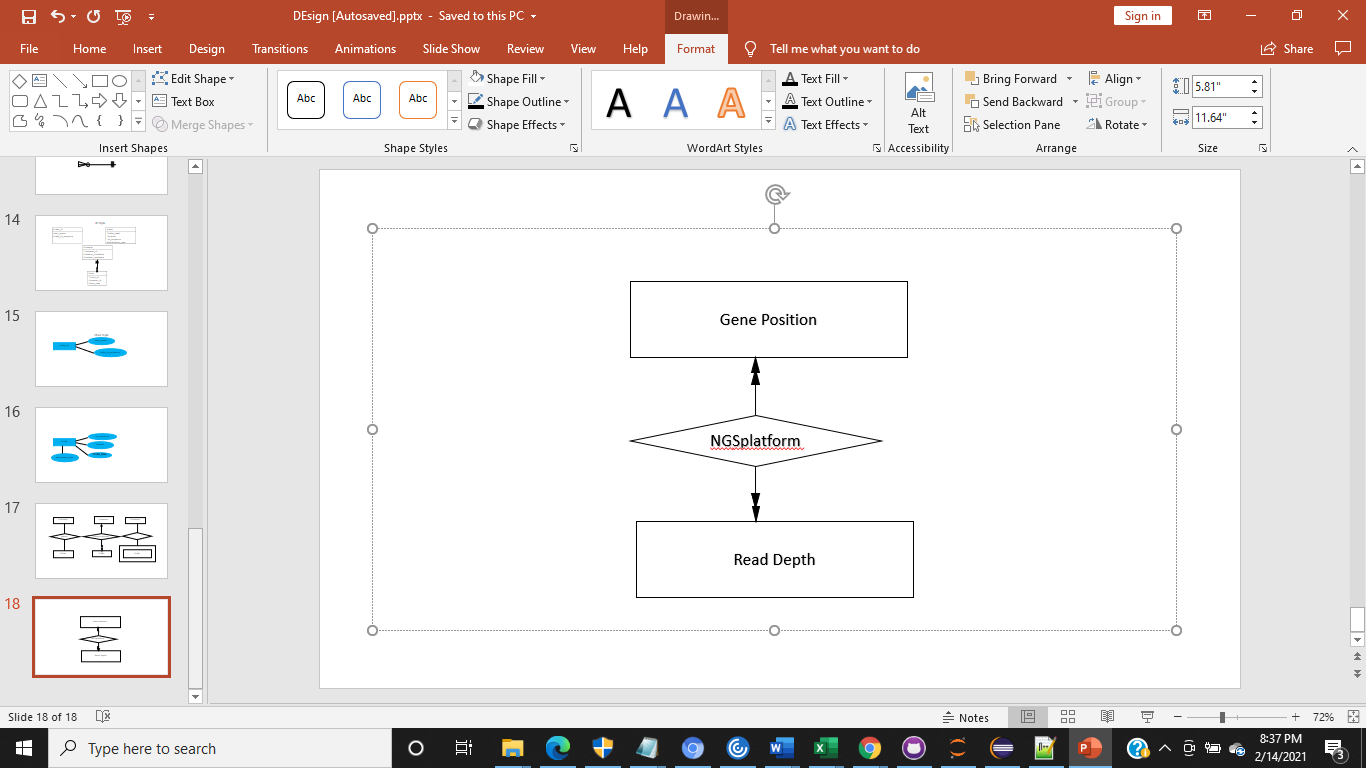
1 is placed by

M places

Order







1. (20) Practice using the application (10 points for Chen version, 10 points for IE version).

Please provide a graphical image of the following, one for each of the two representation modes.

Entity: Cluster // has attributes: Cluster\_id (a unique identifier), size (units are nucleotides), maxDistance (units are nucleotides)

IE Version:

|  |  |
| --- | --- |
| **Cluster** | |
| **PK/Char**  Num  Num | **cluster\_id**  size  maxDistance |

Size and max distance units are nucleotides

Chen Version:

Cluster

Size and max distance units are nucleotides

Entity: Test // has attributes: Test\_id (a unique identifier), gap (units are nucleotides), SequenceVariant (position + substitution), intraDistance (nucleotide units), PositionHead (genome position Chromosome#+location), PositionTail (genome position Chromosome# + position), lengthMin(nucleotide units), lengthMax (nucleotide units), Specs (link to a protocol – text blob).

IE Version:

|  |  |
| --- | --- |
| **Test** | |
| **PK/Char**  Num  Char  Num  Char  Char  Num  Num  Char | test\_id  gap  sequenceVariant  intraDistance  positionHead  positionTail  lengthMin  lengthMax  specs |

Gap, intraDistance, lengthMin, and lengthMax units are nucleotides

Chen Version:

Test

Entity: miRNA // has attributes: miRNA\_id (a unique identifier), GB\_AccessionID (candidate identifier), GuideSequence (string limited to A,C,G,T), GS\_length (nucleotide units, length <120), PassengerSequence(string limited to A,C,G,T) , PS\_length (nucleotide units)

IE Version:

|  |  |
| --- | --- |
| **miRNA** | |
| **PK/Char**  Char  Char  Num  Char  Num | miRNA\_id  gb\_AccessionID  guideSequence  gs\_length  passengerSequence  ps\_length |

GS\_length and PS\_length

units are nucleotides

Chen Version:

miRNA

GS\_length and PS\_length units are nucleotides

Entity: Species // has attributes: Tax\_abbrev (CHAR, limited to list in Taxonomy ontology, unique identifier), SpeciesName (candidate key, limited to list in Taxonomy ontology, genus<dot>species), CommonName (free text, may be null), Phylum (limited to Taxonomy ontology), Class (limited to Taxonomy ontology), Order (limited to Taxonomy ontology), Family (limited to Taxonomy ontology).

IE Version:

|  |  |
| --- | --- |
| **Species** | |
| **PK/Char**  Char  Char  Char  Char  Char  Char | tax\_abbrev  speciesName  commonName  phylum  class  order  family |

Chen Version:

species

Entity: Disease // has attributes: ICD\_ID (references classification using International Classification of Diseases, the unique identifier), DiseaseName (From ICD, potential alternate unique identifier), DiseaseDescription (free text).

IE Version:

|  |  |
| --- | --- |
| **Disease** | |
| **PK/Char**  Char  Char | ICD\_ID  diseaseName (unique)  diseaseDescription |

Chen Version:

Disease

Entity: Reference // has attributes: DOI\_ID (unique identifier is the digital object identifier), JournalName (text), JournalVolume (int), JournalIssue (may be month or number), JournalPage (INT – could be a range to cover first and last page), JournalRef (ISSN is a referenced catalog of journal identifiers), PublicationAuthors (multi-valued text), PublicationDate (day-month-year is composite), PublicationTitle (Varchar), PublicationLength(Computed INT)

IE Version:

|  |  |
| --- | --- |
| **Reference** | |
| **PK/Char**  Char  int  Num  int  Char  Char  Date  Char  int | DOI\_ID  journalName  journalVolume  journalIssue  journalPage  journalRef  publicationAuthors  publicationDate  publicationTitle  publicationLength |

Chen Version:

Reference

Relationships: each Test (an assay that measures expression of miRNA genes) includes a Cluster of miRNA genes that are within a certain distance of one or more disease-associated protein-coding genes. There must be at least one assay for a Cluster to exist. A Cluster is existence-dependent on the Test, and this is a strong relationship. One Test may have many Clusters, a given Cluster belongs to many Tests. Each Cluster contains one to many miRNA genes, and each miRNA gene can be included in zero to many Clusters. Each miRNA comes from one and only one Species, but each Species has one to many miRNAs. Each miRNA is tested against zero to many Diseases and each Disease can be affected by zero to many miRNAs. The References are related to both miRNA reports (many to many) and to Disease reports (many-to-many).

|  |  |
| --- | --- |
| **Test** | |
| **PK/Char**  Num  Char  Num  Char  Char  Num  Num  Char | **test\_id**  gap  sequenceVariant  intraDistance  positionHead  positionTail  lengthMin  lengthMax  specs |

|  |  |
| --- | --- |
| **Cluster** | |
| **PK(char)**  Num  Num  **PK** | **cluster\_id**  size  maxDistance  **test\_id** |

has (strong relationship)

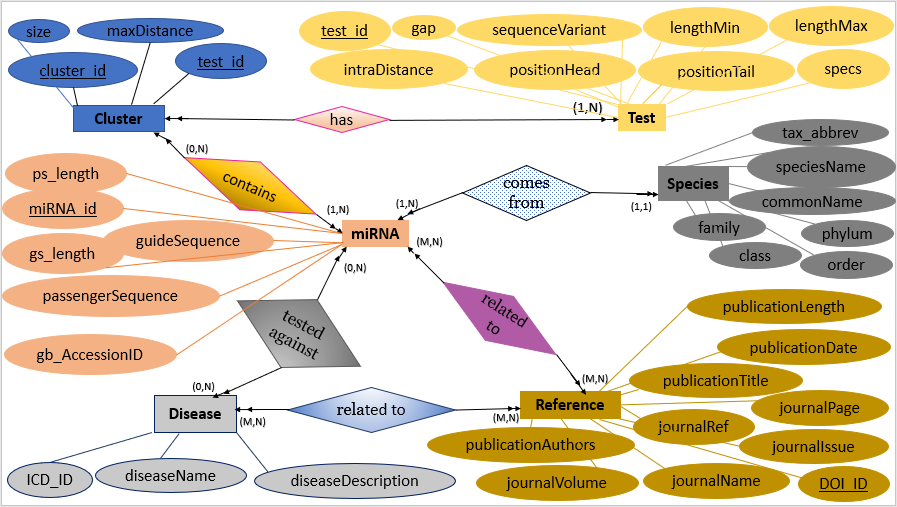
contains

|  |  |
| --- | --- |
| **Species** | |
| **PK/Char**  Char  Char  Char  Char  Char  Char | tax\_abbrev  speciesName  commonName  phylum  class  order  family |

|  |  |
| --- | --- |
| **miRNA** | |
| **PK/Char**  Char  Char  Num  Char  Num | miRNA\_id  gb\_AccessionID  guideSequence  gs\_length  passengerSequence  ps\_length |

|  |  |
| --- | --- |
| **Reference** | |
| **PK/Char**    Char  int  Num  int  Char  Char  Date  Char  int | DOI\_ID  journalName  journalVolume  journalIssue  journalPage  journalRef  publicationAuthors  publicationDate  publicationTitle  publicationLength |

|  |  |
| --- | --- |
| **Disease** |  |
| **PK/Char**  Char  Char | ICD\_ID  diseaseName (unique)  diseaseDescription |



Not handled – propose a solution (conceptual level). The mRNA gene whose regulation by miRNA is investigated in the assay is not modeled, and there should be a way to indicate that only miRNA that are expressed in a sample can affect the Disease state.

Proposed solution for mRNA, miRNA and Disease

|  |  |
| --- | --- |
| **mRNA** | |
| **PK/Char**  **FK/Char**  Num  Num  **FK/Char** | **mRNA\_id**  miRNA\_Id  sequenceMergeStart  sequenceMergeEnd  **ICD\_ID** |

|  |  |
| --- | --- |
| **Disease** | |
| **PK/Char**  Char  Char | **ICD\_ID**  diseaseName(unique)  diseaseDescription |

|  |  |
| --- | --- |
| **miRNA** | |
| **PK/Char**  Char  Char  Num  Char  Num | miRNA\_id  gb\_AccessionID  guideSequence  gs\_length  passengerSequence  ps\_length |