The University of Texas at Austin
McCombs School of Business

MIS 381N – INTRO. TO DATABASE MANAGEMENT

Logical Design

Tayfun Keskin
Visiting Clinical Professor, The University of Texas at Austin, McCombs School of Business
Associate Teaching Professor, University of Washington Seattle, Foster School of Business

QUESTIONS

Any questions

before we begin …

HOMEWORK 1 - DB DESIGN

- You will design a hotel reservation system
- Use an ER diagramming tool (Lucidchart, Visio, Draw.io)
- Label entities/tables clearly with logical names
- Divide data properly (avoid redundancy)
- Identify the primary and foreign keys (mark as PK or FK)
- You must use Crow's Foot method
- **Save as PDF and submit**

AGENDA

Lecture	Detailed Models
Hands-on exercises	From a spreadsheet From a memo
Looking Forward	Homework 1 Quiz

REVIEW QUESTION

What is an ER Diagram?

ENTITY RELATIONSHIP DIAGRAM

- A diagram that uses tables, fields and relationships to plan a database

QUESTION

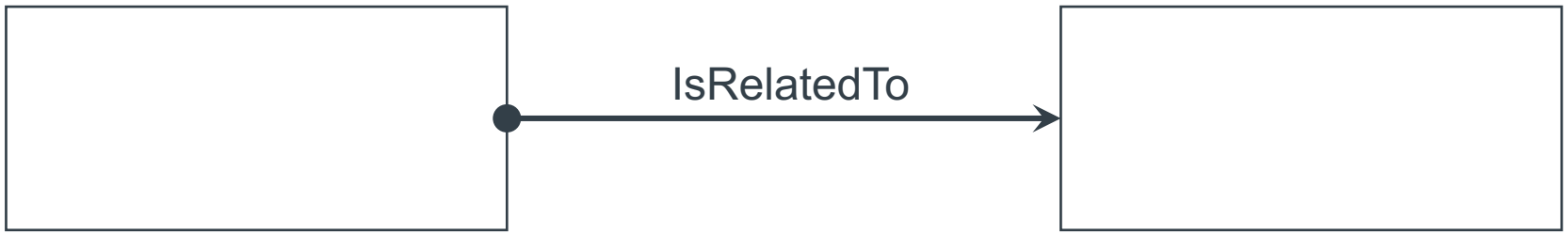
Do we have to use a software for ER diagramming?

NAMING TABLES/FIELDS

- Use the plural name of the entity… i.e.: Customers
- Name attributes with upper camel case… i.e.: CamelCase
- Avoid spaces and special characters
- Do not make the assumption that anything for a data element is unique unless otherwise proven (go with a synthetic PK)

NAMING RELATIONSHIPS

- Verb
- Capital start letter
- Cardinality indicated at the ends of the line



RECAP

- A “model” is a representation of the real world and is built for a specific purpose
- An “entity-relationship” model is a model of the data and the rules that govern the data within a business enterprise
- An entity corresponds to a physical thing in the real-world. Examples: Student, Building, Machine, etc.
- Attributes describe an entity. Example: A Student may be described by their name, gender, and date of birth.

RECAP

- Relationship describe how an entity “relates” to other entities. Example: A student resides in a building and is trained to operate a machine
- An entity maps to a table and attributes map to columns
- Each table should have at least one column that has a unique value across all rows – one such column is designate as a primary key
- When two entities are related, the primary key from one of the corresponding table is a foreign key in the other table

QUESTION

Do terms “**relation and relationship**”

mean the same thing?

HANDS ON EXERCISE 1

A Service Company – From a Speadsheet

CONCEPTUAL DESIGN – 10 MINS

1. What are the entities of the data below (i.e., the tables we'll create)

2. How do these identified entities relate?

EmpName	EmpAddress	EmpPhone	ProjectNbr	ServiceDesc	Cost	ScheduledStart	CompletionDate	ClientName	ClientPhone	Address
Joe	341 Rainy Drive	331-9873	852.1	Plant annuals	$40	4/10/2009	4/10/2009	Lehigh	887-2098	5204 Merri Lane
Joe	341 Rainy Drive	331-9873	852.2	Mow & trim yard	$65	4/10/2009	4/10/2009	Lehigh	887-2098	5204 Merri Lane
Joe	341 Rainy Drive	331-9873	864.1	Trim shrubs	$25	4/10/2009	4/10/2009	Clark	384-0136	1804 Jupiter Drive
Sam	6548 Bell Road	337-4472	880.1	Fertilize yard	$45	4/11/2009	4/11/2009	Lehigh	887-2098	8216 Oak Street
Sam	6548 Bell Road	337-4472	875.1	Mow & trim yard	$50	4/10/2009	4/12/2009	Marvin	382-3176	329 Taylor Road
Beth	84087 48th Street	689-2834	867.1	Trim shrubs	$95	4/11/2009	4/11/2009	Parker	348-9227	62491 Ivy Trail
Beth	84087 48th Street	689-2834	867.2	Fertilize yard	$30	4/11/2009	4/11/2009	Parker	348-9227	62491 Ivy Trail
Mark	1922 Upland Rd.	844-3957	867.3	Plant annuals	$65	4/11/2009	4/12/2009	Parker	348-9227	62491 Ivy Trail
Beth	84087 48th Street	689-2834	870.1	Plant annuals	$35	4/10/2009	4/10/2009	Clark	384-0136	1755 Orange Street
Mark	1922 Upland Rd.	844-3957	870.2	Mow & trim yard	$45	4/11/2009	4/11/2009	Clark	384-0136	1755 Orange Street

HANDS ON EXERCISE 2

A manufacturing company

CREATE AN ERD TO SHOW HOW YOU WOULD TRACK THIS INFORMATION

A manufacturing company produces products. The following product information is stored: product name, product ID and quantity on hand. These products are made up of many components. Each component can be supplied by one or more suppliers. The following component information is kept:

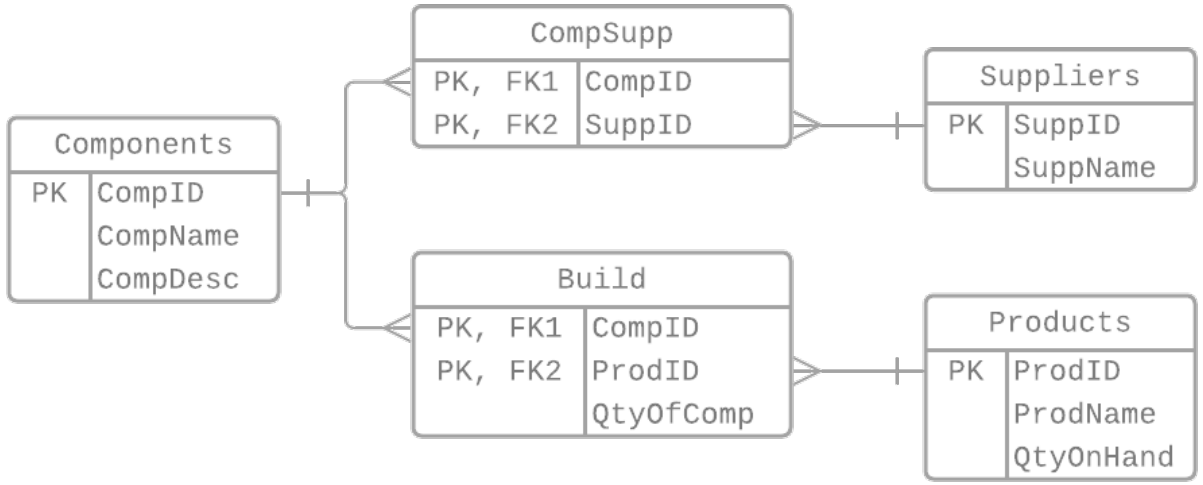
- component ID, name, description, suppliers who supply them, and products in which they are used.

Show entity names, primary keys, attributes for each entity, relationships between the entities and cardinality.

LOGICAL DESIGN

- **Step 1 & 2:** Identify some data elements & subdivide into useful form. Just make assumptions and don't worry about spending too much time on this. List them.
- **Step 3:** Identify the tables (i.e. entities) and assign columns (i.e. attributes) for each. Draw these in a box.
  - TIP: After you identify the tables, can you start to classify the relationships between the tables? This can help you figure out where the FKs will go or the need for bridge tables.
- **Step 4:** What are the primary keys, foreign keys, and remaining relationships?

ERD ANSWER



- Component(CompID, CompName, Description) PK=CompID
- Product(ProdID, ProdName, QtyOnHand) PK=ProdID
- Supplier(SuppID, SuppName) PK = SuppID
- CompSupp(CompID, SuppID) PK = CompID, SuppID
- Build(CompID, ProdID, QtyOfComp) PK= CompID, ProdID

LOOKING FORWARD

Check Canvas… regularly

Read chapters 1 and 9

Homework 1

Quiz 2

The University of Texas at Austin
McCombs School of Business

MIS 381N – INTRO. TO DATABASE MANAGEMENT

Normalization

Tayfun Keskin
Visiting Clinical Professor, The University of Texas at Austin, McCombs School of Business
Associate Teaching Professor, University of Washington Seattle, Foster School of Business

AGENDA

Lecture	Normalization
What's Next	Homework 1 Oracle Server Setup
Quiz	5 questions 15 minutes, on Canvas

QUESTION

Can we immediately design a database that:

- Has no redundant data elements
- Easy to query
- Segmented and granular

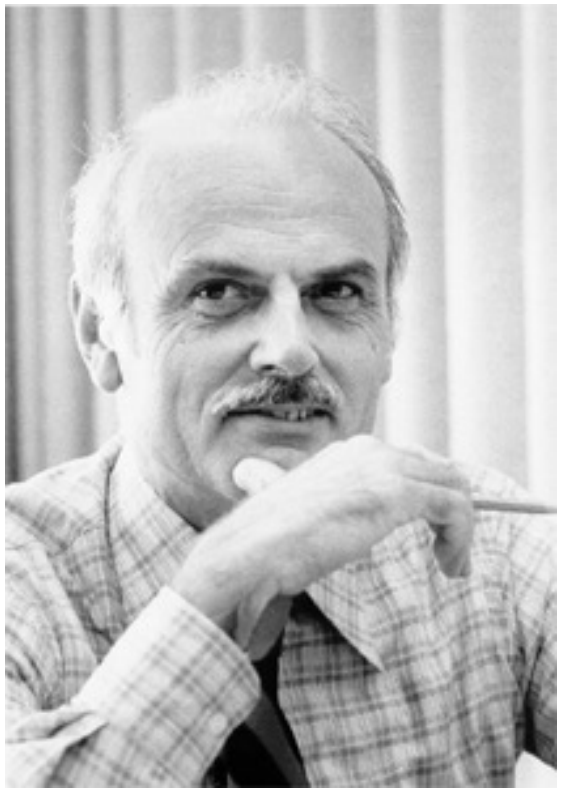
NORMALIZATION

- **Definition 1:** the process of removing redundant data from relational tables by decomposing the tables into smaller tables
- **Definition 2:** the process organizing data in a database
- The goal of normalization is to have relational tables free of redundant data and that can be correctly modified with consistency

NORMAL FORM

- A **normal form** is a process that evaluates each relation against defined criteria and removes the multivalued, joins, functional and trivial dependency from a relation
- If any data is updated, deleted or inserted, it does not cause any problem for database tables and help to improve the relational table's integrity and efficiency.

EDGAR F. CODD



- In the early 1970s Edgar Codd defined three rules for organizing data in a database
- Goal:
  - Reduce redundancy
  - Improve integrity

FIRST THREE NORMAL FORMS

There are six normal forms. The first three are considered critical and most designers stop after that because the costs can outweigh the benefits.

- First Normal Form (1NF) – The value stored at the intersection of each row and column must be a scalar value, and a table must not contain any repeating columns.
- Second Normal Form (2NF) – The primary key should be non-composite and every non-key column must depend on the entire primary key.
- Third Normal Form (3NF) – Every non-key column must depend only on the primary key

THE NEXT FOUR NORMAL FORMS

- Boyce-Codd (BCNF) – A non-key column can't be dependent on another non-key column.
- Fourth (4NF) – A table must not have more than one multivalued dependency, where the primary key has a one-to-many relationship to non-key columns.
- Fifth (5NF) – The data structure is split into smaller and smaller tables until all redundancy has been eliminated.
- Domain-key (DKNF) or sixth (6NF) – Every constraint on the relationship is dependent only on key constraints and domain constraints, where a domain is the set of allowable values for a column.

NORMALIZATION

Benefits:

- Reduce redundant data
- Improve data quality
- Increase the questions that can be answered
- Make the model easier to read and understand

Costs:

- Normalization results in more tables
- Sometimes this leads to
  - A model that is harder to read & understand
  - A database with degraded performance

NORMALIZATION EXAMPLE

- The invoice data with a column that contains repeating values

	VENDOR_NAME	INVOICE_NUMBER	ITEM_DESCRIPTION
1	Cahners Publishing	112897	VB ad, SQL ad, Library directory
2	Zylka Design	97/522	Catalogs, SQL Flyer
3	Zylka Design	97/533B	Card revision

- The invoice data with repeating columns

	VENDOR_NAME	INVOICE_NUMBER	ITEM_DESCRIPTION_1	ITEM_DESCRIPTION_2	ITEM_DESCRIPTION_3
1	Cahners Publishing	112897	VB ad	SQL ad	Library directory
2	Zylka Design	97/552	Catalogs	SQL flyer	(null)
3	Zylka Design	97/553B	Card revision	(null)	(null)

FIRST NORMAL FORM - 1NF

- The invoice data in first normal form

	VENDOR_NAME	INVOICE_NUMBER	ITEM_DESCRIPTION
1	Cahners Publishing	112897	VB ad
2	Cahners Publishing	112897	SQL ad
3	Cahners Publishing	112897	Library directory
4	Zylka Design	97/522	Catalogs
5	Zylka Design	97/522	SQL flyer
6	Zylka Design	97/533B	Card revision

FIRST NORMAL FORM - 1NF

- First normal form with keys added

	INVOICE_ID	VENDOR_NAME	INVOICE_NUMBER	INVOICE_SEQUENCE	ITEM_DESCRIPTION
1	1	Cahners Publishing	112897	1	VB ad
2	2	Cahners Publishing	112897	2	SQL ad
3	3	Cahners Publishing	112897	3	Library directory
4	4	Zylka Design	97/522	1	Catalogs
5	5	Zylka Design	97/522	2	SQL flyer
6	6	Zylka Design	97/533B	1	Card revision

SECOND NORMAL FORM - 2NF

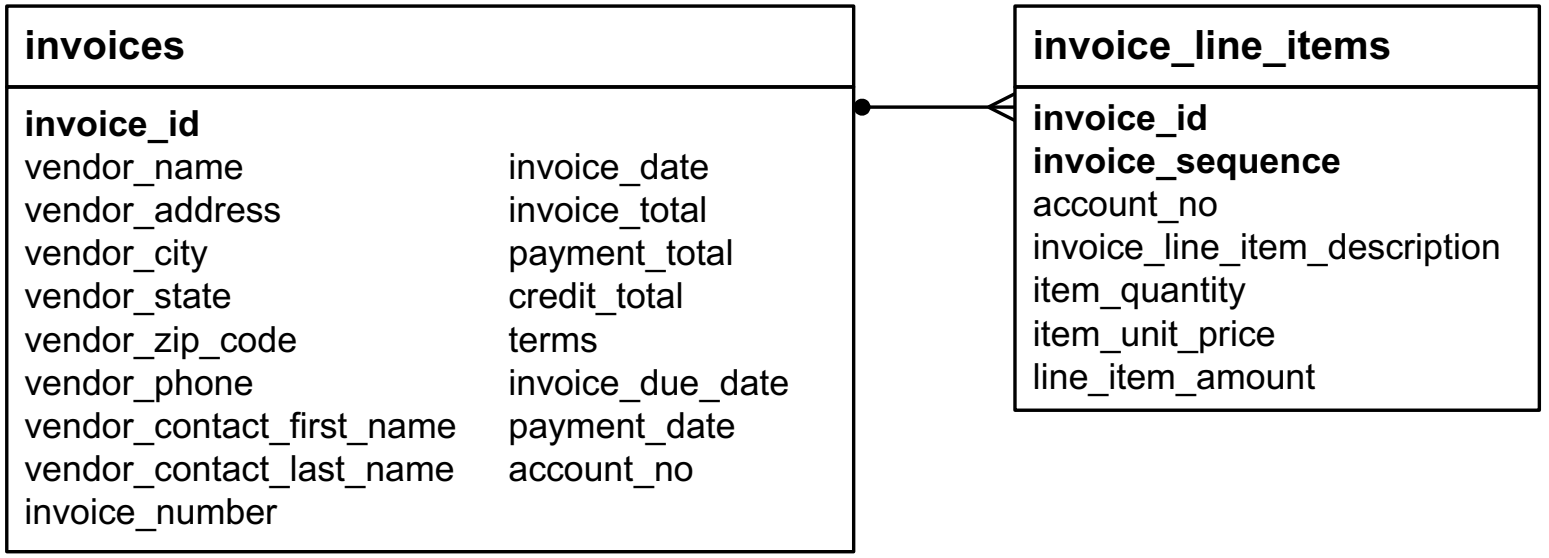
- The invoice data in second normal form

	INVOICE_NUMBER	VENDOR_NAME	INVOICE_ID
1	11287	Cahners Publishing	1
2	97/522	Zylka Design	2
3	97/533B	Zylka Design	3

	INVOICE_ID	INVOICE_SEQUENCE	ITEM_DESCRIPTION
1	1	1	VB ad
2	1	2	SQL ad
3	1	3	Library directory
4	2	1	Catalogs
5	2	2	SQL flyer
6	3	1	Card revision

SECOND NORMAL FORM - 2NF

- The accounts payable system in second normal form

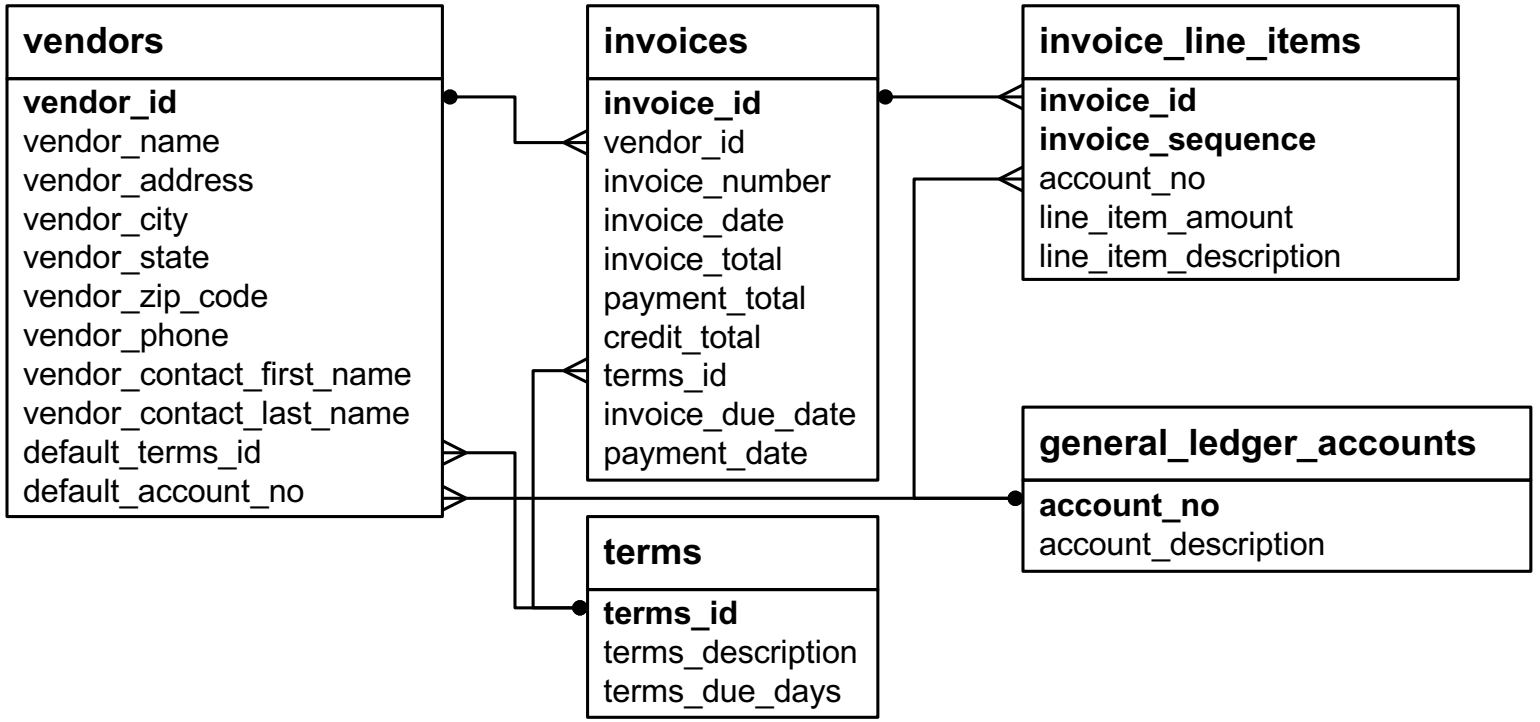


QUESTIONS ABOUT THE STRUCTURE

1. Does the vendor information (vendor_name, vendor_address, etc.) depend only on the invoice_id column?
2. Does the terms column depend only on the invoice_id column?
3. Does the account_no column depend only on the invoice_id column?
4. Can the invoice_due_date and line_item_amount columns be derived from other data?

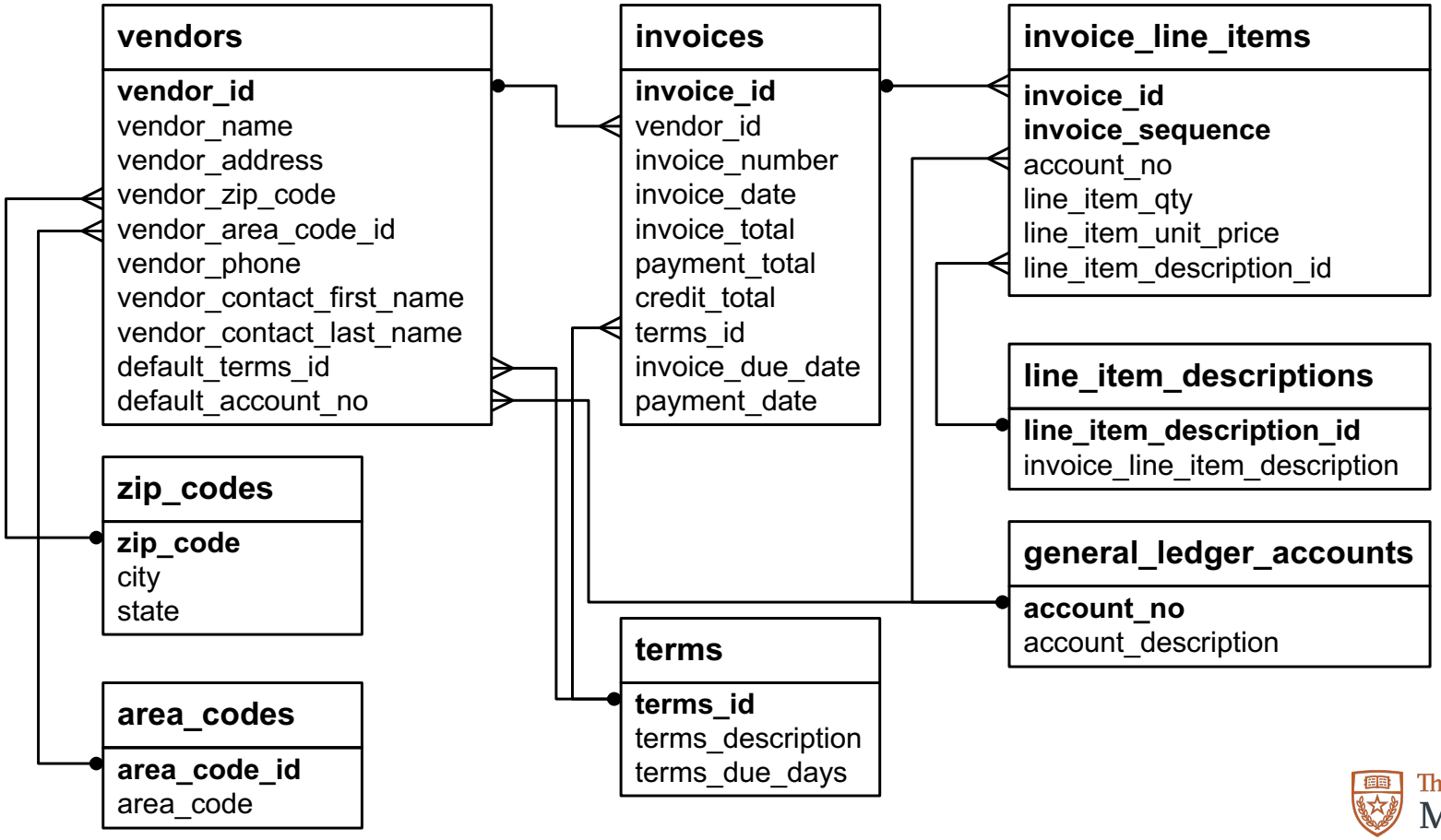
THIRD NORMAL FORM - 3NF

- The accounts payable system in third normal form



FIFTH NORMAL FORM - 5NF

- The accounts payable system in fifth normal form



PROS AND CONS OF NORMALIZATION

Pros:

- Reduce redundant data – leads to increased data integrity
- Increase the number of questions that can be asked
- Segment sensitive data
- Grant granular permission

Cons:

- More difficult for business to understand
- Queries require joins
- Update performance can be reduced
- Insert performance?

QUESTION

Is normalization the ultimate goal for designing/using a database?

DENORMALIZATION

In real-world settings, sometimes it is necessary to denormalize data and accept the disadvantages of data redundancy. When this is done, it is primarily for two reasons:

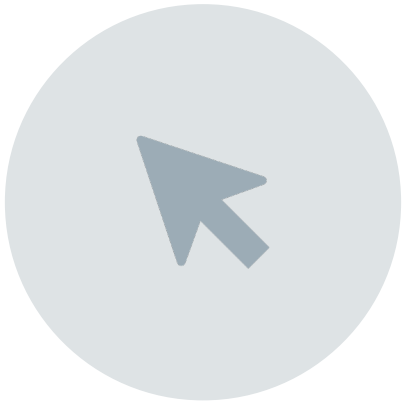
1. Performance – Data is combined in one table so that transactional applications that are reading and updating the data do not need to perform any joins.
2. Analysis – Data is combined in one table so that the data can be analyzed by individuals without knowledge of the structure of the data.

LOOKING FORWARD

- Quiz 2 – don’t go anywhere
- Install SQL Developer and connect to Oracle Server (there is an announcement)
- Homework 1 – due Monday

QUIZ

GO TO CANVAS



CLICK ON QUIZZES

PASSWORD:
****

QUIZ RULES

ALLOWED:

- Book
- Notes
- Assignments

NOT ALLOWED:

- Internet search
- Friends, chat, messaging, phone
- DB expert uncle

THANK YOU