

Chapter 10

How to design a database

Objectives

Applied

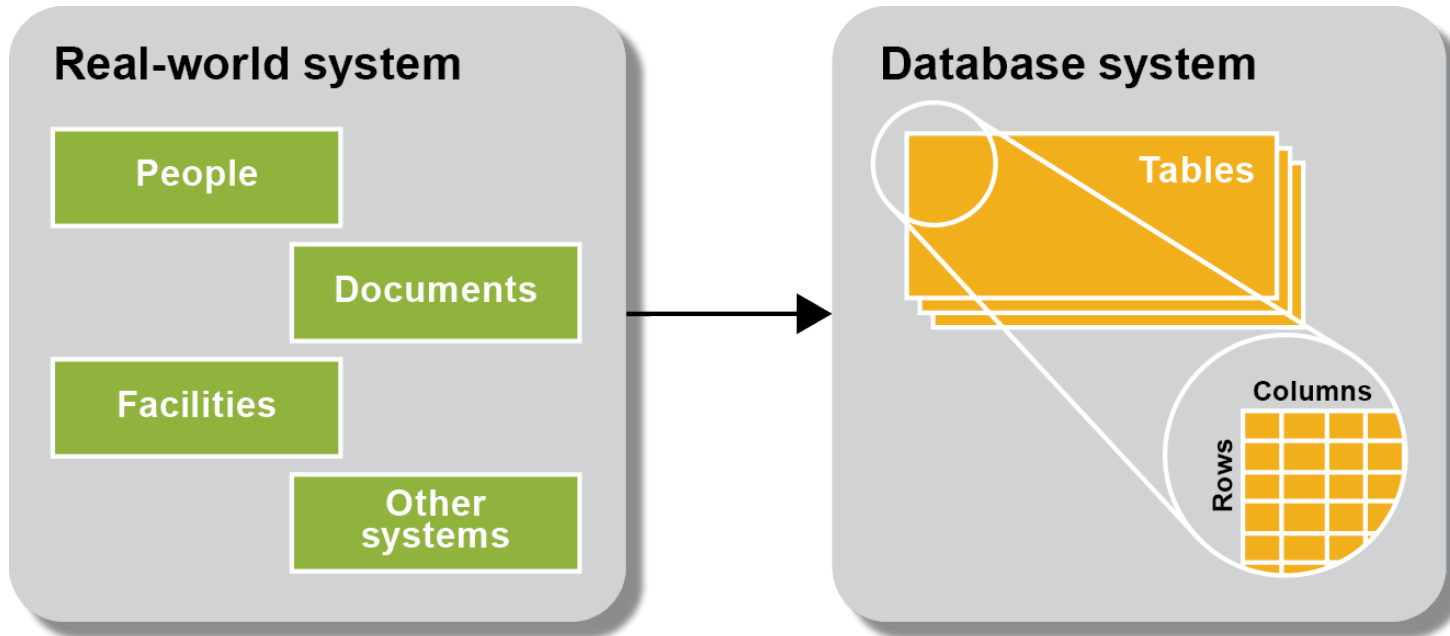
- Given the specifications for a database, identify the tables, columns, keys, relationships, and indexes for the database.
- Given the tables for an unnormalized database, normalize the structure to the third normal form.
- Use MySQL Workbench to create or work with an EER model for a database and any EER diagrams that are associated with that model.

Objectives (cont.)

Knowledge

- Give three criteria for when a column should be indexed.
- Describe referential integrity.
- Explain how MySQL uses declarative referential integrity to prevent deletion, insertion, and update problems.
- Explain how normalizing a database to the third normal form affects database performance.

A database system is modeled after a real-world system



The six basic steps for designing a data structure

Step 1: Identify the data elements

Step 2: Subdivide each element into its smallest useful components

Step 3: Identify the tables and assign columns

Step 4: Identify the primary and foreign keys

Step 5: Review whether the data structure is normalized

Step 6: Identify the indexes

Invoice used to identify data elements

Acme Fabrication, Inc.

<i>Custom Contraptions, Contrivances and Confabulations</i>	Invoice Number:	I01-1088
1234 West Industrial Way East Los Angeles California 90022	Invoice Date:	10/05/11
800.555.1212 fax 562.555.1213 www.acmefabrication.com	Terms:	Net 30

Part No.	Qty.	Description	Unit Price	Extension
CUST345	12	Design service, hr	100.00	1200.00
457332	7	Baling wire, 25x3ft roll	79.90	559.30
50173	4375	Duct tape, black, yd	1.09	4768.75
328771	2	Rubber tubing, 100ft roll	4.79	9.58
CUST281	7	Assembly, hr	75.00	525.00
CUST917	2	Testing, hr	125.00	250.00
		Sales Tax		245.20

Your salesperson:	Ruben Goldberg, ext 4512
Accounts receivable:	Inigo Jones, ext 4901

\$7,557.83

PLEASE PAY THIS AMOUNT

Thanks for your business!

The data elements on the invoice document

Vendor name

Vendor address

Vendor phone number

Vendor fax number

Vendor web address

Invoice number

Invoice date

Invoice terms

Item part number

Item quantity

Item description

Item unit price

Item extension

Vendor sales contact name

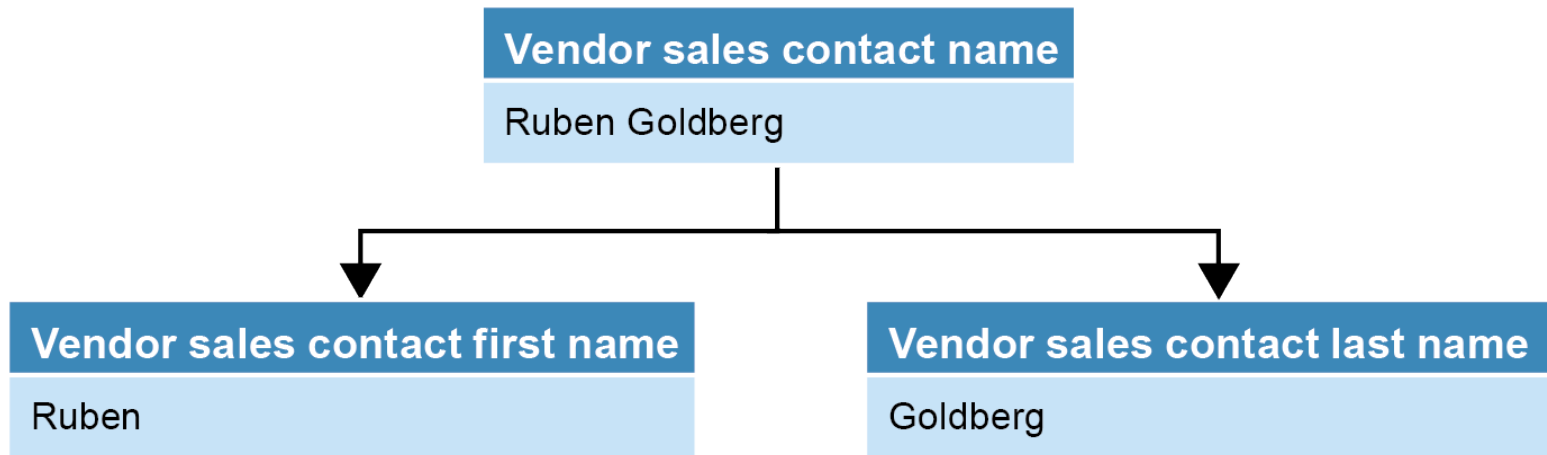
Vendor sales contact extension

Vendor AR contact name

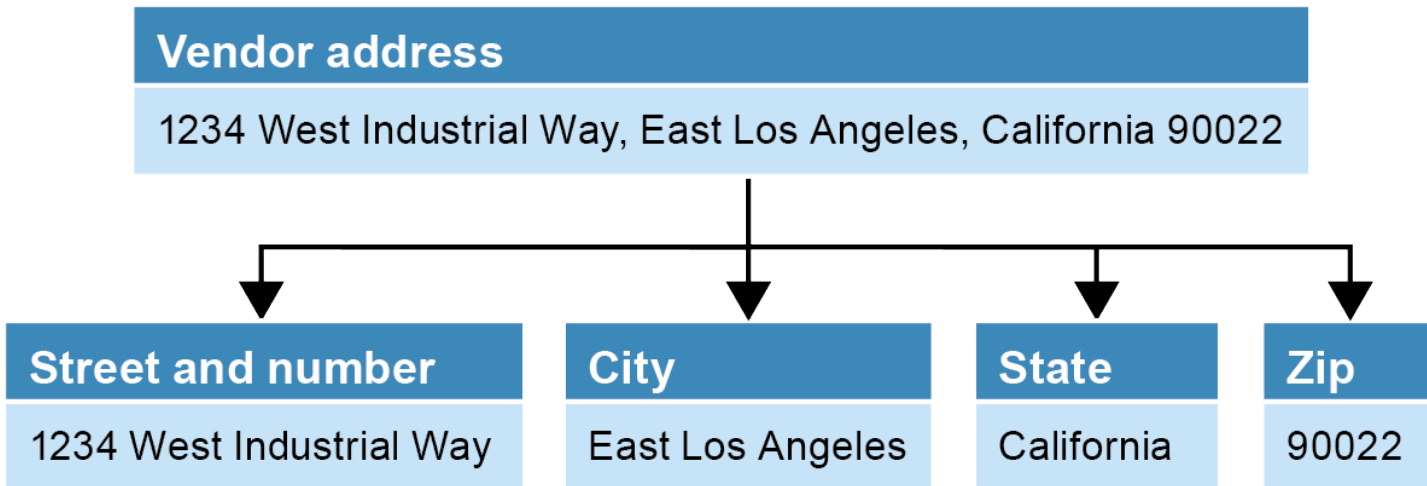
Vendor AR contact extension

Invoice total

A name that's divided into first and last names



An address that's divided into its components



Possible tables and columns for an A/P system

Vendors	Invoices
Vendor name	Invoice number*
Vendor address	Invoice date
Vendor city	Terms*
Vendor state	Invoice total
Vendor zip code	<i>Payment date</i>
Vendor phone number	<i>Payment total</i>
Vendor fax number	<i>Invoice due date</i>
Vendor web address	<i>Credit total</i>
Vendor contact first name	<i>Account number*</i>
Vendor contact last name	
Vendor contact phone	
Vendor AR first name	

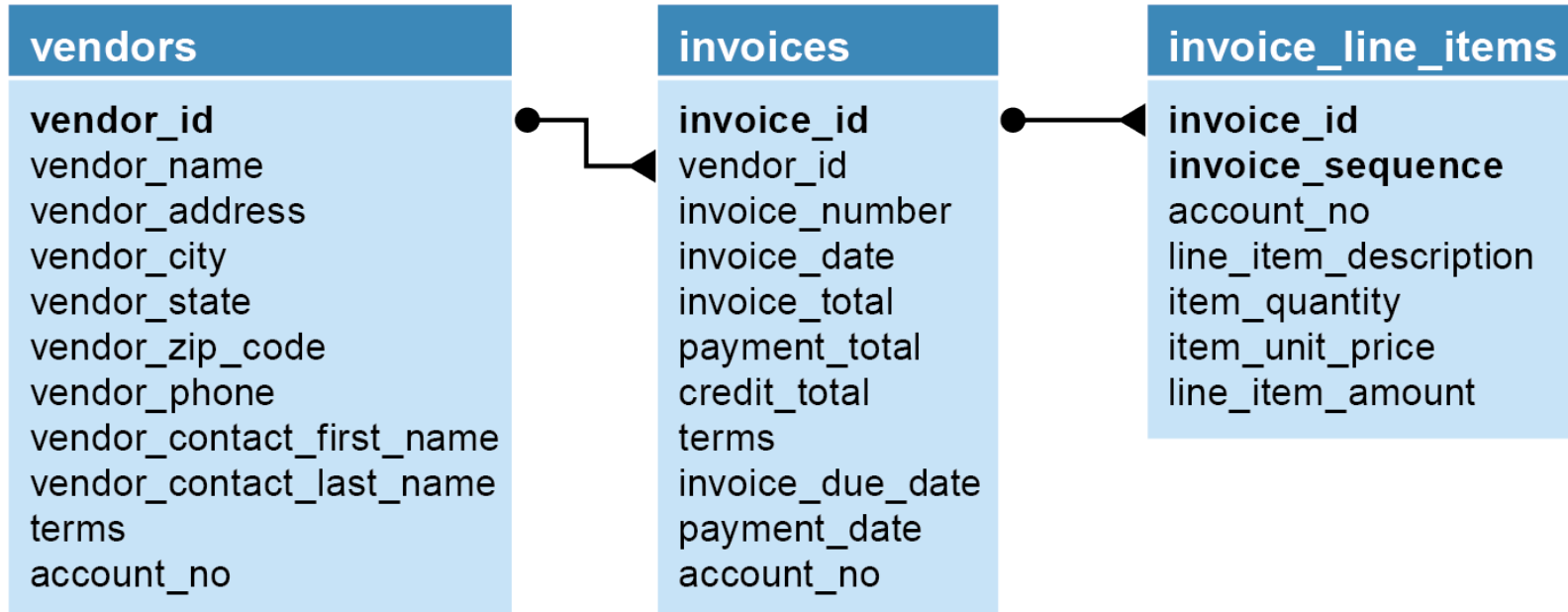
Possible tables and columns for an A/P system (continued)

Vendors	Invoice line items
Vendor AR last name	Invoice number*
Vendor AR phone	Item part number
<i>Terms*</i>	Item quantity
<i>Account number*</i>	Item description
	Item unit price
	Item extension
	<i>Account number*</i>
	<i>Sequence number</i>

The notation for identifying data elements

- Data elements that were previously identified but aren't needed are crossed out.
- Data elements that were added are displayed in italics.
- Data elements that are related to two or more entities are followed by an asterisk.

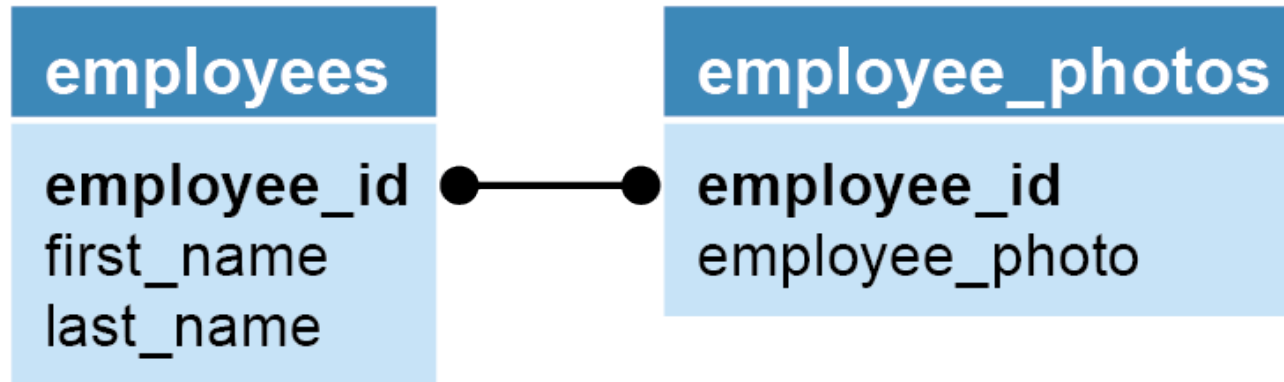
The relationships between the tables



Two tables with a many-to-many relationship



Two tables with a one-to-one relationship



Operations that can violate referential integrity

Deleting a row from the primary key table

If the foreign key table contains one or more rows related to the deleted row

Inserting a row in the foreign key table

If the foreign key value doesn't have a matching primary key value in the related table

Updating the value of a foreign key

If the new foreign key value doesn't have a matching primary key value in the related table

Updating the value of a primary key

If the foreign key table contains one or more rows related to the row that's changed

Terms to know

- Entity
- Attribute
- Instance
- Entity-relationship (ER) modeling
- Referential integrity
- Declarative referential integrity (DRI)
- Foreign key constraint
- Orphaned row

Two tables that need to be normalized

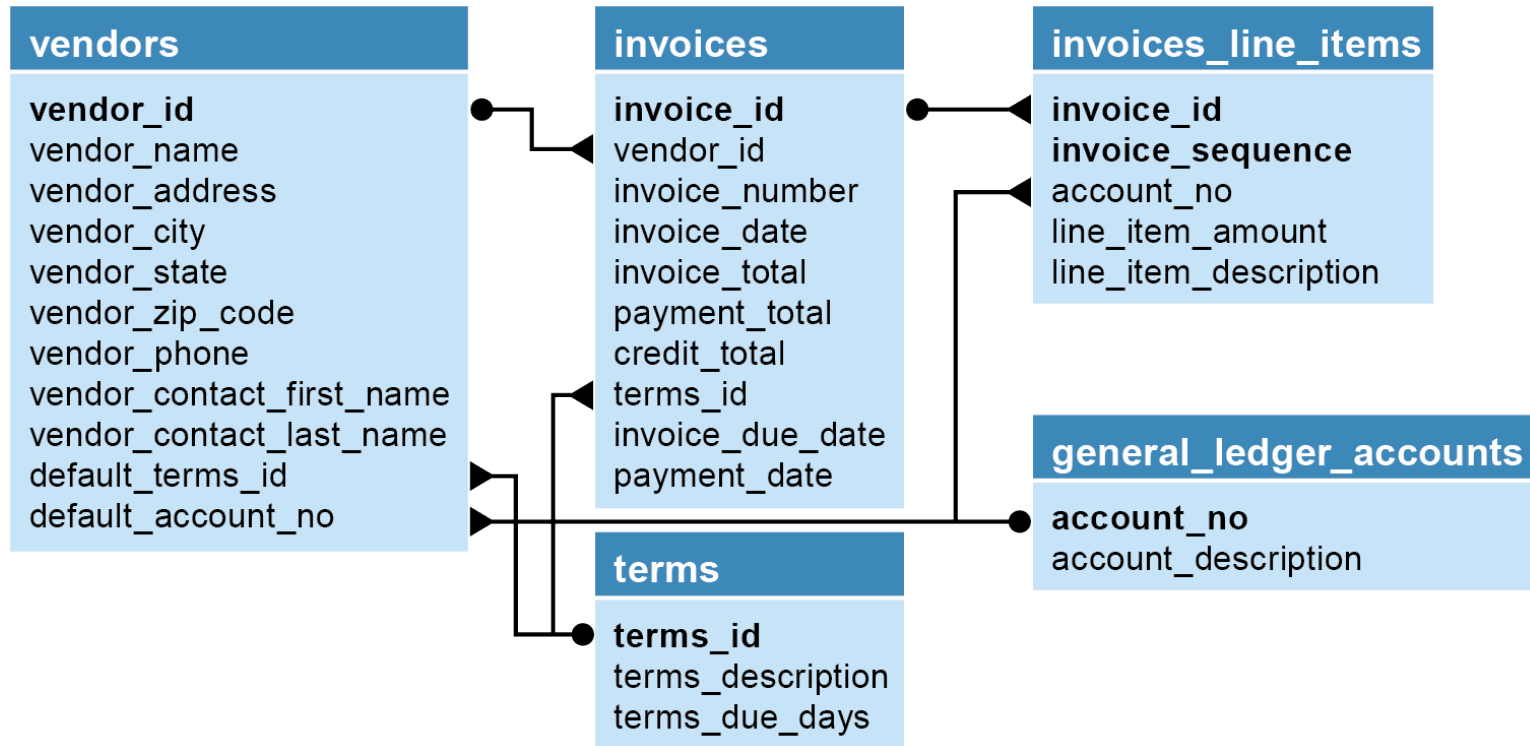
A table that contains repeating columns

	vendor_name	invoice_number	item_description_1	item_description_2	item_description_3
►	Cahners Publishing	112897	VB ad	SQL ad	Library directory
	Zylka Design	97/552	Catalogs	SQL flyer	NULL
	Zylka Design	97/553B	Card revision	NULL	NULL

A table that contains redundant data

	vendor_name	invoice_number	item_description
►	Cahners Publishing	112897	VB ad
	Cahners Publishing	112897	SQL ad
	Cahners Publishing	112897	Library directory
	Zylka Design	97/522	Catalogs
	Zylka Design	97/522	SQL flyer
	Zylka Design	97/533B	Card revision

The accounts payable system in third normal form



Terms to know

- Normalization
- Data redundancy
- Unnormalized data structure
- Normalized data structure
- Normal form

When to create an index

- When the column is used frequently in search conditions or joins
- When the column contains a large number of distinct values
- When the column is updated infrequently

The first three normal forms

Normal form	Description
First (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 (2NF)	Every non-key column must depend on the entire primary key.
Third (3NF)	Every non-key column must depend only on the primary key.

Note

- Most designers stop at the third normal form.

The next four normal forms

Normal form	Description
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 <i>multivalued dependency</i> , 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 <i>domain</i> is the set of allowable values for a column.

The benefits of normalization

- More tables, and each table has an index on its primary key. That makes data retrieval more efficient.
- Each table contains information about a single entity. That makes data retrieval and insert, update, and delete operations more efficient.
- Each table has fewer indexes, which makes insert, update, and delete operations more efficient.
- Data redundancy is minimized, which simplifies maintenance and reduces storage.

Unnormalized invoice data

The invoice data with a column that contains repeating values

	vendor_name	invoice_number	item_description
▶	Cahners Publishing	112897	VB ad, SQL ad, Library directory
	Zylka Design	97/522	Catalogs, SQL Flyer
	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
▶	Cahners Publishing	112897	VB ad	SQL ad	Library directory
	Zylka Design	97/552	Catalogs	SQL flyer	NULL
	Zylka Design	97/553B	Card revision	NULL	NULL

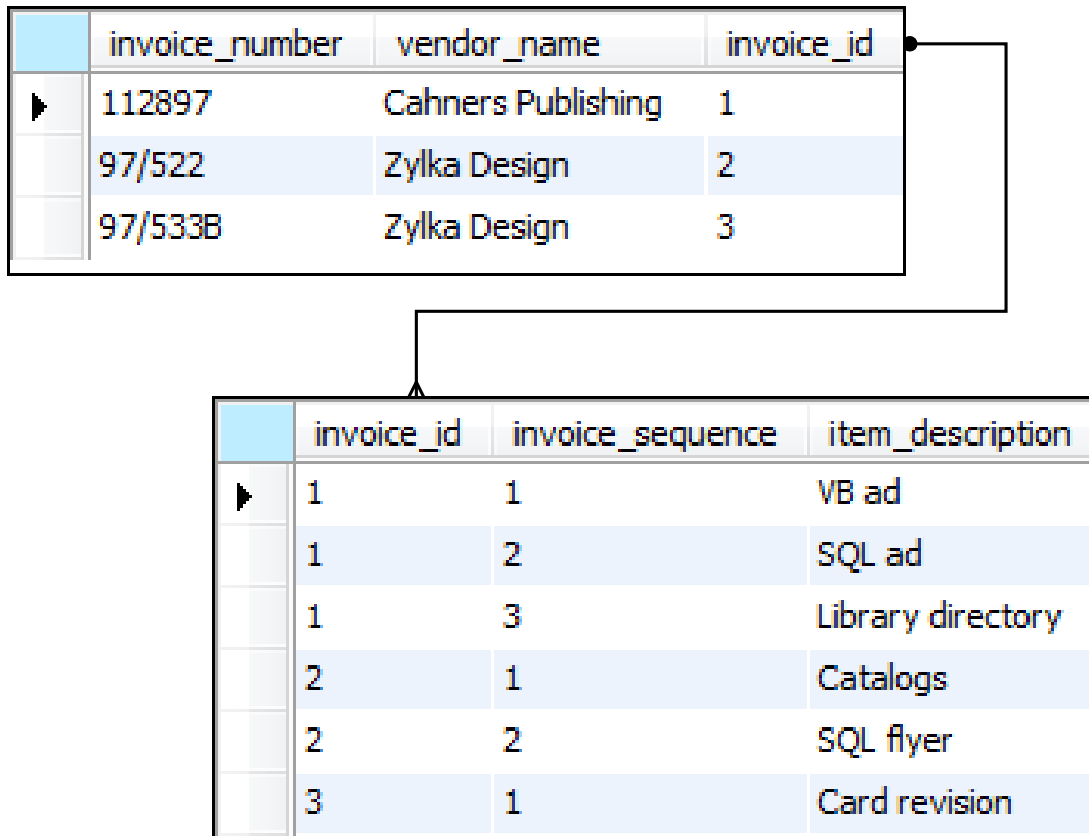
The invoice data in first normal form

	vendor_name	invoice_number	item_description
►	Cahners Publishing	112897	VB ad
	Cahners Publishing	112897	SQL ad
	Cahners Publishing	112897	Library directory
	Zylka Design	97/522	Catalogs
	Zylka Design	97/522	SQL flyer
	Zylka Design	97/533B	Card revision

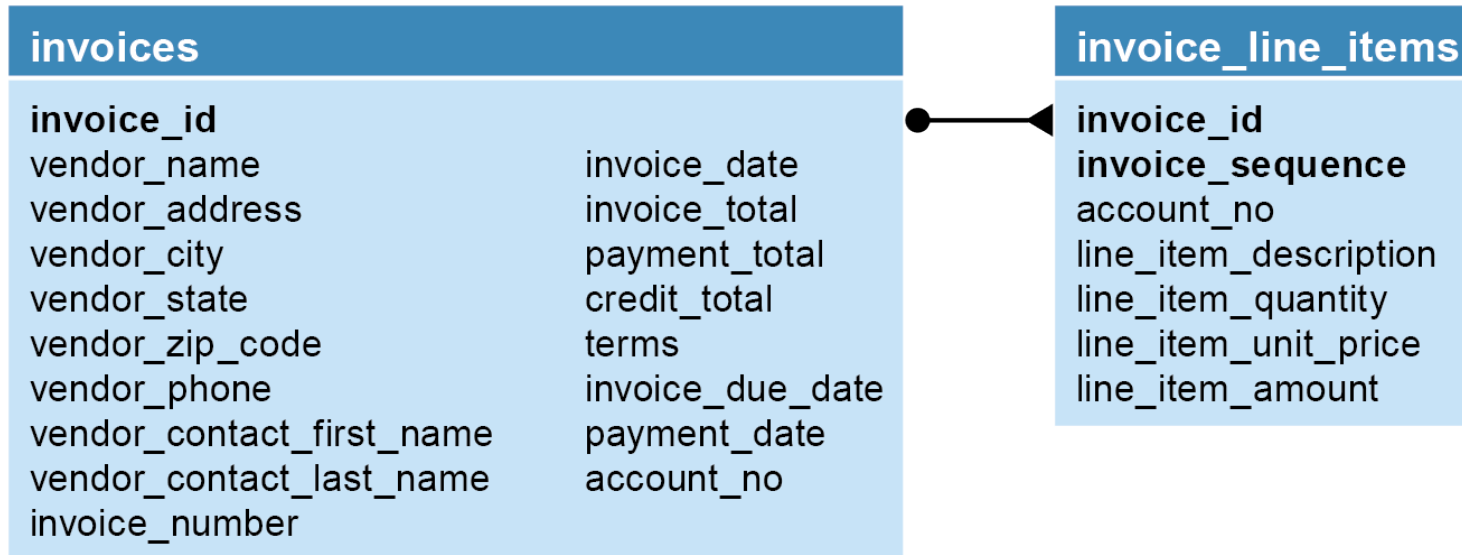
The invoice data in first normal form with keys added

	invoice_id	vendor_name	invoice_number	invoice_sequence	item_description
►	1	Cahners Publishing	112897	1	VB ad
	1	Cahners Publishing	112897	2	SQL ad
	1	Cahners Publishing	112897	3	Library directory
	2	Zylka Design	97/522	1	Catalogs
	2	Zylka Design	97/522	2	SQL flyer
	3	Zylka Design	97/533B	1	Card revision

The invoice data in second normal form



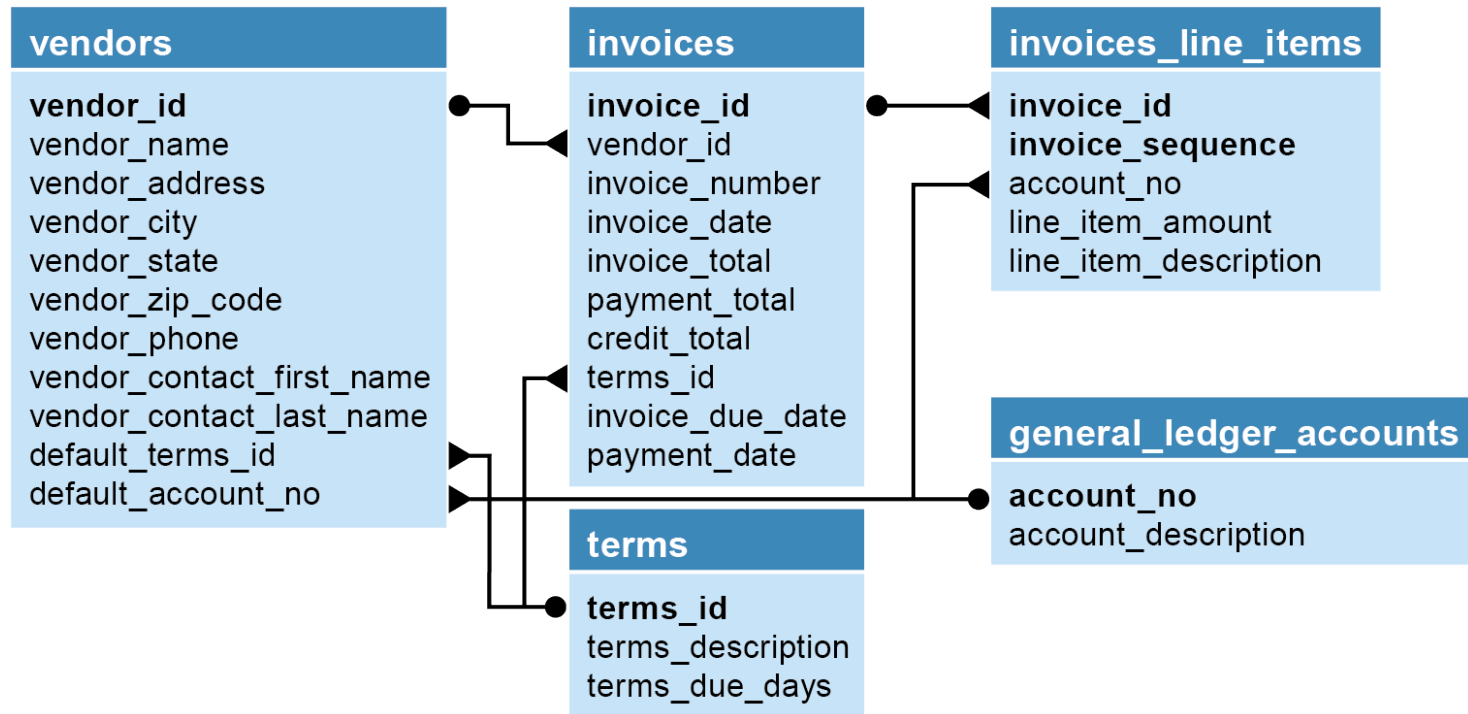
The A/P system in second normal form



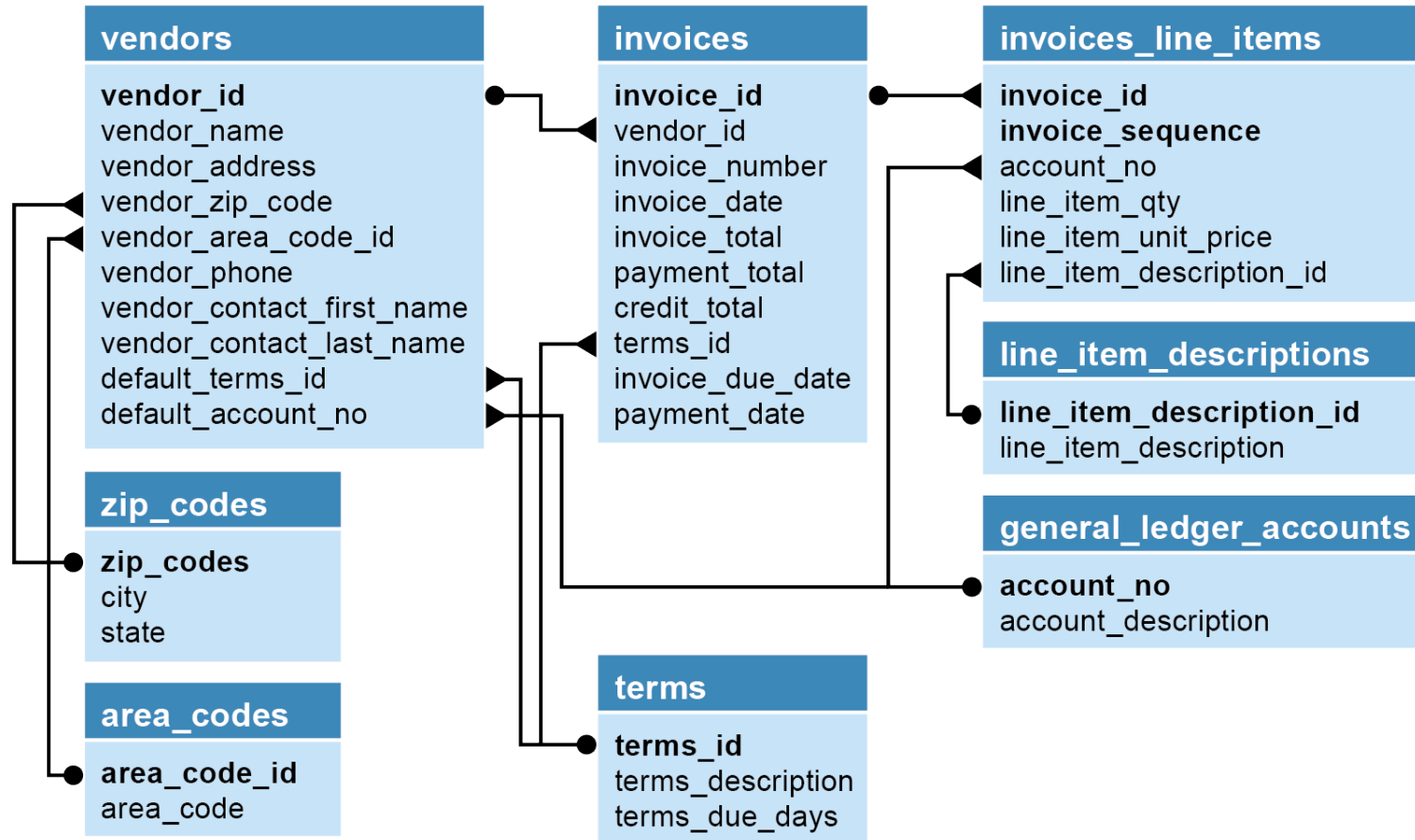
Questions about the structure

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?

The accounts payable system in third normal form



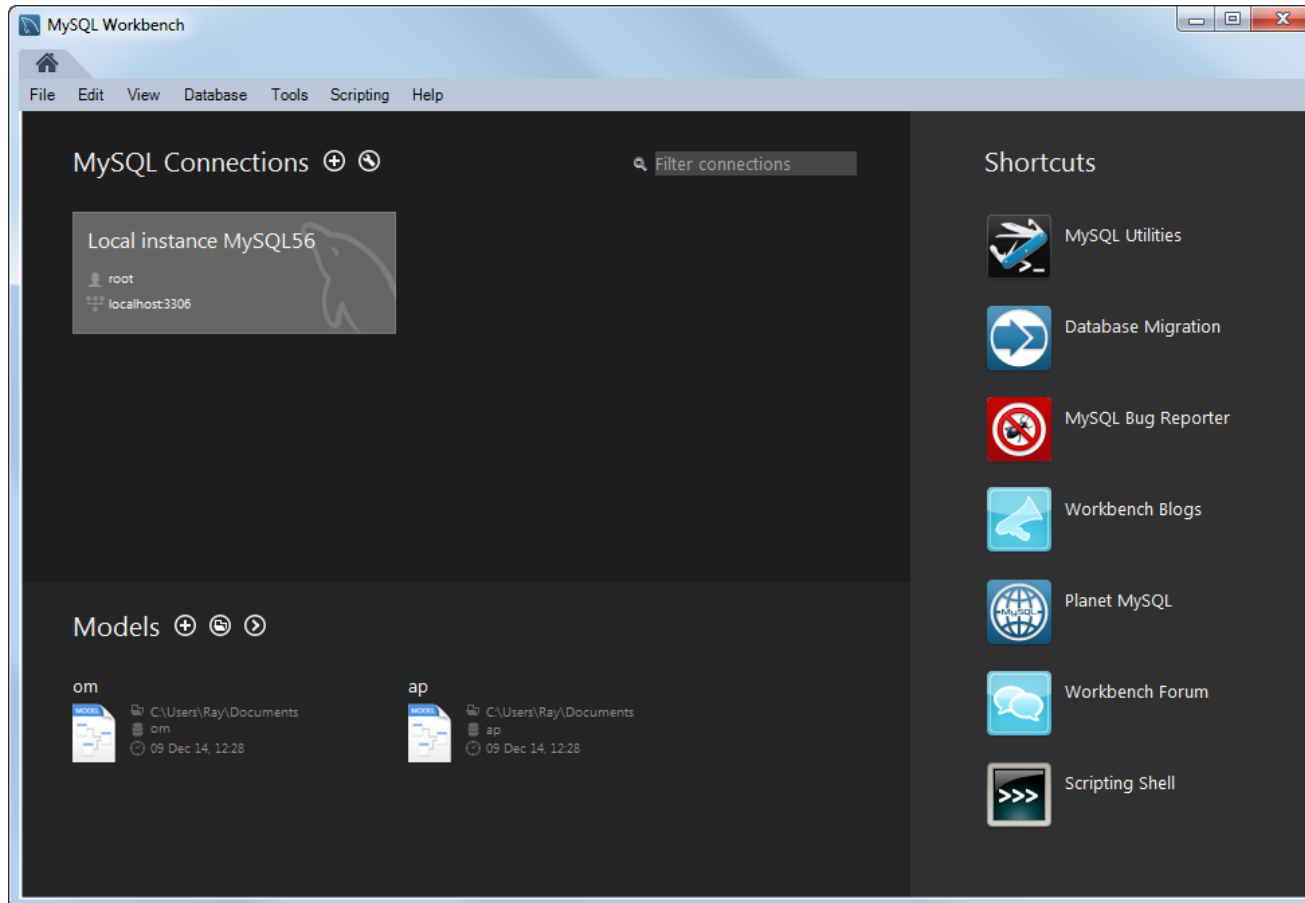
The accounts payable system in fifth normal form



When to denormalize

- When a column from a joined table is used repeatedly in search criteria.
- If a table is updated infrequently.
- Include columns with derived values when those values are used frequently in search conditions.

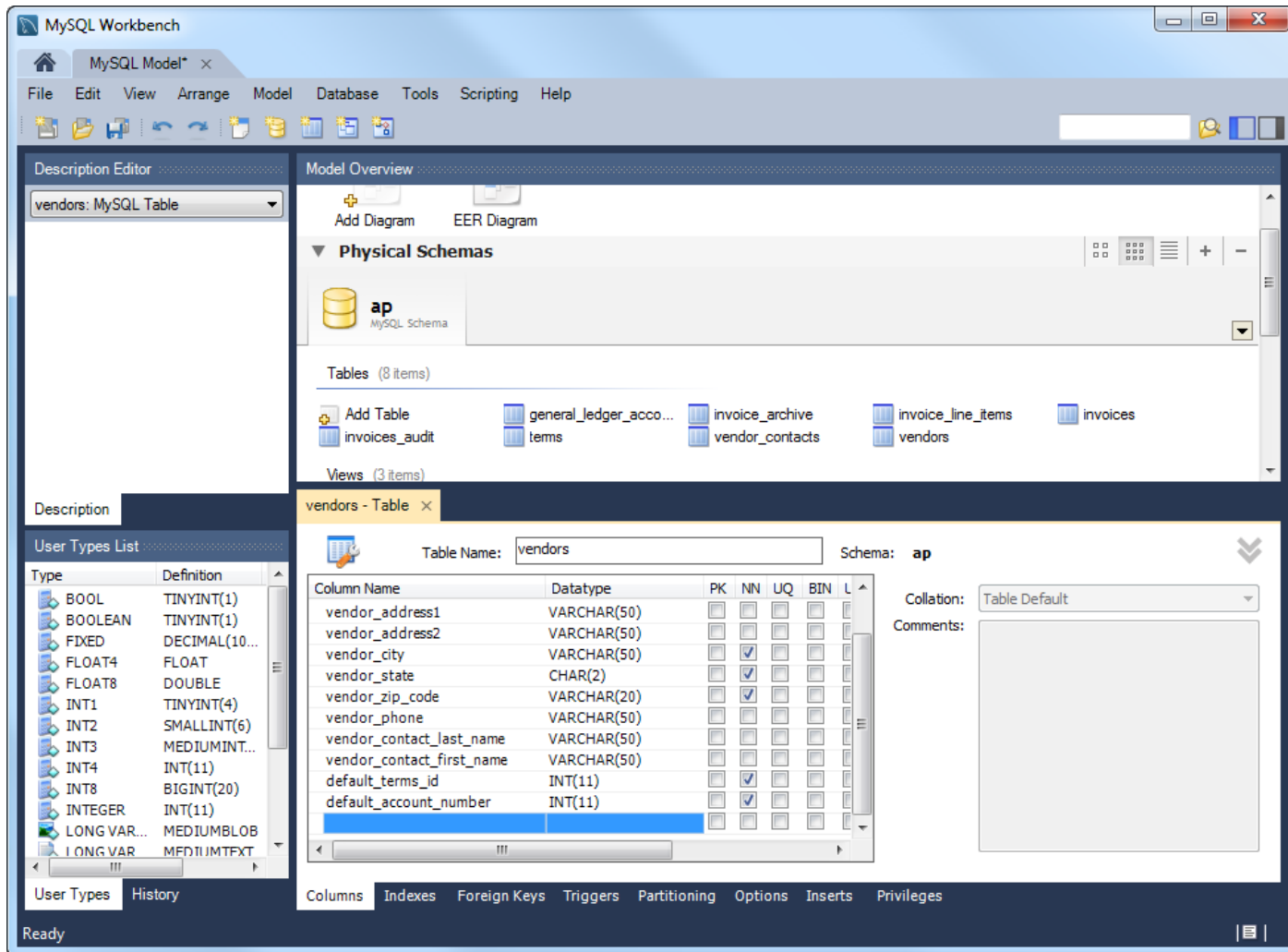
The home page of MySQL Workbench



Operations you can perform from the Home tab

- Open an existing EER model
- Create a blank EER model
- Create an EER model from an existing database
- Create an EER model from a SQL creation script
- Remove an EER model from the list

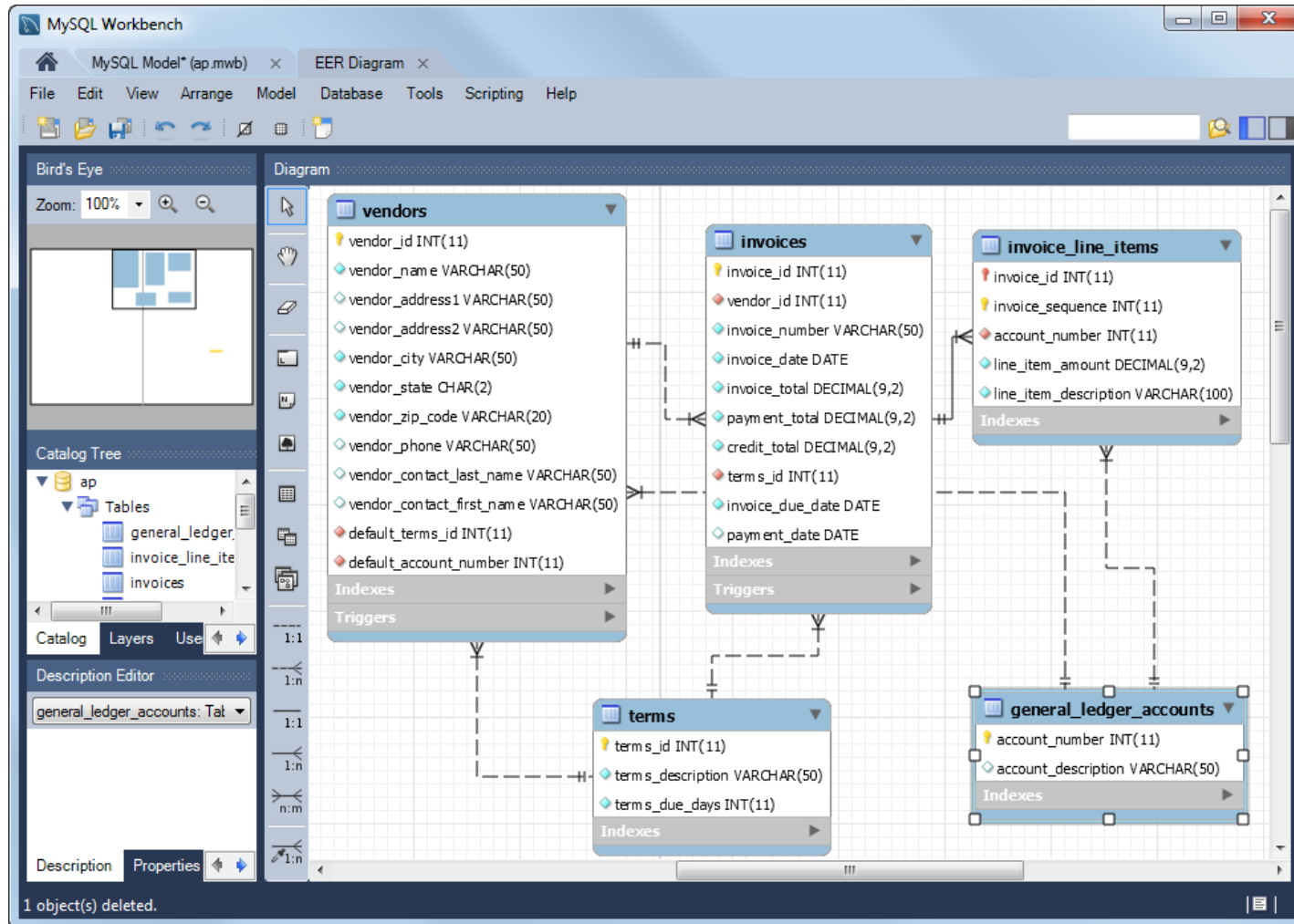
The EER model for the AP database



Operations for working with an EER model

- Edit a table
- Add a new table
- Delete a table
- Display a diagram
- Create a new diagram
- Create a database creation script from the model

The EER diagram for the AP database



Operations for working with an EER Diagram

- Add an existing table to the diagram
- Add a new table to the diagram
- Display the model for a table
- Define the relationship between two tables
- Edit and delete relationships
- Remove a table from the diagram (and, optionally, the model)