

IS 475/675 Database Design Project

What is the purpose of this part of the project?

The objective of this part of the project is to design a database to support the requested operations of the company described in this document. This project requires you to:

- Identify the data required to be stored for the application.
- Identify the entities, attributes, primary keys, relationships and foreign keys.
- Create a logical data model in third normal form.
- Write a memo briefly explaining the logical data model.

What are the deliverables from this part of the project?

There are two deliverables for this part of the project: A memo to me and a data model for the application in third normal form.

The data model should be documented using a logical entity-relationship diagram (ERD). The ERD must be produced using Visio or another computer modeling tool. Please **do not** turn in an ERD drawn by hand.

Describe the data model with a business-style memo. This memo will be written to me, as your instructor so that I can understand any assumptions you made to complete the model. The memo should:

1. **Describe the data model.** When describing your data model, you can assume I know how to read an ERD. You can't assume that I understand your abbreviations or what was in your mind when you designed the database. Explain how your database design will produce the information desired by the client. Review the managerial and operational questions that will depend on the database and explain how your data model will store the data necessary to eventually answer those questions.
2. **Explain the assumptions.** If you need to assume anything (existence of data in another system that you will link to, business rules, etc.) please explain those assumptions.
3. **Describe any limitations of the data model.** All data models have limitations. Explain any data needed by this organization that you have elected not to store. Make sure you include your reasoning for **not** storing any requested data, if you make the decision to not store requested data. Explain any problems that you anticipate might occur in storing data for a long time in your database.

Use a business memo format and incorporate easy-to-understand headers and bullet points. I really like headings/sub-headings and bullet points. The memo should be brief – about 2 pages in length.

What is the application?

Replica Toys designs, develops, manufactures, and markets sophisticated, expensive, automated toys for children. Replica Toys sells directly to the public, but most of their toy sales are made through distributors such as Toys-R-Us, World of Toys and independent toy stores. The company's toys are relatively expensive (the least expensive toy sold is priced at \$250 and the most expensive toy sells for \$2,875) and are also frequently sold through exclusive, specialized, independent toy stores. Replica Toys sells few toys when compared to other toy manufacturers because their market includes only those people who are looking for sophisticated, expensive, very high quality toys.

The company's first product line included a set of miniature ride-able motor vehicles. Each toy is a complex replica of an adult vehicle. Their ride-able toys include all-terrain vehicles, trains, trucks, luxury sedans, sports cars, and vintage automobiles. The product line includes approximately 35 unique ride-able vehicles at any one time. The model number indicates the general model of a toy and could be considered a product identifier. For example, a "Chevy 4 door SUV" is model CSUV7. Each individual toy, however, is uniquely identified by a serial number. Replica Toys will sell more than one model CSUV7, each with its own unique serial number.

Replica Toys wants to create a database **to track post-sales information**. The database will be used by two different departments in the organization:

- 1) **Marketing Department.** The Marketing Department will use the database to answer questions about customer purchases. The Marketing Department wants to store customer registration information. They will use registration data to better understand why customers elect to purchase toys from the company. They will use the return data to understand why customers return toys from the company. They may conduct customer surveys in the future, and want a way to store data from those surveys. The Marketing Department plans to gather reviews on their products from public sources and store those reviews in the database.
- 2) **Quality Control.** The Quality Control group within the Manufacturing Department will use the database to track problems with toys that have been sold. The Quality Control group wants to store problem reports, testing and resolution documents.

The company wants to consolidate all post-sales returns, registration, service and customer data in a single database. This database does not contain customer order data, but in the future it might link to the customer order database. This database does not contain inventory data, but in the future it might link to the inventory database. The purpose of this particular database, however, is to store post-sales data. **It is NOT to store order data or inventory data** beyond what is necessary to support the processes required by the Marketing Department and the Quality Control group.

Registration Data

In order to keep track of customers and purchases, Replica Toys asks all customers to fill out the registration form shown on the next page as Figure 1. Customers are usually compliant with that request because Replica Toys offers free service for two years to all those who submit the form. Replica Toys also provides savings offers to current customers for future purchases, if those customers fill out the registration form. Replica Toys has many return customers – their customers tend to buy toys from them many different times.

Registration forms will be filled out online, using the database you are designing as the place to store the data collected from customers. Registration forms are completed by the person who purchased the toy. Most toys are sold through distributors. The company anticipates using the serial number of a toy as the

primary key because each toy can be uniquely identified by its primary key. Business rules about the registration data are provided on the next page.

Buyer Information		Purchase Information	
First Name	Bert	Model Number	JCSUV7
Last Name	Swanson	Serial Number	19029388-11
Address	7883 Cordova Ave.	Date of Purchase	03/04/2016
City	Rancho Cucamonga	Distributor	Claremont Toy World
State	CA	Price	485.99
Zip	91209	Primary User of Toy	
Email	bswanson@gmail.com	Age	3
Primary Phone	714-232-1111	Gender	F
Background Information about the Purchase			
Where did you first learn about toys from Replica Toys? (Check only one)		<input type="checkbox"/> Advertisement in print <input type="checkbox"/> Advertisement on the web <input checked="" type="checkbox"/> Advertisement on TV <input type="checkbox"/> Friend's recommendation <input type="checkbox"/> In-store display <input type="checkbox"/> Catalog <input type="checkbox"/> Other	
What features were you most interested in when buying the toy? (Check all that apply)		<input checked="" type="checkbox"/> Type of toy (car, jeep, etc.) <input type="checkbox"/> Size <input checked="" type="checkbox"/> Color <input checked="" type="checkbox"/> Speed <input checked="" type="checkbox"/> Quality of design <input type="checkbox"/> Level of replication from original <input type="checkbox"/> Safety features <input type="checkbox"/> Cost <input type="checkbox"/> Sound features <input type="checkbox"/> Other	
How are you related to the primary user of the toy? (Check only one)		<input type="checkbox"/> Parent <input type="checkbox"/> Grandparent <input checked="" type="checkbox"/> Aunt/Uncle <input type="checkbox"/> Friend <input type="checkbox"/> Other Relative <input type="checkbox"/> Other	
Do you anticipate buying similar toys in the future? (Check only one)		<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> Don't know	

Figure 1: Sample Registration Form

Business Rules:

- A registration will contain only one answer to the question “Where did you first learn about Replica Toys?”
- A registration may contain multiple answers to the question “What features were you most interested in when buying the toy?” The Marketing Department anticipates that additional features may be added as they determine which features are most important to customers.

- A registration will contain only one answer to the question “How are you related to the primary user of the toy?”
- Each toy has a unique serial number. A serial number has one and only one registration – Replica Toys does not keep track of re-sale registrations for its toys. A serial number determines the model number of a toy. The model number determines the name and description of a given toy. The company keeps track of the name, description, and recommended price of each model (even though that information is not shown on the registration form in Figure 1), so that data should be in this database.
- A given registration can have only one gender checked.
- The company does not keep track of the name of the intended primary user, since that user is most likely a child. The age of the intended primary user is interesting to the Marketing Department because they want to know the ages at purchase of the recipients of each particular model. The “buyer” is the person who purchased the toy. The buyer (customer) is usually an indulgent relative (parent, grandparent, aunt, uncle) of the intended user of the toy.

Return Data

Replica Toys has not kept track of return data in the past, but the Marketing Department wants to understand why people are returning toys. The Marketing Department wants to store data about returns made from toy buyers. They requested that their distributors ask their customers (people buying the toys) to provide an explanation when making a return. This data will be gathered from the cash register systems of the distributors and sent to the Marketing Department at Replica Toys. For those people buying toys directly online from Replica Toys, the data will be gathered from an online form. Return data that the Marketing Department wants to store will include: Serial Number, Date purchased, date returned, reason for return (they plan to have about 10 standard reasons why people would return the product), and a notes area where people can type in any comments they have about the returned item. Personnel in the Marketing Department anticipate they will be able to persuade their distributors to type in any comments that the customers made about the toy when returning the product. Since this system is not in place yet, they don’t have a sample form to provide for you to understand the data required.

Survey Data

The Marketing Department anticipates using registration data to occasionally survey past customers. The folks in the Marketing Department aren’t sure how they will conduct the surveys yet, but they know they will need a place to store the data once they do conduct surveys. Surveys would not be related to a specific registration, they would be related to a customer. The folks in the Marketing Department want to be able to store the questions for a survey, possible answers to questions for a survey, and a specific answer given to a specific question by a specific customer. They really aren't sure exactly what the surveys would look like right now, but here are a few samples of questions and answers that they hope to be able to store for a customer:

Question	Answer
1. How many rideable toys have you purchased in total?	5
2. How many of those rideable toys have you purchased from Replica Toys?	2
3. How would you rate the quality of the toys you purchased from Replica Toys in comparison to those purchased from other companies?	Much better
4. How likely are you to purchase another toy from Replica Toys?	Somewhat likely

Review Data

The Marketing Department anticipates gathering data from available sources to capture customer reviews of their products. They plan to relate a review to a given model of a toy. The data that they will store for each review is: Model number, review date, review source (i.e. Yelp, Google), and review text.

Quality Control Data

One of the greatest concerns about the products sold by Replica Toys is that they be safe for children. The products are fairly complicated and it is possible that a toy could harm a child if the design is flawed or if the implementation of the design is not correct. As a result, the company monitors the problems reported with toys very closely. If there are any problems with a toy, Replica Toys notifies customers and distributors directly. “Distributors” are what Replica Toys calls the “place of purchase” as indicated on the form in Figure 1.

To keep track of problems reported for a product, employees at the company fill out the problem report form shown below as Figure 2. Whenever a customer or distributor calls in with a problem, an employee at Replica Toys fills out the form in Figure 2 with detailed information about the problem. There is a standard set of problems (problem types) used to fill out the problem description area of the form in Figure 2, but each problem reported is currently typed in as text by the employee filling out the form. Problems are related to a specific serial number, but a specific serial number may have more than one problem reported.

Date/Time of Report	04/02/2016 9:35AM	Report ID	760-22931
Model Number	JCSUV7	Serial Number	19029388-11
Returned?	Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>	Return Date	
Person Reporting Problem (Check only one)	<input checked="" type="checkbox"/> Customer <input type="checkbox"/> Employee of Replica Toys <input type="checkbox"/> Distributor <input type="checkbox"/> User <input type="checkbox"/> Other		
Complaint Made	<input type="checkbox"/> WebForm <input type="checkbox"/> Phone <input checked="" type="checkbox"/> Email <input type="checkbox"/> Twitter <input type="checkbox"/> Store		
First Name	Bert		
Last Name	Swanson		
Address	7883 Cordova Ave.		
City	Rancho Cucamonga		
State	CA		
Zip	91209		
Phone Number	714-232-1111		
Email	bswanson@gmail.com		
Type of Problem (Check only one)	<input type="checkbox"/> Inadequate finish – toy does not look good <input type="checkbox"/> Inadequate operation – toy does not operate correctly <input checked="" type="checkbox"/> Poor operation – toy tips over during use <input type="checkbox"/> Inadequate operation quality – toy broke during use <input type="checkbox"/> Poor operation quality – toy hurt user <input type="checkbox"/> Other		
Description of Problem	When driving around corners, the Chevy Tahoe seems top heavy and tips over unexpectedly. Happens most often on uneven terrain.		
Injury	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No		
Description of Injury			

Figure 2. Sample Problem Report Form

Problem reports can be initiated by a distributor, an employee of Replica Toys, or a customer. A problem report is initiated by only one person. The person who initiates a problem report could fill out multiple problem reports. Multiple problem reports could be filled out for a given toy (uniquely identified by a serial number).

It is not necessary to have a registration form stored to create a problem report. For example, a customer does not have to register a toy to call the company with a problem with a specific toy. While this post-sales system includes both Marketing and Quality Control data, they are only tangentially related to each other.

Once a problem report is filed, Replica Toys does testing and evaluates the validity of the complaint. Employees fill out reports for each test performed in relation to a problem report. An example of the Test form is provided in Figure 3.

Problem Report ID	760-22931	Date and Time of Report	04/08/2016 8:30AM
Model Number	JCSUV7	Serial Number	N/A
EmployeeID	35002	Test Type	Operational
Test Description	Used computer simulation of vehicle with 40 pound rider 3 MPH; 5MPH; 10MPH; 15MPH test runs. Tried it on qualified uneven terrain including potholes. Tried it on unqualified terrain (sand).		
Test Results	Slight outside tilt; no tipping at any speed on qualified terrain. Tipping occurs when vehicle is used on sand.		
Recommended Resolution	No problems with expected performance. Provide additional documentation to Marketing about intended operating environment for JCSUV7. Recommend inclusion of additional information in operating manual stating: don't operate on sand.		
Test complete?	Yes.		

Figure 3. Test Form

More than one test form may be completed per problem report. A test form is not filled out unless initiated by a problem report; there must be a problem report stored in the database before a test form can be completed. It is possible that the recommended resolution of a test form is to perform another test. When that occurs and another test must be performed, the “test complete” question on the form is answered as “no”.

Managerial and Operational Questions

Two disparate groups (Marketing and Quality Control) each have different types of questions they hope to answer from information obtained from this database. Both groups have data from other databases in the organization – Marketing gets most of its data from the ordering database, while Quality Control uses the manufacturing database. But neither of those databases stores post-sales registration, return, survey, and service data. Here are samples of the questions that the two groups hope to answer from this database:

- How old is the average user for each of the different styles of vehicles?
- What is the preferred purchase by gender?
- What are the most important features to a grandparent purchasing a vehicle for a grandchild?
- How do grandparents hear about our toys?
- Which toy models are returned the most often?
- What reason is given for returns for those toy models returned the most often?
- Which user age group has the most returns?
- What is the relationship of the customer to the intended user for those toys returned the most often?
- What is the relationship to the primary user for our different models? In other words, do grandparents or parents buy the most SUVs? Which type of buyer buys the most sports cars?
- Which model has the most problem reports? Which model has the least problem reports?
- Who reports the most problems – distributors or customers?
- What is the gender and age group of the primary user with the most and least problem reports?
- What is the current status of a specific problem report?
- What tests did we run on a specific date?
- Are there any problem reports that have not had any tests run?
- Which problem reports are not complete (have not been issued a “yes” on the “test complete” question of a test form?)
- What is the count of injury reports by model?
- What is the count of injury reports by model by age of primary user?
- What is the count of injury reports by model by location of buyer? Or by location of distributor?
- When did we receive the most current injury report for model CVR722? (or any other specific model)

IS 475/675 Database Design and Implementation Project – Parts 2 & 3

What is the purpose of these two parts of the project?

The objectives of these parts of the project are to revise your logical model and validate that model through the implementation of a prototype database for Replica Toys. These parts of the project require you to:

- Evaluate and **revise** as necessary the logical data model submitted for part 1. An important final deliverable for this project is a complete logical data model in third normal form.
- Design a prototype physical data model that will generate the result tables listed in this document.
- Create a prototype physical data model using the SQL Server database management system. Create only those tables necessary to answer the required questions. **Do not create every table on your ERD – it will be too time-consuming!!** Feel free to include only those attributes needed for a particular query. In other words, you do not have to include each of the attributes in a table that are on your “final” data model. A prototype is for model validation, not for final implementation.

Note: Since this is a prototype that must be completed in a short time, you can take short cuts. For example, when you create your tables, include ONLY those attributes in your tables that are necessary to satisfy the queries. I’d prefer that you keep your physical data model in third normal form. **If you have questions, please consult me about what would or would not be an allowable “short cut” for this prototype.**

- Devise a robust test data set to test your data model. A “robust” test data set includes data in various stages of completeness that will allow you to validate both your database design and your queries. For example, you will need problem reports that are in various levels of completion. You will need a problem report that has just been entered (no testing completed), a problem report with one test completed, a problem report with a test completed that is closed, and a problem report that has multiple tests. Populate your tables with your test data set. Again, do not try and create and populate every single table and attribute on the data model – it will take too much time. Instead, focus on creating a really good test data set for those tables you actually need to answer the queries required.
- Write SQL queries to generate the requested result tables. Make sure that each query you write actually returns data in the result table. You may have to adjust your test data set to ensure that you have data to return for each query. A test data set that returns “no rows” as the answer to a query is NOT acceptable!!

What are the deliverables for Part 2?

There are five deliverables for this part of the project:

1. **A memo.** Write a memo to me explaining what you have modified in your logical data model that addresses the issues raised in my response to part 1 of the project. I recommend that you take my list of issues/concerns from part 1 and explain how you fixed or chose not to fix each one. This is a memo to me in my role as your instructor, rather than a memo to me in my role as the “client”.
2. **A revised logical data model.** This deliverable is an ERD. If your logical data model (documented with an ERD) turned in for part 1 did not have any problems noted by me, then simply turn in that ERD. If you found you needed to modify the design to address my concerns or to answer the required questions for this part of the project, then turn in a revised

ERD. You must have a logical data model in third normal form as one of the deliverables for this part of the project.

3. **A copy of your original logical data model with my comments.** Turn in your original ERD (submitted for part 1) with my comments so I can compare the original with the revised model.
4. **A physical data model for the prototype database.** This deliverable will be an ERD of the database implemented for the prototype. Since the prototype should include only those tables necessary to answer the required queries, this ERD will be a sub-set of the revised logical data model. Include **only the tables and attributes** that are actually in your prototype on this ERD. For any tables with concatenated primary keys, you should create single attribute surrogate primary keys for your physical data model. There should be no concatenated primary keys on your physical data model or in your actual prototype database.
5. **A prototype database.** The prototype will include all tables required to answer the questions in part 3. Please note that you do not have to install your entire database as your prototype!!! The prototype does not have to be complete – it just has to work for the required queries and maintain the integrity of your logical design. It is not acceptable to create a prototype that is not in third normal form without prior approval from the instructor, but you do not have to include all the tables in your logical ERD.

You must have enough rows in each table to reasonably test your queries. I recommend looking at the queries you are required to write before creating the test data so that the data you create will effectively test the queries. Do NOT simply repeat data from row to row – create a sample data set that will provide a good test of your SQL code. I also recommend creating “meaningful” data rather than nonsensical data to better test queries. In other words, create real names, registrations, and toys where applicable. The number of rows in the transaction table (registrations) should be larger than the number of rows in the parent tables. **Look at the queries before creating the tables.**

One of the factors used to evaluate your project is the robustness of your test data set. You must create a test data set that will fully test the parameters of your queries.

Make sure you look at the requested queries before creating and populating your tables so that you don't create tables you don't need.

This fifth deliverable must be documented as follows:

CREATE TABLE statements. Turn in all SQL statements used to create tables.
SELECT * FROM each_tbl. Turn in the complete contents of all tables in the prototype database.
Please do NOT turn in any SQL INSERT statements.

****Include the name of the database in the memo where your prototype tables are stored.**

What are the deliverables for Part 3?

1. **Memo.** Write the memo as though you are writing to a client who has requested the database prototype. Avoid writing this to “your teacher” – make the assumption that you have been hired by Replica Toys to create a database for post-sales product registration and problem reporting, and then write your memo to that client. Make sure you cover the following in the memo:
 - **An explanation of the data model.** When describing your data model, you can assume that the reader understands how to read an ERD. Explain how your database design will produce the information desired by the client. **Include both your logical and physical ERD's as appendices to the memo.**

- **A description of any processes that must be instituted to use the new database.** How do you envision the organization using your database as part of an entire system? What additional components must be built to put this database into actual operation?
 - **A recommendation for the next step in the implementation of the system.** Imagine you are being paid to develop this prototype for Replica Toys. Use the memo to help "sell" your next set of services by outlining what tasks should occur next in the "recommended next step" section. To write the recommendation, think back to a class where you learned about the systems development life cycle (SDLC) or re-read chapter 1 of your database text to understand what should happen after you design a database. What happens after you present a prototype to your client? What should you do next to get agreement that the data model is a good representation of the data required for the application? This is what you need to write as the recommendation for the next step.
2. **SQL Queries.** Turn in the code and the output for the queries below using the same format as used on the SQL homework assignments (HW#6,7,8). Please place the code before the output generated in the printed output for each requested query. Provide documentation/comments for all views so that I understand the purpose of the views in your code.
 3. **Electronic version of the SQL Queries.** Upload a file (only one file, not separate files for each query!) with the SQL code for your queries to the IS475/ReplicaToys directory on the h: assignment uploads drive. Include the code for any views you created. Use the last name of one (or all) of the team members in the name of the file. I may want to test your queries with your database, so I need an electronic copy of the SQL code for the queries. Do not include screen shots (pictures) of your result tables or screen shots from SQL Server in this file. I need to be able to execute the SQL code in this file.
 4. **The name of the database where your test tables reside.** I have access to all databases on our server, so I don't need a login but I need to know the name of the database to look at your database and run your queries online.

What are the required SQL queries?

The SQL queries are designed to test some aspects of the registration part of your database – there are no queries required for quality control part of the database. The sample formats provided below include “fake” data – I made it up just so that I could provide a sample of the sort of data that would be in the result table. You will, of course, make up your own data that is relevant to the organization.

1. List all registrations currently in the database. Include the serial # of the toy, the model #, description of the model, first and last name of the buyer, date of purchase, price, distributor, and the relationship of the buyer to the user of the toy. There should be one row in the result table per registration in the database.

Serial Number	Model Number	Model Description	Buyer Name	Purchase Date	Price	Distributor	Relationship of Buyer to User
19029388-11	JCSUV7	Jeep Cherokee Grand SUV	Swanson, Bert	3/4/2016	485.99	Claremont Toy World	Aunt/Uncle
19109381-12	FFS488	Ferrari 488 GTB	Garcia, Santiago	4/1/2016	512.16	Claremont Toy World	Grandparent

2. Summarize registrations by the relationship of the purchaser to the primary user of the toy. Provide a count of the number of toys, a sum of the price paid, average price paid, and a count of the distinct models registered. The data below is completely imaginary. While your data will look different, the structure of the output should look something like this:

Relationship to User	Count of Registrations	Count of Distinct Models	Sum of Price	Average Price
Parent	3	3	1290	430
Grandparent	12	5	5920	493.333
Aunt/Uncle	6	2	3525	587.5
Friend	0	0	0	0
Other Relative	2	2	660	330
Other	2	1	760	380

3. Display the information above only for the “relationship to user” that has the largest count. For example, in the data above, only the “grandparent” line would be displayed.
4. Similar to query #2, summarize the registrations by model #. Provide a count of the number of toys, a sum of the price paid, average price paid, earliest registration date, and latest (most recent) registration date.
5. Which model is purchased most often by grandparents?

6. We want to analyze which features are most important to people registering toys. Create a count of the features that are most important to all registrations. The result table might look something like this assuming that there are a total of 25 registrations in your dataset. The result table below might be best sorted in order by the percentage (highest to lowest). Right now, it is sorted in the same order as the list of features on the original registration form.

	Count of Times Feature is Mentioned	% of registrations mentioned
Type of Toy	12	48%
Size	3	12%
Color	7	28%
Speed	1	4%
Quality of Design	3	12%
Level of replication from original	21	84%
Safety Features	24	96%
Cost	3	12%
Sound Features	0	0%
Other	0	0%

7. Modify query # 6 to display only the features mentioned when the relationship to the purchaser is a grandparent.
8. What are the three most important features (as measured by count) to a grandparent purchasing a vehicle for a grandchild?