CMPE 2400

Database Normalization

Database Design

- An important part of Database design consists of deciding which tables our database should have.
- We want our tables to be in such as way as to provide for efficient insert and retrieval of data while minimizing replication and redundancy.
- For this we need to perform normalization of our database.
- There are 2 main approaches:
  1. Start with a few intuitive (trial) tables, then follow normalization criteria to refine the table structures.
  2. Start by building one big table, then apply normalization criteria to divide more tables and move the data, accordingly to the new tables.

Normalization

- Normalizing our database means converting our tables to a number of normal forms.
- In this course, we are interested to convert up to the 3rd normal form
- We'll go through the following optimization steps
  1. Identify or include a primary key. Atomize data items. Remove any repeating attribute groups. This creates the first normal form (1NF)
  2. Remove partial dependencies on composite primary keys. This creates the second normal form (2NF)
  3. Remove transitive (non-key) dependencies. This creates the third normal form (3NF)

0 NF – No Optimization

- Consider the following tables that store information about students and courses they are following

Personal Info	Address	Course1 Code	Course1 Name	Pre-req1	Grade 1	Teacher1 Id	Teacher1 name	Course2 Code	Course2 Name	Pre-req2	Grade 2	Teacher2 Id	Teacher2 Name	Course3 code	Course3 name	Pre-req 3	Grade 3	Teacher3 Id	Teacher 3 Name
Marvin Pearl	211, 176 Street, Edmonton, T6H5B6	CMPE 2000	Web Technologies	CMPE 1300	A	1	Justin Clause	CMPE 2400	Databases	CMPE 1666	A-	2	Jessica Farmer	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
Marjolie Caspian	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1100	Basic Programming	Grade 12 Maths	B+	1	Justin Clause	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer						
Jonny Shipper, Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause
Marvin Pearl, Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause

Converting to 1NF

- 1NF requires that each value in the table be *atomic*
  - For example, name should be stored as first name and last name
  - Marital status should be stored separately
- 1NF requires that there should not be *repeating attribute groups*

Converting To 1NF

- Step1 – Identify a primary key. If there is no column that can be the primary key, create one. In this case, we'll add student id

Student Id	Personal Info	Address	Course1 Code	Course1 Name	Pre-req1	Grade 1	Teacher1 Id	Teacher 1 name	Course2 Code	Course2 Name	Pre-req2	Grade 2	Teacher2 Id	Teacher2 Name	Course3 code	Course3 name	Pre-req 3	Grade 3	Teacher3 Id	Teacher 3 Name
2020123	Marvin Pearl	211, 176 Street, Edmonton, T6H5B6	CMPE 2000	Web Technologies	CMPE 1300	A	1	Justin Clause	CMPE 2400	Databases	CMPE 1666	A-	2	Jessica Farmer	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2020546	Marjolie Caspian	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1100	Basic Programming	Grade 12 Maths	B+	1	Justin Clause	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer						
2021234	Jonny Shipper, Married	431, Gold Crescent, Edmonton, V7W5T6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause
2021456	Marvin Pearl, Married	3021 Vista Street, Edmonton, B7A7F4	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause

Converting To 1NF

- Step 2- Atomize the items. No multi-valued columns – Use first name, last name and marital status

Student Id	First Name	Last Name	Marital Status	Address	Course1 Code	Course 1 Name	Pre-req1	Grade 1	Teacher1 Id	Teacher1 name	Course2 Code	Course2 Name	Pre-req 2	Grade 2	Teacher 2 Id	Teacher2 Name	Course 3 code	Course 3 name	Pre-req 3	Grade 3	Teacher3 Id	Teacher 3 Name
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2000	Web Technologies	CMPE 1300	A	1	Justin Clause	CMPE 2400	Databases	CMPE1666	A-	2	Jessica Farmer	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1100	Basic Programming	Grade 12 Maths	B+	1	Justin Clause	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer						
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause

Converting To 1NF

- Step 3- No repeated attribute groups

Student Id	First Name	Last Name	Marital Status	Address	Course Code	Course Name	Pre-req	Grade	Teacher Id	Teacher name
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2000	Web Technologies	CMPE 1300	A	1	Justin Clause
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2400	Databases	CMPE1666	A-	2	Jessica Farmer
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1100	Basic Programming	Grade 12 Maths	B+	1	Justin Clause
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause

Converting To 1NF

- Step 4: Identify a new primary key such that its values are unique in the table

Student Id	First Name	Last Name	Marital Status	Address	Course Code	Course Name	Pre-req	Grade	Teacher Id	Teacher name
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2000	Web Technologies	CMPE 1300	A	1	Justin Clause
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2400	Databases	CMPE1666	A-	2	Jessica Farmer
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1100	Basic Programming	Grade 12 Maths	B+	1	Justin Clause
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6	CMPE 2000	Web Technologies	CMPE 1300	A-	1	Justin Clause
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 2800	Advanced C#	CMPE 1600	B+	3	Tyler Grande
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4	CMPE 1201	Digital Logic	Grade 12-Physics	A-	2	Jessica farmer
2021456	Marvin	Pearl	Married	3021 Vista Street,	CMPE	Web	CMPE 1300	A-	1	Justin

Converting To 2NF

- 2NF Requirements – Each non-key column should depend on the entire (composite) primary key
  - Eliminate Partial Dependencies
- Steps
  - remove the partially dependent column from the original table and place into a new table
  - copy the key column(s) that the non-key column was dependent on into the new table; this will be PK for the new table AND also now FK in the original table (have created a PARENT/CHILD, also called One-to-many relationship)

Converting To 2NF

Student Id	First Name	Last Name	Marital Status	Address
2020123	Marvin	Pearl	Single	211, 176 Street, Edmonton, T6H5B6
2020546	Marjolie	Caspian	Single	321, 167 Ave, Edmonton,R5T6Y4
2021234	Jonny	Shipper	Married	431, Gold Crescent, Edmonton, V7W 5T6
2021456	Marvin	Pearl	Married	3021 Vista Street, Edmonton, B7A 7F4

Course Code	Course Name	Pre-req	Teacher Id	Teacher name
CMPE 2000	Web Technologies	CMPE 1300	1	Justin Clause
CMPE 2400	Databases	CMPE1666	2	Jessica Farmer
CMPE 2800	Advanced C#	CMPE 1600	3	Tyler Grande
CMPE 1100	Basic Programming	Grade 12 Maths	1	Justin Clause
CMPE 1201	Digital Logic	Grade 12-Physics	2	Jessica farmer

Student Id	Course Code	Grade
2020123	CMPE 2000	A
2020123	CMPE 2400	A-
2020123	CMPE 2800	B+
2020546	CMPE 1100	B+
2020546	CMPE 1201	A-
2021234	CMPE 2800	B+
2021234	CMPE 1201	A-
2021234	CMPE 2000	A-
2021456	CMPE 2800	B+
2021456	CMPE 1201	A-
2021456	CMPE 2000	A-

Converting to 3NF

- Remove Transitive Dependencies, i.e No non-key column should depend on another non-key column
- Steps
  - remove the dependent column(s) and place it in a new table
  - Copy the non-key column(s) that the other non-key column(s) depended on and place into the new table as the primary key; the matching columns in the original table must now be made FK (to the PK in the new table)

Converting to 3NF

Course Code	Course Name	Pre-req	Teacher Id
CMPE 2000	Web Technologies	CMPE 1300	1
CMPE 2400	Databases	CMPE1666	2
CMPE 2800	Advanced C#	CMPE 1600	3
CMPE 1100	Basic Programming	Grade 12 Maths	1
CMPE 1201	Digital Logic	Grade 12 Physics	2

Teacher Id	Teacher name
1	Justin Clause
2	Jessica Farmer
3	Tyler Grande

Additional Information

- The videos at the links below discuss converting up to 3rd Normal Form.

1. https://www.youtube.com/watch?v=jgUeOjlmOOw
2. https://www.youtube.com/watch?v=9L10Q1nAfyg
3. https://www.youtube.com/watch?v=_K7fcFQowy8

Normalization- A practical approach

- The previous way of converting to 1NF provides a systematic approach.
- However, for small systems, we can have a different approach that can provide us with more than 1 table from the 1NF.
- We'll again consider the example of students and courses that we had seen before.
- This approach is particularly useful when we work from a document
- The next few slides go through the required steps.

Normalization- A practical approach

- Step 1-List all the available attributes (including multi-valued ones), with a mention repeated groups. This will be our 0NF
- Student:
  - Personal info (first name, lastname, marital status)
  - Address
  - Course Info (repeated)
    - Course Code
    - Course Name
    - Pre-req
    - Teacher Id
    - Teacher Name (first name +last name)
    - Grades

Normalization- A practical approach

- Step 2 – Atomize
- Student:
  - first name
  - Last name
  - marital status
  - Address
  - Course Info (repeated)
    - Course Code
    - Course Name
    - Pre-req
    - Teacher Id
    - Teacher - first name
    - Teacher - last name
    - Grades

Normalization- A practical approach

- Step 3- Check if a primary key exists, else add one
- Student:
  - Student Id (PK)
  - first name
  - Last name
  - marital status
  - Address
  - Course Info (repeated)
    - Course Code
    - Course Name
    - Pre-req
    - Teacher Id
    - Teacher - first name
    - Teacher - last name
    - Grades

Normalization- A practical approach

- Step 4 – Create a table with the non-repeated attributes and another table for the repeated attributes.
- The first table will have as primary key the key that we had identified. Identify a suitable primary key for the second table and also make a copy of the initial primary key to the second table (It will form part of the PK but also be an FK)

Students

StudentId
First name
Last Name
Address

Courses

StudentId
Course Code
Course Name
Prereq
Teacher Id
Teacher First name
Teacher last Name
Grades

- These tables satisfy 1NF.

Normalization- A practical approach

- Converting to 2NF.
- The previous tables satisfy 1NF. However, in the second table, we note that except for grades which depend on the full composite primary key. All the remaining columns depend only on Course Code.
- To convert to 2NF, we remove partial dependencies. So, we move all the course information to a different table with Course code as PK. We rename the initial table. Course code also becomes an FK in the initial table.

Students

StudentId
First name
Last Name
Address

Students-Courses

StudentId
Course Code
Grades

Courses

Course Code
Course Name
Prereq
Teacher Id
Teacher First name
Teacher last Name

- Our tables are now in 2NF

Normalization- A practical approach

- Converting to 3NF
- To convert to 3NF, we remove transitive dependencies
- In this case, we remove teacher first name and last name form the courses table and move them into a new table with Teacher Id as PK. Teacher Id becomes a foreign key in the courses table.

Students

StudentId
First name
Last Name
Address

Students-Courses

StudentId
Course Code
Grades

Courses

Course Code
Course Name
Prereq
Teacher Id
Teacher First name
Teacher last Name

Teachers

Teacher Id
Teacher First name
Teacher last Name