FINAL PROJECT

Rishabh Taneja

Northeastern University

Data Warehousing and SQL

FALL ‘18

**Part 1) Normalization**

* 1. **For each fact variable in your fact table, what type of fact is it? Additive, semi-additive, or non-additive?**

The facts are ‘copay’ and ‘insurancepaid’. They are additive since they can be summed up across every dimension related to the fact table.

* 1. **In your fact table, describe the *grain* in one sentence. What does each fact row represent?**

The fact\_table has a grain of day, considering fill\_date.

The fact\_table has a grain of amount paid by the member, considering copay.

The fact\_table has a grain of amount of insurance paid, considering insurancepaid.

**Part 2) Primary and Foreign Key Setup in MySQL**

ALTER TABLE `abc`.`dim\_brand\_details`

CHANGE COLUMN `drug\_brand\_generic\_code` `drug\_brand\_generic\_code` INT NOT NULL ,

CHANGE COLUMN `drug\_brand\_generic\_desc` `drug\_brand\_generic\_desc` VARCHAR(100) NOT NULL ,

ADD PRIMARY KEY (`drug\_brand\_generic\_code`);

ALTER TABLE `abc`.`dim\_drug\_details`

CHANGE COLUMN `drug\_ndc` `drug\_ndc` INT(11) NOT NULL ,

CHANGE COLUMN `drug\_name` `drug\_name` VARCHAR(100) NOT NULL ,

CHANGE COLUMN `drug\_form\_code` `drug\_form\_code` CHAR(2) NOT NULL ,

CHANGE COLUMN `drug\_brand\_generic\_code` `drug\_brand\_generic\_code` INT(11) NOT NULL ,

ADD PRIMARY KEY (`drug\_ndc`);

ALTER TABLE `abc`.`dim\_drug\_details`

ADD INDEX `drug\_form\_code\_idx` (`drug\_form\_code` ASC) VISIBLE,

ADD INDEX `drug\_brand\_generic\_code\_idx` (`drug\_brand\_generic\_code` ASC) VISIBLE;

ALTER TABLE `abc`.`dim\_drug\_details`

ADD CONSTRAINT `drug\_form\_code`

FOREIGN KEY (`drug\_form\_code`)

REFERENCES `abc`.`dim\_form\_details` (`drug\_form\_code`)

ON DELETE RESTRICT

ON UPDATE CASCADE,

ADD CONSTRAINT `drug\_brand\_generic\_code`

FOREIGN KEY (`drug\_brand\_generic\_code`)

REFERENCES `abc`.`dim\_brand\_details` (`drug\_brand\_generic\_code`)

ON DELETE RESTRICT

ON UPDATE CASCADE;

ALTER TABLE `abc`.`dim\_form\_details`

CHANGE COLUMN `drug\_form\_code` `drug\_form\_code` CHAR(2) NOT NULL ,

CHANGE COLUMN `drug\_form\_desc` `drug\_form\_desc` VARCHAR(100) NOT NULL ,

ADD PRIMARY KEY (`drug\_form\_code`);

ALTER TABLE `abc`.`dim\_member\_details`

CHANGE COLUMN `member\_id` `member\_id` INT(11) NOT NULL ,

CHANGE COLUMN `member\_first\_name` `member\_first\_name` VARCHAR(100) NOT NULL ,

CHANGE COLUMN `member\_last\_name` `member\_last\_name` VARCHAR(100) NOT NULL ,

CHANGE COLUMN `member\_birth\_date` `member\_birth\_date` DATE NOT NULL ,

CHANGE COLUMN `member\_age` `member\_age` INT(11) NOT NULL ,

CHANGE COLUMN `member\_gender` `member\_gender` CHAR(2) NOT NULL ,

ADD PRIMARY KEY (`member\_id`);

ALTER TABLE `abc`.`fact\_fact\_table`

CHANGE COLUMN `fact\_id` `fact\_id` INT(11) NOT NULL ,

CHANGE COLUMN `member\_id` `member\_id` INT(11) NOT NULL ,

CHANGE COLUMN `fill\_date` `fill\_date` DATE NOT NULL ,

CHANGE COLUMN `copay` `copay` INT(11) NOT NULL ,

CHANGE COLUMN `insurancepaid` `insurancepaid` INT(11) NOT NULL ,

CHANGE COLUMN `drug\_ndc` `drug\_ndc` INT(11) NOT NULL ,

ADD PRIMARY KEY (`fact\_id`),

ADD INDEX `member\_id\_idx` (`member\_id` ASC) VISIBLE,

ADD INDEX `drug\_ndc\_idx` (`drug\_ndc` ASC) VISIBLE;

ALTER TABLE `abc`.`fact\_fact\_table`

ADD CONSTRAINT `member\_id`

FOREIGN KEY (`member\_id`)

REFERENCES `abc`.`dim\_member\_details` (`member\_id`)

ON DELETE RESTRICT

ON UPDATE CASCADE,

ADD CONSTRAINT `drug\_ndc`

FOREIGN KEY (`drug\_ndc`)

REFERENCES `abc`.`dim\_drug\_details` (`drug\_ndc`)

ON DELETE RESTRICT

ON UPDATE CASCADE;

* 1. **What are the primary keys you designated for each of your tables? For each PK, is it a natural key or a surrogate key?**

1. Dim\_Member\_details – The primary key designated is ‘member\_id’ and it is a natural key as it identifies each row in the table uniquely.
2. Dim\_Drug\_details – The primary key designated is ‘drug\_ndc’ and it is a natural key as it identifies each row in the table uniquely.
3. Dim\_Brand\_details – The primary key designated is ‘drug\_brand\_generic\_code’ and it is a natural key as it identifies each row in the table uniquely.
4. Dim\_Form\_details – The primary key designated is ‘drug\_form\_code’ and it is a natural key as it identifies each row in the table uniquely.
5. Fact\_Fact\_table – The primary key designated is ‘Fact\_id’ and it is a surrogate key.
   1. **What are the foreign keys you designated for each of your tables? For each FK, which table did you reference where that FK is listed as the PK?**
6. Dim\_Drug\_details – The foreign keys designated here are ‘drug\_form\_code’ and ‘drug\_brand\_generic\_code’ which are referenced to the tables ‘dim\_brand\_details’ and ‘dim\_form\_details’ respectively.
7. Fact\_Fact\_table – The foreign keys designated here are the ‘member\_id’ and ‘drug\_ndc’ which are referenced to the tables ‘dim\_member\_details’ and ‘dim\_drug\_details’ respectively.
   1. **For each FK, what did you tell MySQL to in case of deletion or update (CASCADE, SET NULL, or RESTRICT)? Why did you select the option that you did for each FK?**
8. ‘member\_id’

Delete – I used ‘Restrict’ keyword as it will then delete the information of the members from the table ‘dim\_member\_details’ but will restrict the system to delete the records from the table ‘fact\_fact\_table’. This is also used to retain the historical information.

Update – I used ‘Cascade’ keyword as it will also update the member information across the ‘fact\_fact\_table’ if any record is updated in the ‘dim\_member\_details’.

1. ‘drug\_ndc’

Delete - I used ‘Restrict’ keyword as it will then delete the information of the drugs from the table ‘dim\_drug\_details’ but will restrict the system to delete the records from the table ‘fact\_fact\_table’. This is also used to retain the historical information of the drug details even after the details are deleted.

Update - I used ‘Cascade’ keyword as it will also update the drug information across the ‘fact\_fact\_table’ if any record is updated in the ‘dim\_drug\_details’.

1. ‘drug\_form\_code’

Delete - I used ‘Restrict’ keyword as it will then delete the information of the forms from the table ‘dim\_form\_details’ but will restrict the system to delete the records from the table ‘dim\_drug\_details’. This is also used to retain the historical information of the drug and form details even after the details are deleted.

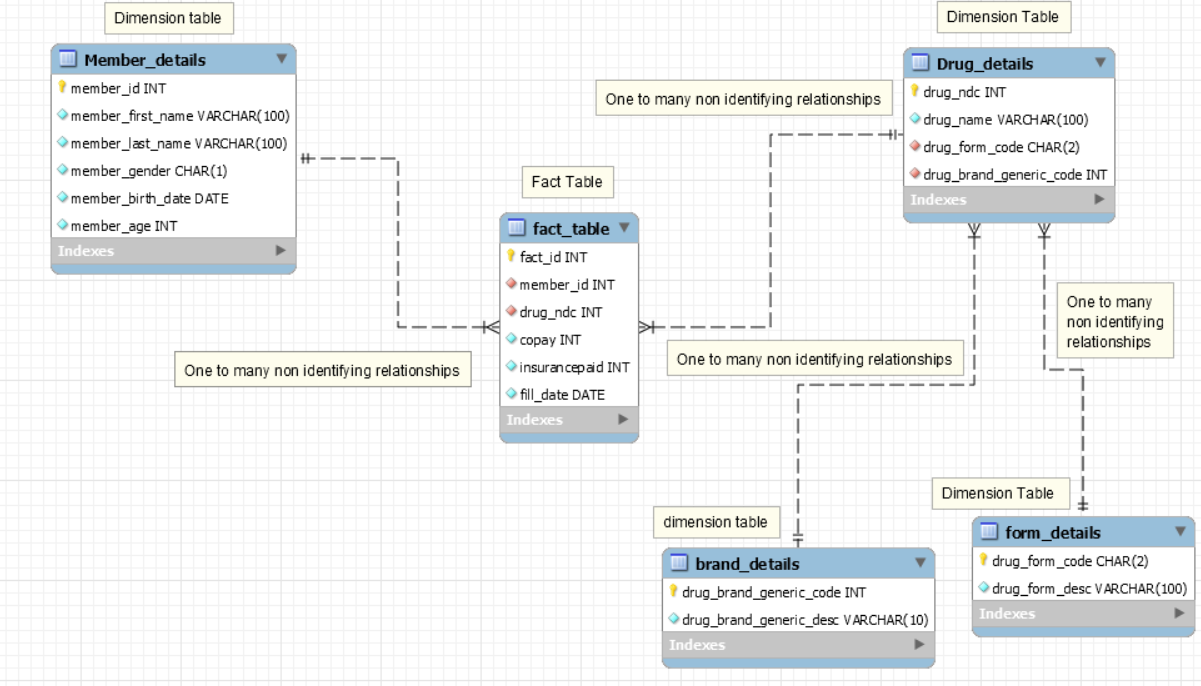
Update - I used ‘Cascade’ keyword as it will also update the drug information across the ‘dim\_drug\_details’ if any record is updated in the ‘dim\_form\_details’.

1. ‘drug\_brand\_generic\_code’

Delete - I used ‘Restrict’ keyword as it will then delete the information of the forms from the table ‘dim\_brand\_details’ but will restrict the system to delete the records from the table ‘dim\_drug\_details’. This is also used to retain the historical information of the drug brand details even after the details are deleted.

Update - I used ‘Cascade’ keyword as it will also update the drug information across the ‘dim\_drug\_details’ if any record is updated in the ‘dim\_brand\_details’.

**Part 3) Entity Relationship Diagram**



**Part 4) Analytics and Reporting**

SELECT

ddd.drug\_name, COUNT(fft.fill\_date) AS no\_of\_prescription

FROM

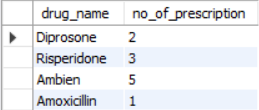
fact\_fact\_table fft,

dim\_drug\_details ddd

WHERE

fft.drug\_ndc = ddd.drug\_ndc

GROUP BY drug\_name;



1. How many prescriptions were filled for the drug Ambien?

**5**

SELECT

COUNT(fft.fill\_date) AS no\_of\_prescription,

COUNT(DISTINCT fft.member\_id) AS distinct\_members,

SUM(fft.copay) AS sum\_of\_copay,

SUM(fft.insurancepaid) AS sum\_of\_insuarance\_paid,

CASE

WHEN dmd.member\_age < 65 THEN 'Less than 65'

WHEN dmd.member\_age > 65 THEN 'greater than 65'

END AS age\_group

FROM

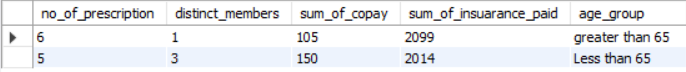
dim\_member\_details dmd,

fact\_fact\_table fft

WHERE

fft.member\_id = dmd.member\_id

GROUP BY age\_group;



1. How many unique members are over 65 years of age?

**1**

1. How many prescriptions did they fill?

**6 no. of prescription for members greater than 65 years old (if this is the follow up question of the above question then answer is 6)**

**5 no. of prescription for less than 65 years old**

**Total = 11**

CREATE TABLE fact\_final AS SELECT \* FROM

fact\_fact\_table

ORDER BY fill\_date DESC;

SELECT

ff.member\_id,

dmd.member\_first\_name,

dmd.member\_last\_name,

ddd.drug\_name,

fill\_date,

ff.insurancepaid

FROM

fact\_final ff,

dim\_drug\_details ddd,

dim\_member\_details dmd

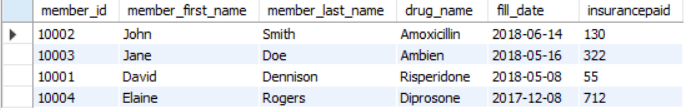
WHERE

ff.drug\_ndc = ddd.drug\_ndc

AND ff.member\_id = dmd.member\_id

GROUP BY ff.member\_id

ORDER BY fill\_date DESC;



1. **For member ID 10003**, what was the drug name listed on their most recent fill date?

**Ambien**

1. How much did their insurance pay for that medication?

**$322**