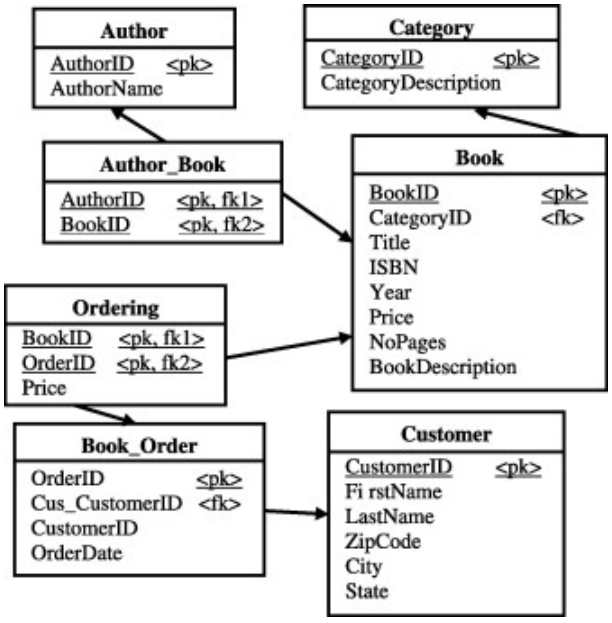
Design star schema for relatonal database

Asked 4 years, 11 months ago Modified 4 years, 11 months ago Viewed 2k times

1

I'm a newbie to data warehousing and I've been reading articles and watching videos on the principles but I'm a bit confused as to how I would take the design below and convert it into a star schema. in this example i assume that the fact table is (order-orderitem-book) And the measures is (category-customer-time) My question is about book author how can we put is as measure? Is it allowed to put many to many relationship in star schema ?? And if i am wrong how to draw star schema to this relatonal db?



data-warehouse star-schema

Share Improve this question Follow

edited Dec 9, 2018 at 21:54

asked Dec 9, 2018 at 21:24

J. DOE
47 6

2 Answers

Sorted by: Highest score (default)

You could put a many-to-many relationship within a data warehouse, but many people consider it bad practice to do so - even so far as some data warehousing tools do not permit it to be created at all. Here is how I would create a star-schema from your design:

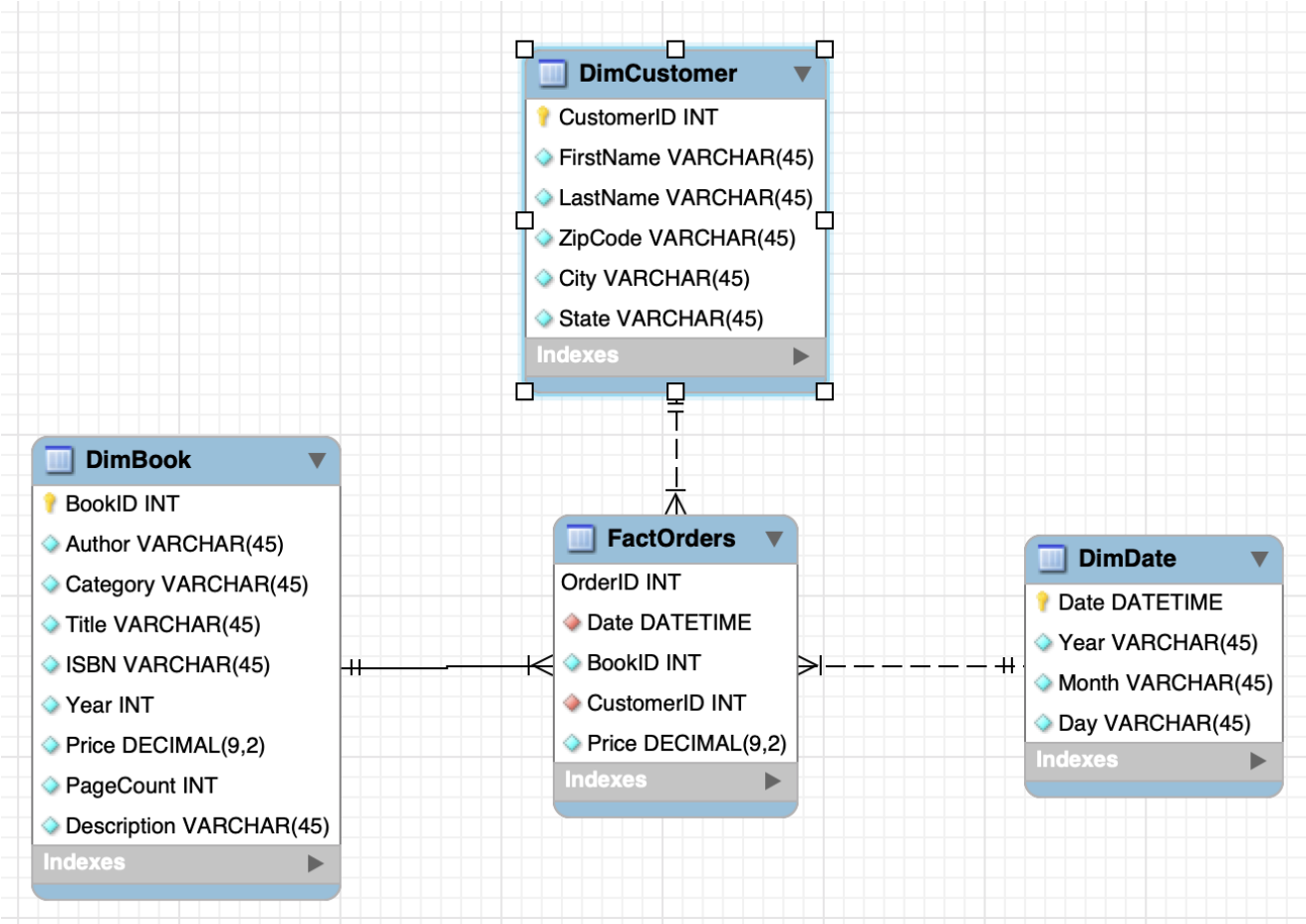
As your `Author` table and `Category` table only have one valuable attribute (the name) I would roll them into the `Book` table which would then become your first dimension. The `Customer` table can stay as-is and become a dimension as well. You would then roll the two `Order` tables into one and create a `Order` fact table which consists of `OrderID`, `Date`, `BookID`, `CustomerID`, `Price` - like so:

```
CREATE TABLE DimBook
(
    BookID      INT          NOT NULL PRIMARY KEY,
    Author      VARCHAR(50)  NOT NULL,
    Category    VARCHAR(50)  NOT NULL,
    Title       VARCHAR(50)  NOT NULL,
    ISBN        VARCHAR(50)  NOT NULL,
    Year        SMALLINT     NOT NULL,
    Price       DECIMAL(9,2) NOT NULL,
    NoPages     SMALLINT     NOT NULL,
    Description VARCHAR(100) NOT NULL
);

CREATE TABLE DimCustomer
(
    CustomerID INT         NOT NULL PRIMARY KEY,
    FirstName  VARCHAR(50) NOT NULL,
    LastName   VARCHAR(50) NOT NULL,
    ZipCode    VARCHAR(20) NOT NULL,
    City       VARCHAR(50) NOT NULL,
    State      VARCHAR(50) NOT NULL
);

CREATE TABLE FactOrders
(
    OrderID    INT          NOT NULL,
    "Date"     DATETIME     NOT NULL,
    BookID     INT          NOT NULL REFERENCES DimBook(BookID),
    CustomerID INT          NOT NULL REFERENCES DimCustomer(CustomerID),
    Price      DECIMAL(9,2) NOT NULL
);
```

You may also want to consider a `Date` dimension which is also commonly found in star-schemas and data warehouses to make searching by dates easier. A very basic implementation is below:

```
CREATE TABLE DimDate
(
    "Date"  DATETIME NOT NULL PRIMARY KEY,
    "Year"  SMALLINT NOT NULL,
    "Month" TINYINT  NOT NULL,
    "Day"   TINYINT  NOT NULL
);
```

Then, just add a foreign key from your `Date` attribute in the fact table to the `Date` key in the `DimDate` table. This would produce something like:



If you need to handle scenarios where a book can have many authors (which frequently happens), there are a couple of ways to do so.

The first, and my recommendation, is to have all of the authors within the `Author` attribute. This would allow you to easily search for all books written by the same combination of authors.

The second approach denormalises the `Author` attribute into its own dimension which is then referenced by the book dimension. This would create a snowflake schema (your question stated you wanted a star schema so I avoided this approach) and would also be slower when trying to search by multiple authors.

Ultimately, it depends on your exact needs and the requirements you are trying to meet. I would personally stick with having all authors in the same attribute as this is the easiest design and meets your star schema requirement.

Share Improve this answer Follow

edited Dec 9, 2018 at 23:31

answered Dec 9, 2018 at 22:43

Mr.Brownstone
13k 4 36 54

1 What if the book has many authors? How can i roll them into book dimention? – J. DOE Dec 9, 2018 at 23:03

I have updated my answer accordingly. – Mr.Brownstone Dec 9, 2018 at 23:25

If i replace the category with publisher and inside it i put city country to make hierarchy on its location. Can i but the publisher on the book dimention in this case? – J. DOE Dec 9, 2018 at 23:50

Yes, that would not be a problem. – Mr.Brownstone Dec 10, 2018 at 0:02

0

So your question is a couple of different questions -

1. `Author` should not be its own dimension, it will just be an attribute of the `Book` dimension.
2. Because a fact table's primary key is a composite key made up of a set of foreign keys, every table that has a many-to-many relationship has to be expressed as a fact table. You'll have to employ the use of bridge tables, but the best way to implement this depends on your needs.
3. I don't think you're wrong in your approach, but just to help you clarify what you're doing, you'll want `Order` as a fact table, and `Book` (which I would move `Author` and `Category` into as attributes) `DateTime` (or `Date` and `Time` separate from each other) and `Customer` as dimensions in your example. All your quantitative data (other than `DateTime`) should be going in `Order` and all your descriptive and qualitative data should be going in your surrounding dimensions.

Share Improve this answer Follow

answered Dec 9, 2018 at 23:09

Rhys

1 3

If i consider book as a dimention how can i put author in book dimention if the book has many authors? – J. DOE Dec 9, 2018 at 23:18

There are a few ways you could do that. You could put all of the authors into the `Author` attribute and separate them (i.e. Author:"John Smith, John Doe, Jane Citizen), or you could snowflake the `Book` dimension so that `Author` becomes an outrigger (you can think of it as a subdimension). Although it isn't pretty, it's probably best to put all of the authors into `Author` and set it as a VARCHAR with a high character limit. – Rhys Dec 9, 2018 at 23:28