Understanding Star and Snowflake Schemas

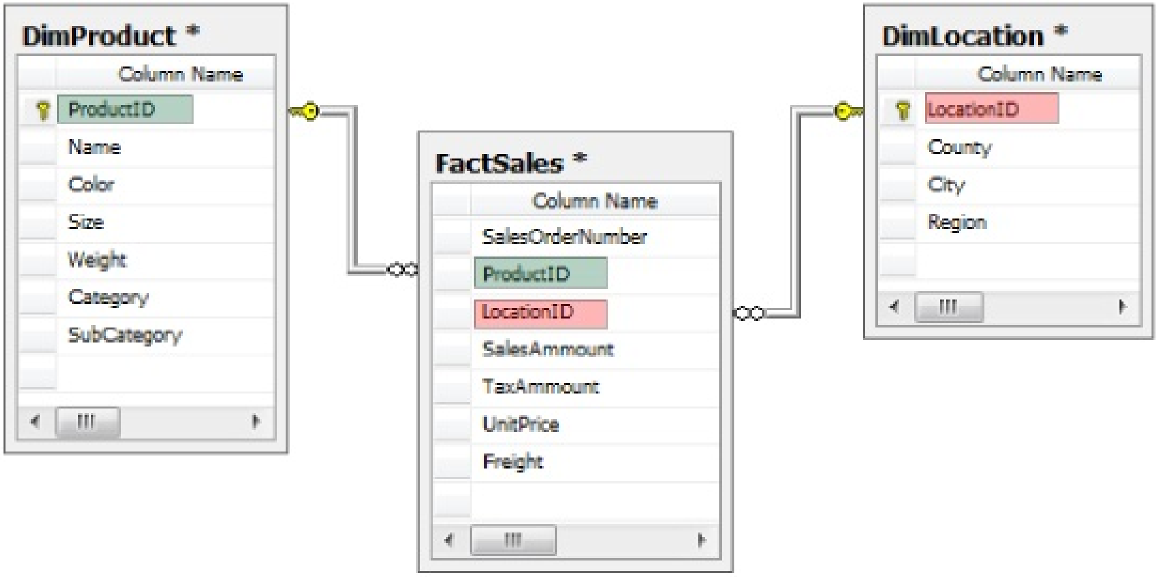
Dimensions Review

This article discusses how to design your model using the star or snowflake schemas. So dimensions are the objects or business entities in your database, there a group of related attributes about that object. Now it's important to make the distinction between dimension tables and the dimensions themselves, when we're creating our unified dimensional model were going to be modelling our dimension tables that ultimately we will create a dimension from so don't get confused thinking they are the same thing. I think when we review the approach to modelling the data you'll see how dimension tables and dimensions are different.

And in the dimensions article we talked about measures now measures are the data values that you want to report on, they are numeric and usually additive and the measures are stored in fact tables. A good way to define your dimensions and fact tables for the first time can usually be in a way that people describe their reports they like to see, for example you might hear a user that says they would like to see a report that shows sales by date and product line. Well in this example the sales data would be a measure and therefore be in your fact table and date and product would be your dimension tables

Model our Database Star or Snowflake

We have two design approaches we can follow and they are either the star or snowflake schemas. The star schema is the simpler of the two, the star schema has a few basic rules each dimension is represented by one dimension table. So here there is a one to one match between the dimension tables that you're modelling and the actual dimensions that will be built into your OLAP cube. The dimension table is also related to or linked to a fact table, so all of your dimension tables in the star schema in your model will be linked to a fact table.



DimProduct *

	Column Name
🔑	ProductID
	Name
	Color
	Size
	Weight
	Category
	SubCategory

FactSales *

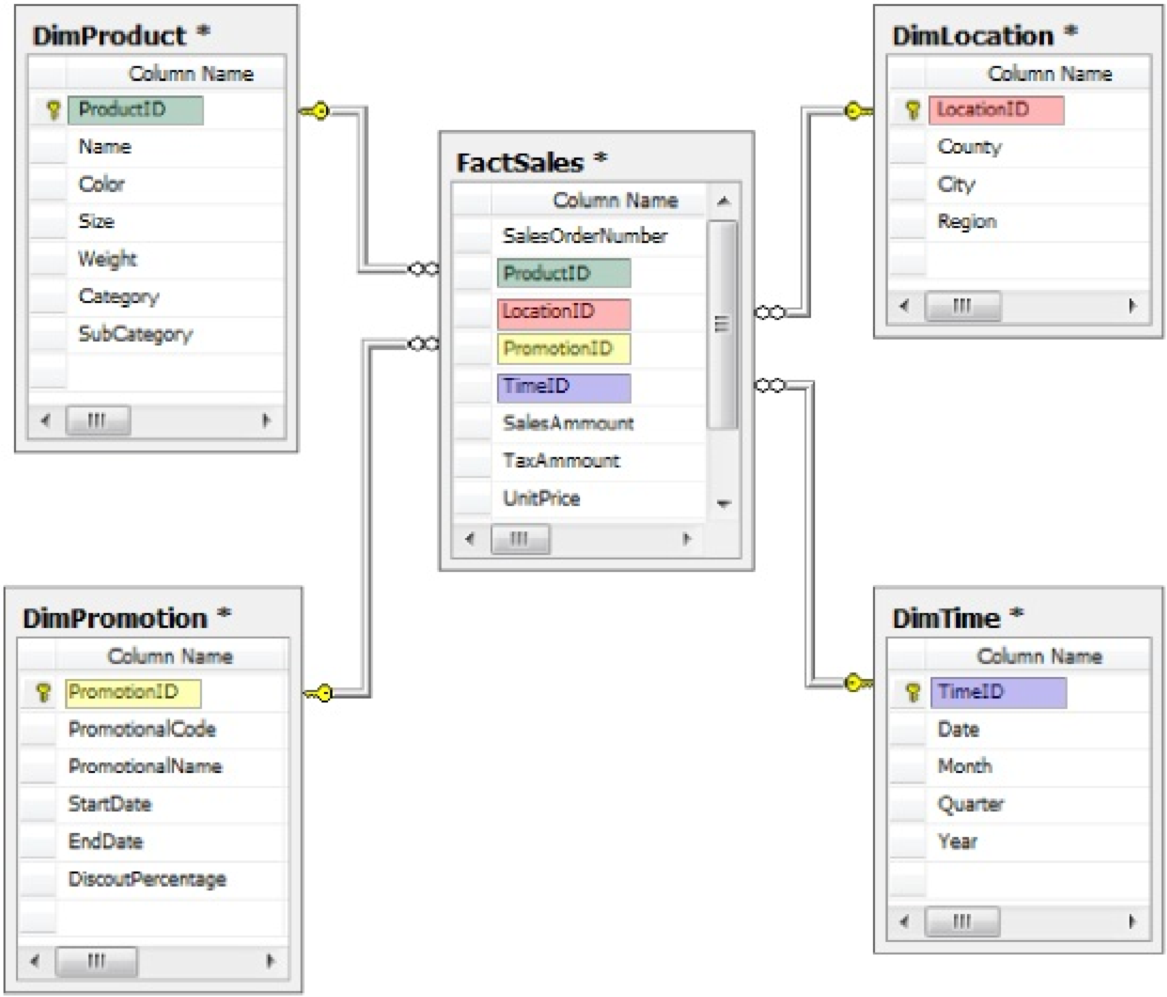
	Column Name
	SalesOrderNumber
	ProductID
	LocationID
	SalesAmmount
	TaxAmmount
	UnitPrice
	Freight

DimLocation *

	Column Name
🔑	LocationID
	County
	City
	Region

You can see in the above diagram we have two dimension tables that are linked to a fact table, we have DimProduct and DimLocation. The fact table here is in the middle of the diagram each dimension table is linked directly to the fact table and each dimension table is represents one dimension. The naming convention you see is widely used when working with dimensional modelling, the dimension tables are prefixed with Dim and the fact tables are prefixed with Fact.

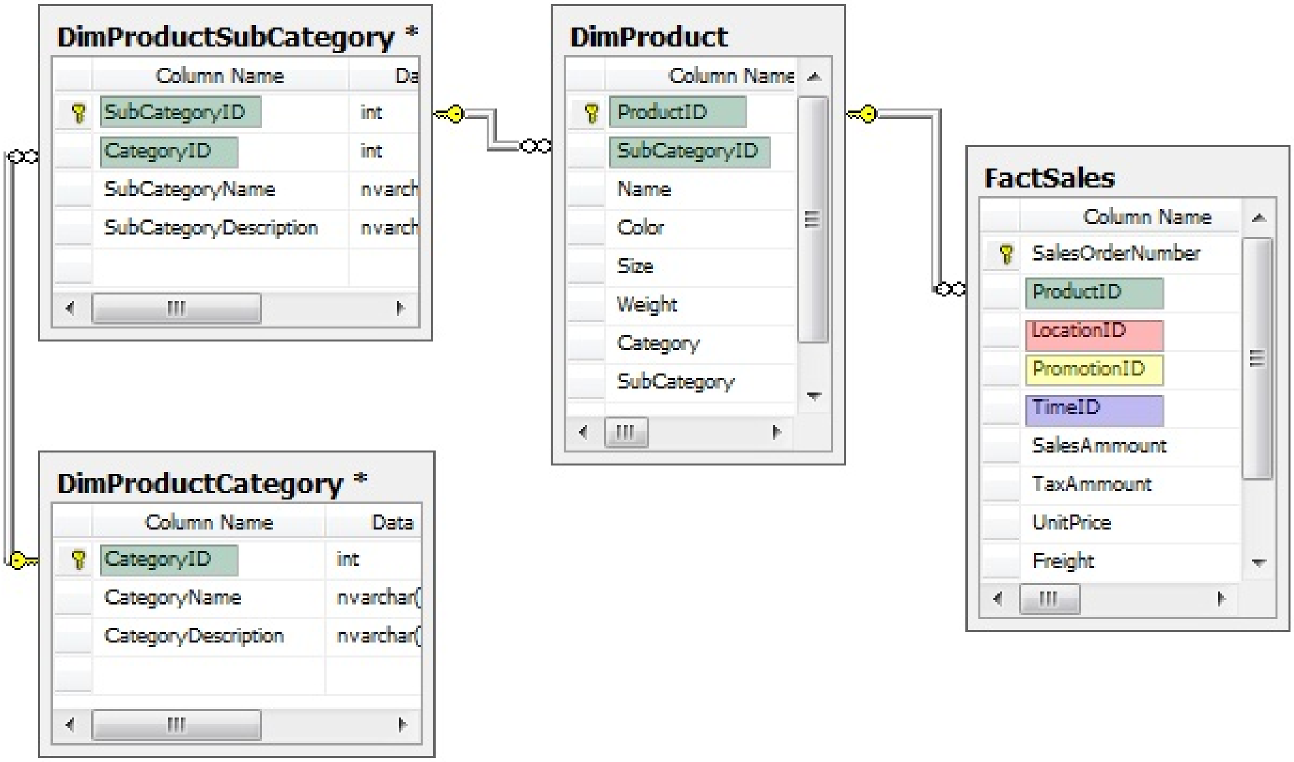
Star Schema



So let expand on the diagram a little so you can see how this resembles a star which is where the star schema got its name from but also notice the relationships between the dimension tables and the fact tables are via primary keys and foreign keys. You can see with the DimProduct dimension at the top left has a primary key of ProductID which can be found in the FactSales table in the middle as a foreign key. So the FactSales table is just a collection of foreign keys to all the dimensions in the star schema and the measures you want to report on and you can see that represented in the above diagram. The fact table in your dimensional model in this case FactSales will most likely have most of the data and the rows as compared to the dimensional tables since its actually holding all of the detail that you will be reporting on. So take some time to study this star schema and really pay particular attention to the relationships between each of the dimension tables and the fact tables. So if look just a little bit longer at the FactSales table you can see that we have the SalesOrderNumber which is the primary key for the fact sales table itself, it’s not related to any dimensions and then we have the four foreign keys each representing the primary key for the four dimensions that its related to DimProduct by ProductID, DimLocation by LocationID, DimPromotion by PromotionID then DimTime by TimeID. The remaining fields’ SaleAmmount, TaxAmmount and UnitPrice are the measures so each of these measures can then be viewed by either of those dimensions.

Snowflake Schema

Next we have the snowflake schema which varies slightly from the star schema, with the snowflake schema the dimensions are represented by more than one dimension table in other words its takes multiple dimension tables to define a dimension. Not all dimension tables are linked or related to the fact table because some of the dimensions will just be related to other dimensions.



DimProductSubCategory *

Column Name	Da
SubCategoryID	int
CategoryID	int
SubCategoryName	nvarch
SubCategoryDescription	nvarch

DimProduct

Column Name
ProductID
SubCategoryID
Name
Color
Size
Weight
Category
SubCategory

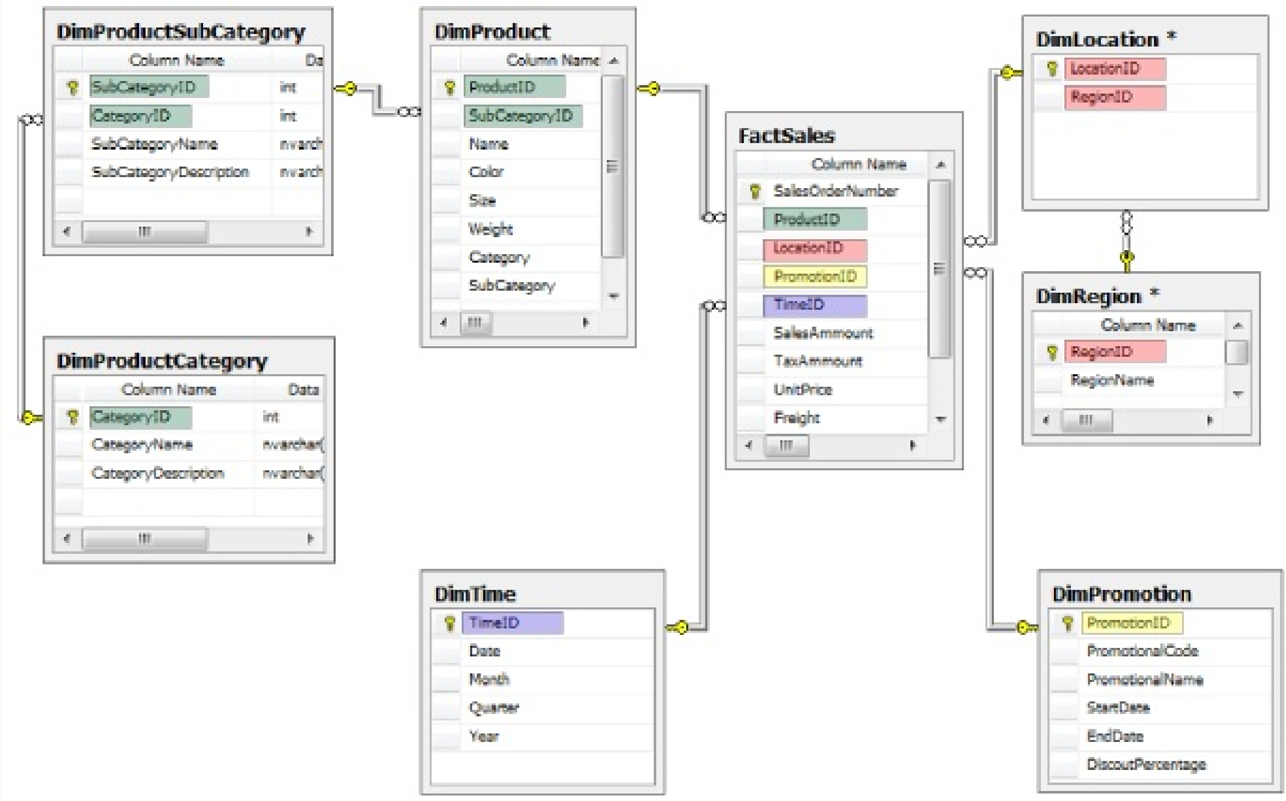
FactSales

Column Name
SalesOrderNumber
ProductID
LocationID
PromotionID
TimeID
SalesAmmount
TaxAmmount
UnitPrice
Freight

DimProductCategory *

Column Name	Data
CategoryID	int
CategoryName	nvarchar(
CategoryDescription	nvarchar(

So you can see in the diagram we have three dimension tables that together represent one dimension, they represent the pPoduct dimension. So with the star schema approach all the information in the three dimension tables shown here would actually be in only one dimension table, this is where the snowflake schema really differs in approach from the star. Also note since the dimension is represented by more than one dimension table not all the dimension tables are related to the fact table. So in the next diagram a few more tables have been added to get a better picture of what the schema will look like and the name snowflake comes from the idea somehow resembles a snowflake. At any rate there are some very key differences between this and the star schema we just pointed out.



Which way you choose to go has in my mind has a lot to do with your personal preference and there's an ongoing debate that continues to discuss which of these is better, honestly there are advantages and disadvantages of both and my experience I've seen the star schema used the most and preferred by many it's simple to work with, but really the choice is up to you and both will work in analysis services.

Conclusion

Star Schema

- Dimensions represented by one Dimension table
- Each Dimension table related to Fact table

Snowflake schema

- Dimensions represented by more than one Dimension table
- Not all Dimension tables related to Fact table

Next article Data Source Views

Related Articles

- SSAS Workflow
- UDM
- Dimensions Modelling
- Schemas

- About
- Contact Us
- Privacy Policy
- Terms of Use