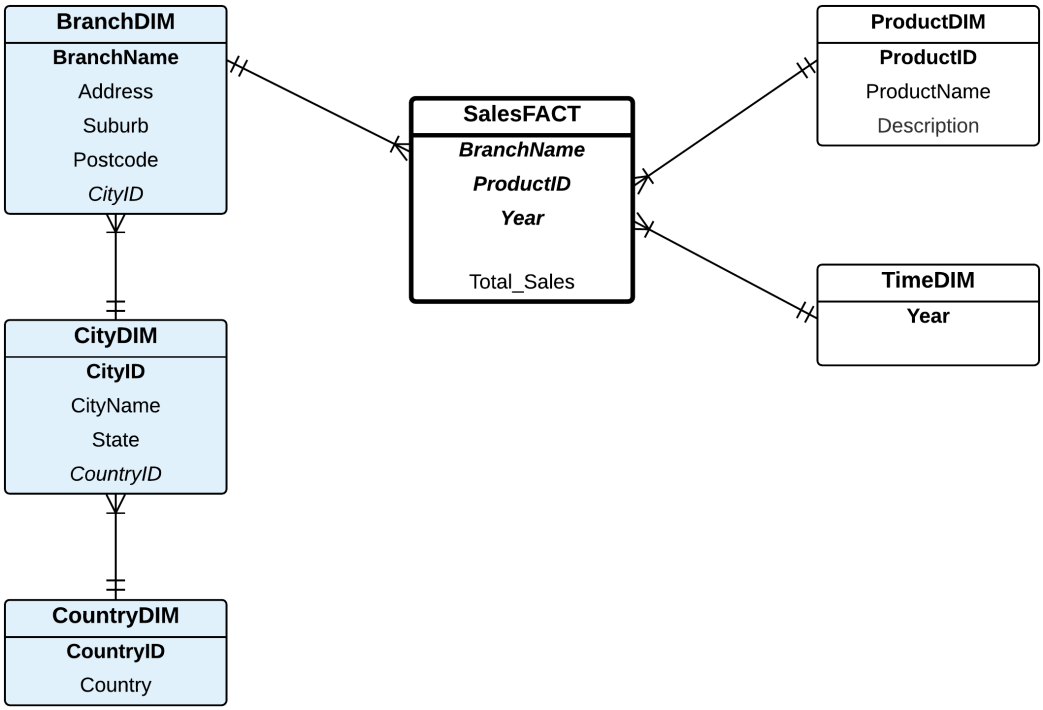
Level of Aggregation, Hierarchy, and Multi-Fact

In this lesson, we are going to explore data warehousing architecture (or level of aggregation), in the context of hierarchy and multi-fact. The following star schema is a simple Sales star schema with hierarchy. The hierarchy is Branch, City, and Country dimensions. There are two other dimensions: Product and Time (Year). The Sales Fact has one fact measure, which is Total Sales.

The Sales Fact shows that the Total Sales is actually based on the Branch granularity. The hierarchy is merely an implementation of the normalization concept in the relational database management systems, in order to minimize repetition and to avoid anomalies. The hierarchy is not meant to be for drilling down or rolling up of the fact measure.



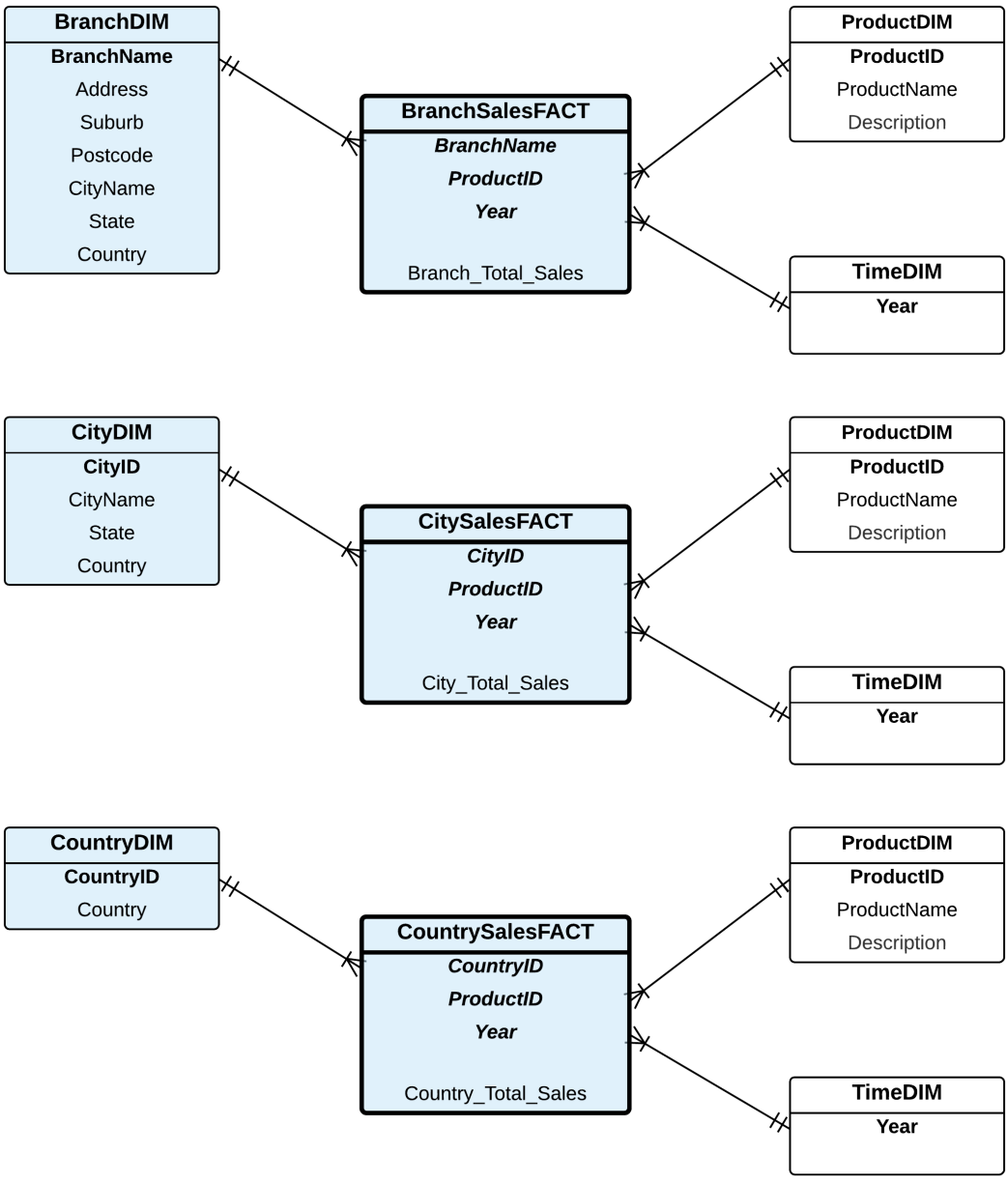
The hierarchy actually shows the level of granularity. The hierarchy should start from the most detail; in this case: Branch, and goes down to more general: City, and then Country.

If we split these three dimensions in the hierarchy into three star schemas, then we will have three fact tables. The three star schemas are shown below. The first star schema uses the Branch Dimension, and therefore, the Total Sales is based on the Branch granularity. In order to differentiate this from the other two star schemas, the fact table is called the Branch Sales Fact, with Branch Total Sales as the fact measure.

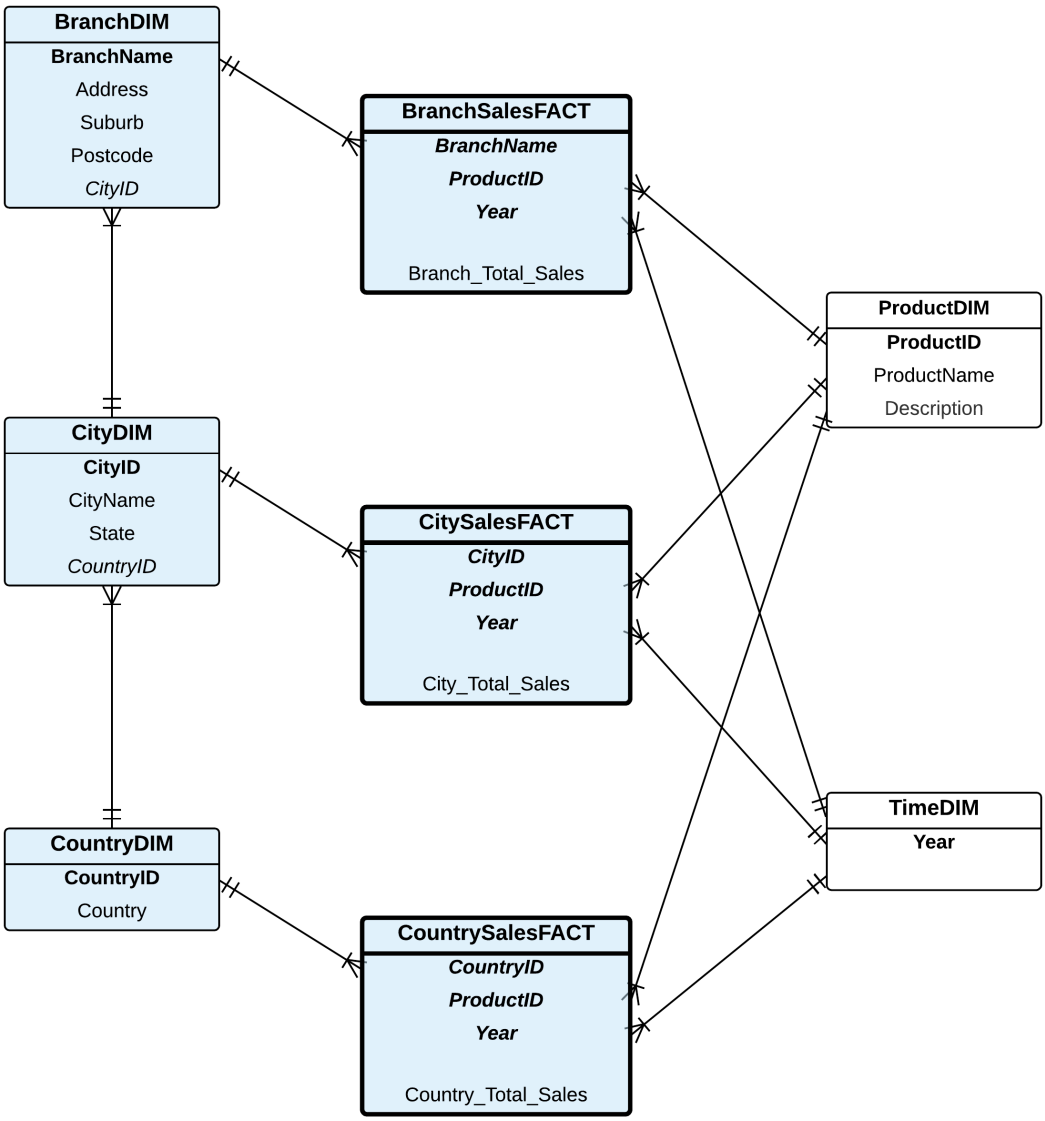
The second star schema uses the City dimension, instead of the Branch. Therefore, the Total Sales fact measure is based on the City granularity. The Fact table is renamed to City Sales Fact with City Total Sales as the fact measure. Note that this star schema (i.e. City Star Schema) is more general than the Branch star schema.

The third star schema uses the Country dimension, which is the most general. The Fact table becomes Country Sales Fact with Country Total Sales as the fact measure.

These three star schemas (i.e. Branch star schema, City star schema, and Country star schema) focus on the same fact measure, which is Total Sales. The difference is that they have different level of granularity or level of aggregation. The Branch star schema is the most detail, whereas the County star schema is the most general. Hence, there are three levels in this data warehousing architecture. Level-1 for the Branch star schema, Level-2 for the City star schema, and Level-3 for the Country star schema.

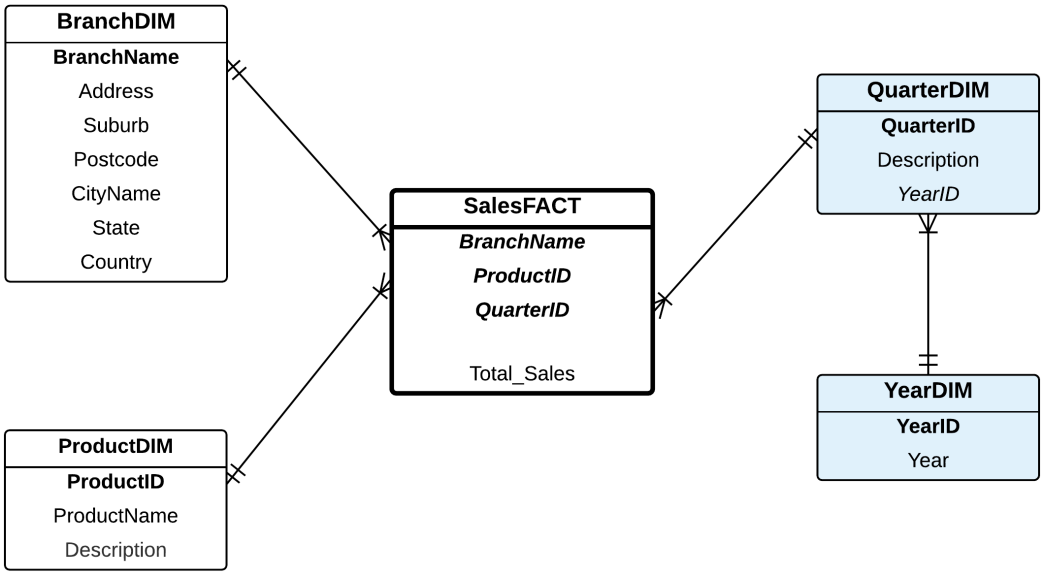


These three levels can actually be joined into a multi-fact star schema through the hierarchy, as shown in the following:
The hierarchy is maintained among the dimensions from different level of granularity in the data warehousing architecture (e.g. Branch, City, and Country dimensions), and the other dimensions (e.g. Product and Time) are shared by the three fact tables.

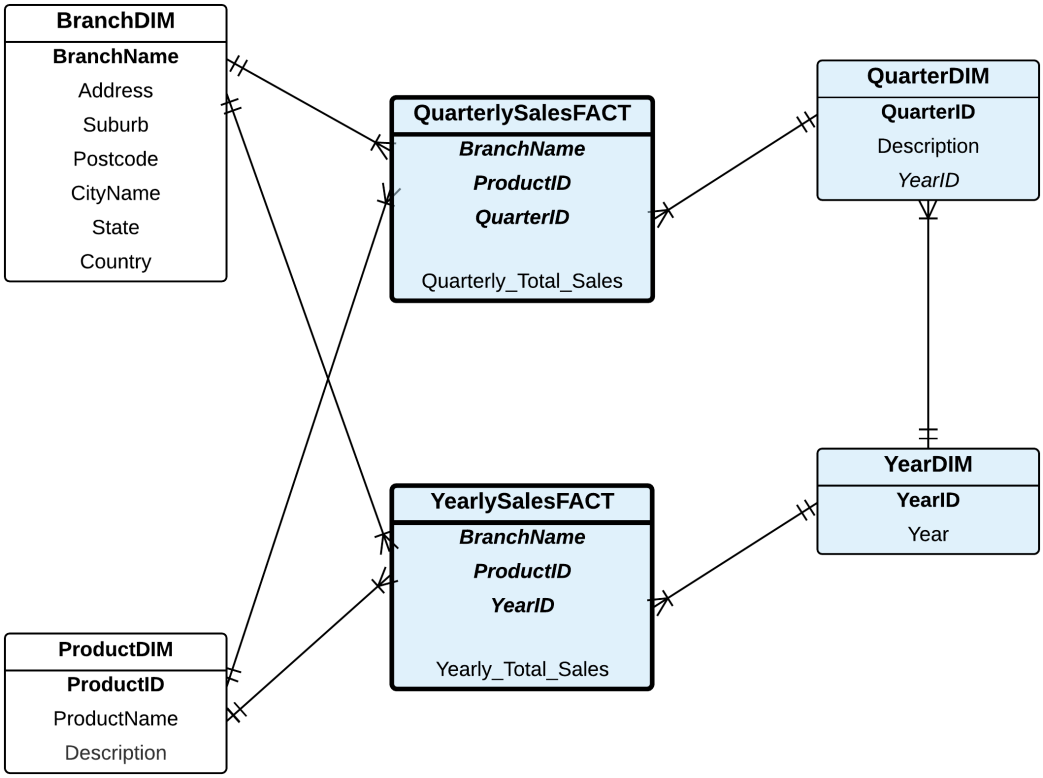


So in short, there is a correlation between the concept of data warehousing architecture (or level of granularity), dimension hierarchy, and multi-fact. The different levels of star schema actually depict the multi-fact, and the dimensions in different levels of star schema can form a hierarchy. As the matter of fact, hierarchy is often used in conjunction with level of granularity in the data warehousing architecture.

It will become more complex when multiple hierarchies exist in a star schema. Supposed, Time dimension has a hierarchy. In the following star schema, the Time dimension is broken down into Quarter and Year dimensions in a hierarchy.

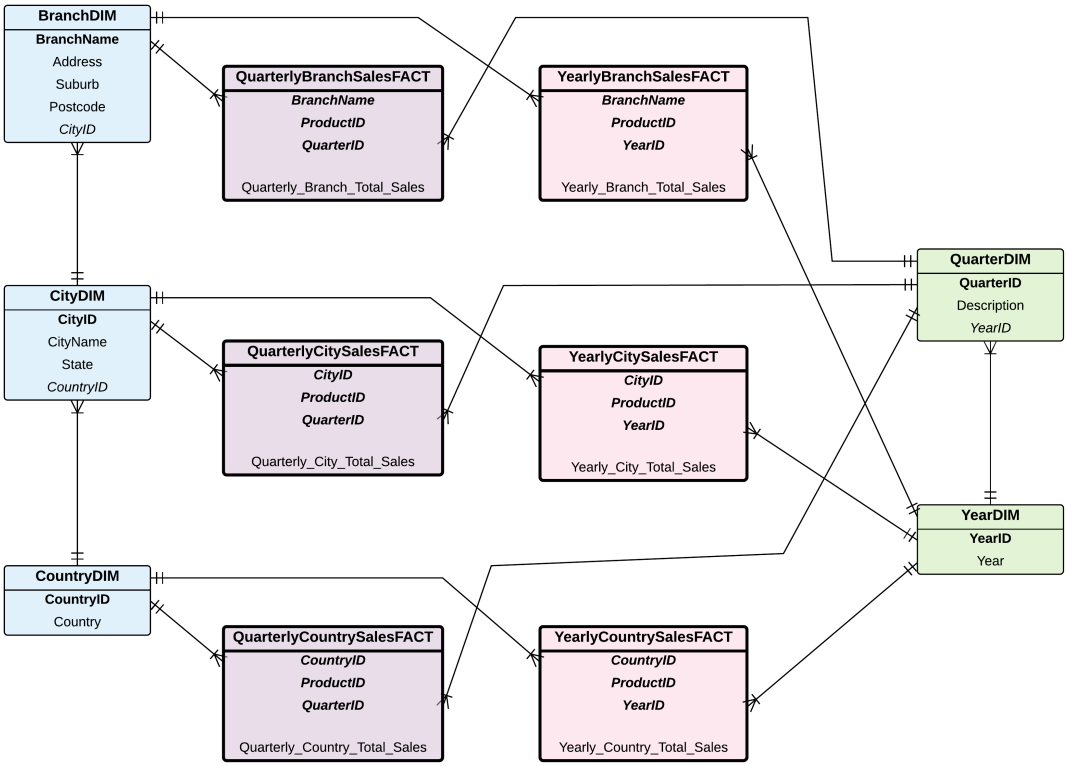


Using the hierarchy, multi-fact, and level of granularity concepts as described above, we can have a star schema consisting of two Fact: one for Quarterly Sales, and the other for Yearly Sales. The hierarchy between Quarter and Year is maintained in the following star schema. So, this is another star schema with Time hierarchy.



If we combine the previous star schema (with the Branch hierarchy) and this star schema (with Time hierarchy), we will end up with 6 Fact tables, as shown in the

following star schema (for simplicity, the Product dimension is not shown in the picture).



Using this combined star schema, drilling down and rolling up the Total Sales can be done by querying the right level of the star schema; as there are six options to choose from Quarterly and Yearly Branch, City, and Country Facts, each with a different granularity of Total Sales.

The numbering system in the data warehousing architecture is not fixed and static. What we know is that Country Sales star schema is more general (has a higher level of aggregation) than City Sales and Branch Sales. The Yearly Sales star schema is more general, and hence has a higher level of aggregation, than the Quarterly Sales star schema. The following six levels of aggregation for the above star schema are **incorrect**:

- Level-6: Yearly Country Sales
- Level-5: Quarterly Country Sales
- Level-4: Yearly City Sales
- Level-3: Quarterly City Sales
- Level-2: Yearly Branch Sales
- Level-1: Quarterly Branch Sales

Take the Quarterly Country Sales and the Yearly City Sales as an example. It is incorrect to say that the Quarterly Country Sales is more general than the Yearly City Sales, because they are not formed by one hierarchy. This is the reason why the above level numbering is incorrect.

When comparing two star schemas, we only need to know whether one star schema is more general than the other, or whether the two star schemas cannot be compared. In the above example, Quarterly Country Sales and Yearly City Sales are not comparable, whereas Yearly Country Sales is definitely more general than Quarterly Country Sales. Therefore we should have two partitions or groups of the level of granularity: one based on time, and the other based on location.