# About Star Schema Data Modeling

A star (or snowflake) schema refers to a way of organizing data in a relational database in order to support OLAP queries. AtScale's virtual cubes rely on star schema data modeling concepts.

About Star Schemas In Relational Databases

In a relational online analytical processing (ROLAP) database, the physical datasets are organized into tables using a star (or snowflake) data model. In a star schema, the tables are organized around a central fact table that represents real world events that have occurred, such as a sales transaction or a web site visit. The fact dataset contains the event records that can be measured, such as a customer's purchases in an online store or an click event logged by the web server.

A fact dataset is then linked to other dimension datasets to provide more context about the fact event. For example, a customer dimension can help you understand which people are the most valuable to your business. A date dimension can help you break down activities by day or week or month.

In a relational database star schema, the dimensional data is usually normalized into separate tables when the data is loaded into the database. A snowflake schema is the same as a star, except that the dimension data is normalized into more granular levels of tables.
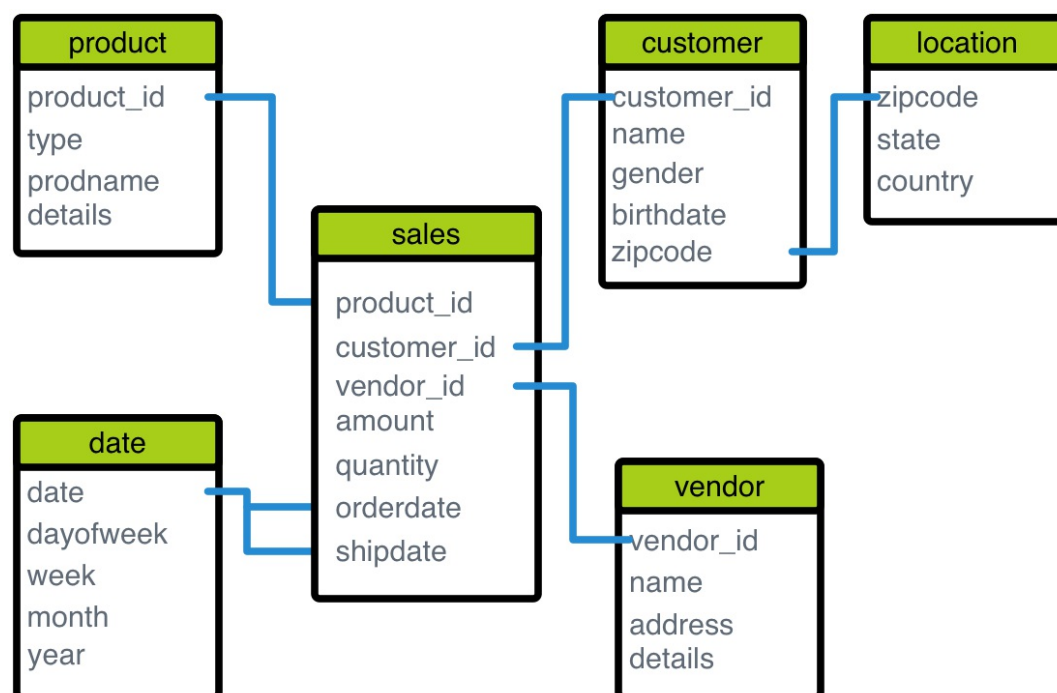
How The Star Schema Relates To AtScale Cubes

When you are designing a cube in AtScale, you always begin with your fact data. An AtScale cube is always based on a central fact dataset (a table in the connected data warehouse containing the cube's event records). This is the dataset on which you create the cube's measures.

The fact dataset is then linked to logical dimensions. Unlike a relational database star schema, dimensions in an AtScale cube do not have to represent physical datasets (although sometimes they do). You can model dimensions on top of a dataset, on columns from the fact dataset, or on columns coming from multiple physical datasets. This virtual star modeling allows you to overlay a star schema on the tables you already have, rather than having to physically normalize the data ahead of time.