* Select the business process to model: a business process is a natural business activity, supported by a data-collection system.
* Declare the grain of the business process: describe *exactly* what an individual fact table row specifies
* Choose dimensions that apply to each fact table row
* Identify the numeric facts that will populate each fact table row

SKUs( Stock-Keeping unit) :  Stock Keeping Units, are product codes that you (and others) can use to search and identify stock on hand from lists, invoices, or order forms.

Data Collection: when a customer checks out at a cash register, the bar codes are scanned directly to the point-of-sale (POS) system of the grocery store

Retail Sales Model- 4 step process:

1. Business process: customer purchases as modeled by the POS system
2. grain: an individual line item on a POS transaction
3. dimensions: store, product, date, and transaction
4. facts: sales quantity, sales dollar amount, cost dollar amount, profit dollar amount

Dimension Tables: if the dimension table also exists in the operational database and has a primary key, the primary key value is stored, but it is *not* used as the surrogate key in the date warehouse

Store Dimension: a geographic dimension a store can be thought of as a location ,stores can be “rolled up” to any geographic attribute, such as zip, state, or city ,stores can also be rolled up to store districts and regions

A **Transaction Dimension** is a **dimension** table containing the textual details of a business **transaction** that is left over in the source **transaction** table, after all the other attributes have been put to other **dimensions**.Columns on these 2 tables are put into 4 main **dimensions**: customer, product, store and date

  A **fact table** consists of the measurements, metrics or **facts** of a business process. It is located at the center of a star schema or a snowflake schema surrounded by dimension **tables**. The primary key of a **fact table** is usually a composite key that is made up of all of its foreign keys.

**Additive facts** are **facts** that can be summed up through all of the dimensions in the **fact** table

A semi-additive fact can be added across some, but not all, dimensions ,bank balances are semi-additive

* Calculated facts: in a data warehouse, it is common to store calculated facts – the storage cost is minor and storing it removes the possibility of user error in making the calculation

Things to avoid :

* + rule of thumb is to have less than 15 dimensions
  + 25 or more dimensions is almost always wrong

Surrogate Keys:

* + surrogate keys should be meaningless
  + do not use “smart” keys where you can tell something about the contents of the row simply by looking at the key
  + surrogate keys should be used for the date dimension, but unlike other surrogate keys, the date dimension keys should be assigned in a meaningful, sequential order