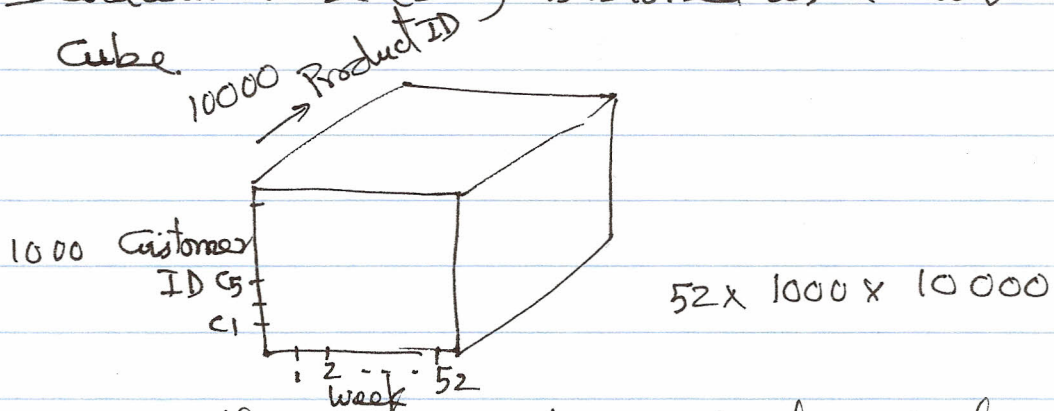


# Data Warehouse

It is a collection of time-variant, non-volatile, subject-oriented data of an organization.

Datawarehouse (DW) is stored as a multidimensional cube.



cell (3, 150, 35)  $\Rightarrow$  int value (sale quantity)  
Week      CustID      ProductID

Since many cells ~~are~~ would be empty,  
Sparse matrix is used for data representation.

Purpose of DW analysis is to predict the behavioral patterns (of customers) from the past data set.

Association      temporal (time series)

## Operations on a DW

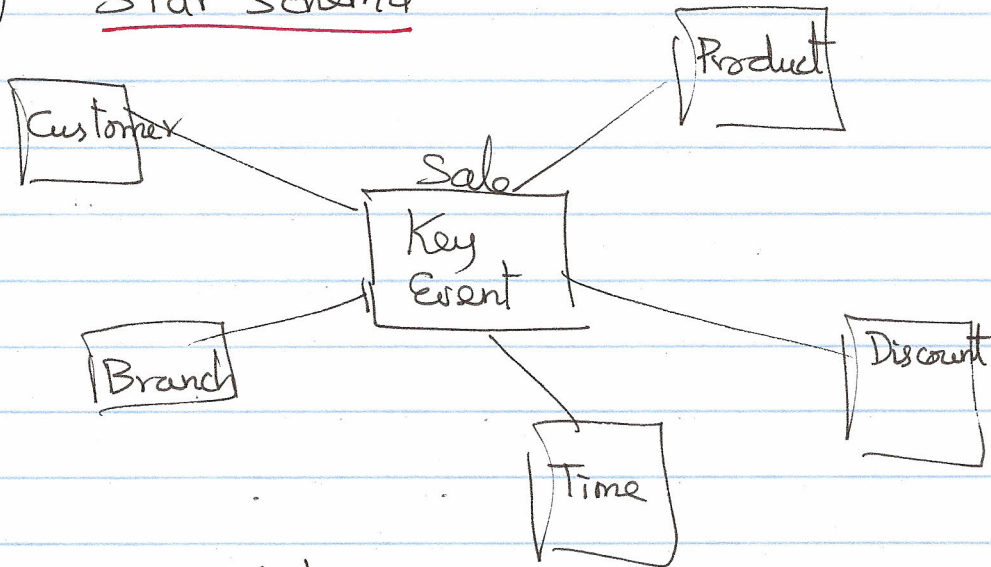
Slicing } reduces the dimensions  
Dicing } 4D  $\rightarrow$  3D, 3D  $\rightarrow$  2D, 2D  $\rightarrow$  1D ...

Drill down      start with a n-D to a specific cell

Roll up      start with a cell to n-D cube

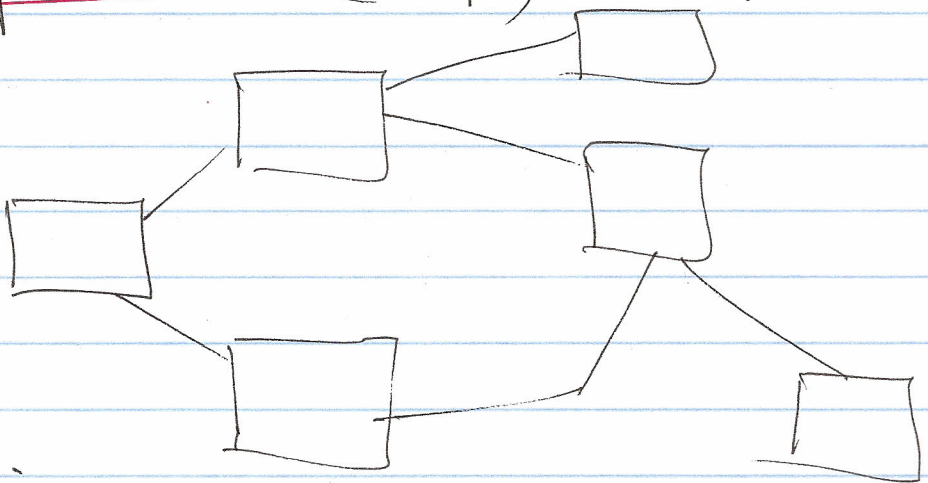
# Data warehouse Model

## a) Star Schema



I-hop

## b) Snowflake Schema (Graph) multi-hop



## Processing

### OLAP

Online Analytical Processing  
Complex operations  
Requires substantial processing

### OLTP

Online Transaction Processing  
Smaller transactions  
Requires quick response  
(short response time)