

Day 0

Just syllabus

Day 1

S subject oriented - particular segment of a business / company

I integrated - multiple data sources integrated

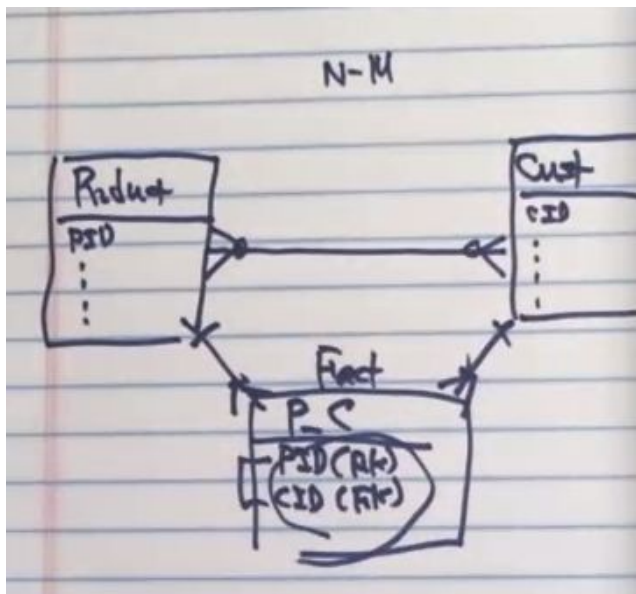
N non-volatile - data is stable; added but not removed

T time-variant - all data is for a particular, identified point in time

Day 2

Why do we create warehouses? More user friendly and efficient / fast

How to implement any to many - put a composite (fact) table in between using an associative identity



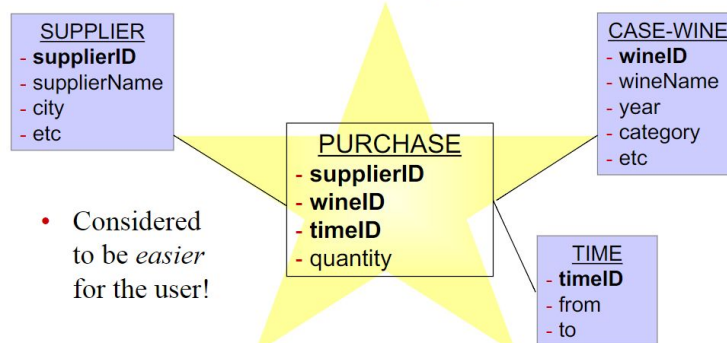
Look for any to many relationships as they will be a candidate for a fact table

Any data map MUST have time or date mapped (the T in SINT - time dimension)

Star schema is then created: fact table in middle with dimensional tables on the ends with one to many relationships

Dimensional Modeling

- Star schema for the SUPPLIER_WINE relationship:



Day 3

Why do we need data warehousing?

- Accessible

- Roll-up / summary / details which can then be sliced, diced to specific data (ex. Getting data just from the state of NY)

- Easy, user friendly, fast, efficient

- Can show just what is important

Possible problems wh

- Duplicate data

- Incorrect or missing data

- Data with differing names (ex. 'Student Id' vs 'Id Number')

Operational systems = computing systems that provide the information necessary to run day to day business activities

Decision support systems = computing systems that provide vital strategic information for effective decision making

Good strategic info:

- Integrated: a single, enterprise-wide viewpoint

- Data integrity: data is correct and accurately represents business rules

- Credible: single source values

- Accessible: ease of access, flexible for intuitive and investigational analysis

- Timely: up to date info is available when needed

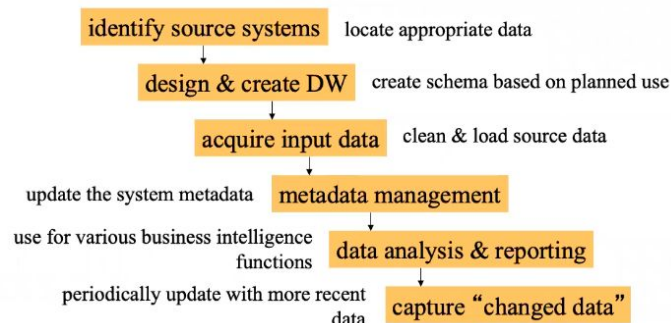
Data warehousing definition - the activities needed to create, maintain, and utilize a data warehouse or data mart

- Creating

- Populating

- Querying (user accessing)

Basic DW'ing Life Cycle



<- Handling dimensional change is important!

Ex: someone buy tv in NY state, person later moves to CA, so we need to handle that they lived in NY when it was bought and not CA - "Slowly changing dimension(al table)"

General Guidelines:

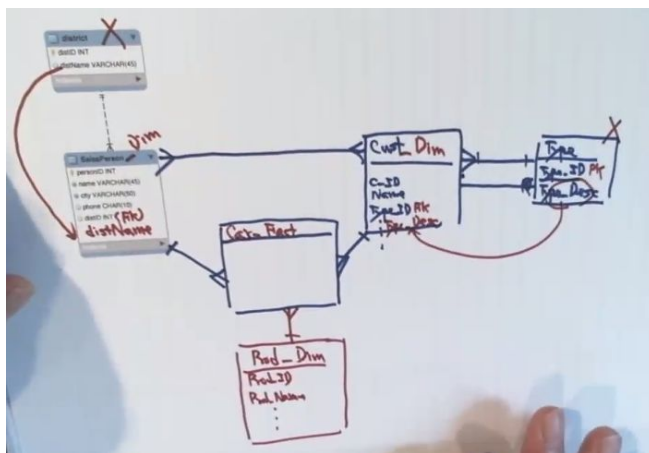
- Build from a clear set of business objectives and user requirements
- Define the architectural framework in advance
- Document all assumptions
- Use the right tools for the job
- Understand the life cycle
- Expect data problems
- Learn from mistakes

Any data map MUST have a time table (remember SINT)

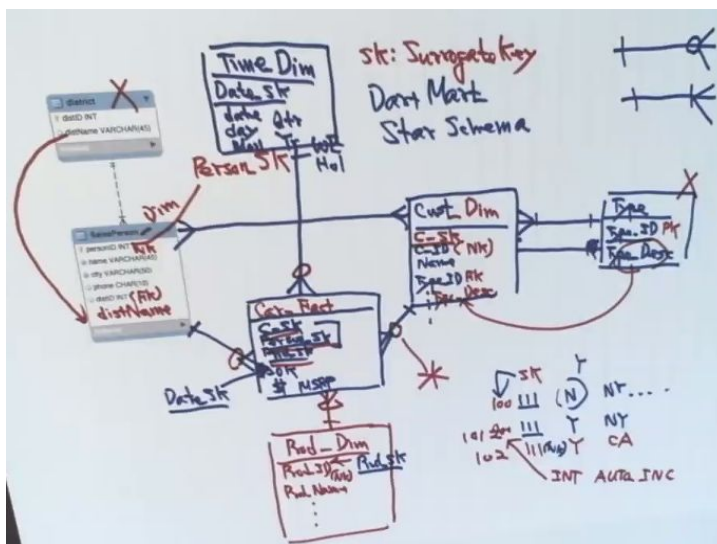
Day 4

Denormalizing a many to many relationship with an associative identity to a Data Mart (star schema):

1. Bring district name into sales person to remove the District table
2. Bring type description into customer to remove the Type table
3. Customer, sales person, and rod are all connected and represented by individual dimensional tables, connected by the "car" fact table



4. Surrogate key represented by '_SK' (which are also the primary keys), all added to the car fact table, auto increments when information changes
5. We NEED a time/date table now (always!)
6. Add the circles to create ANY to one instead of many to one
7. For the original primary keys, we represent them as '_NK', natural keys now



Tips for exercise 1:

Put Type_Desc into customer and get rid of Customer_Type table; this creates 'Customer_DIM' table

'Customer_ID' must now be natural key -> add a surrogate key 'Customer_SK'

Move Cat_Name into Product and get rid of Category table

Customer_ID and Product_ID in Sales table becomes _SK

Create Date dimensional table using the Date from the Sales table, with 'Date_SK' as primary/surrogate key

Change 'many to one' to 'any to one' with open circle on the many side

Bring a hard copy to class!

Major components of DW architecture:

- **Data Acquisition**
 - Extracts data from legacy systems & external sources, consolidates & summarizes the data, and loads it into the Data Storage
- **Data Storage**
 - Contains the integrated data, metadata, and associated software
- **Information Delivery**
 - Allows users to access and analyze data in the warehouse

Data Staging

- An area used to receive data from operational sources and prepare it to be placed into a data warehouse
- **Extracting (E)**: read and understand the source data, and copy the parts that are needed to the data staging area for further work
- **Transforming (T)**: Once the data is extracted, there are many possible transformation steps, including cleaning, purging, combining, creating **surrogate keys**, & building aggregates
- **Loading (L)**: Initial load moves very large volumes of data, & the business conditions determine the refresh cycles

Day 5

In person class going over Exercise 1