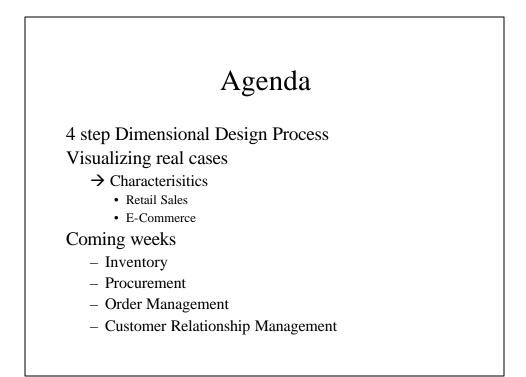
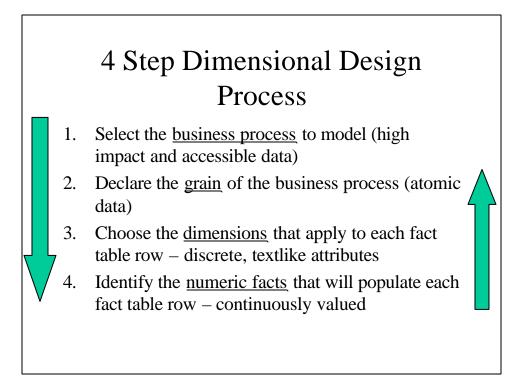
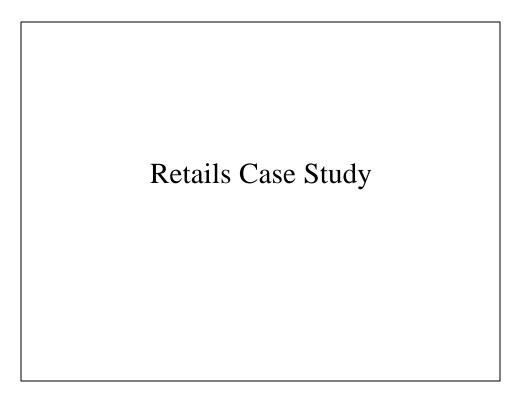
IS660J

Lecture 3 Professor K.M. Burns

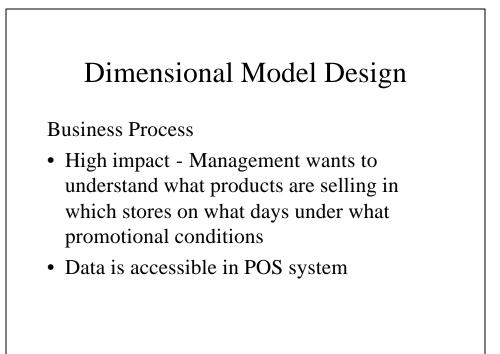






Background

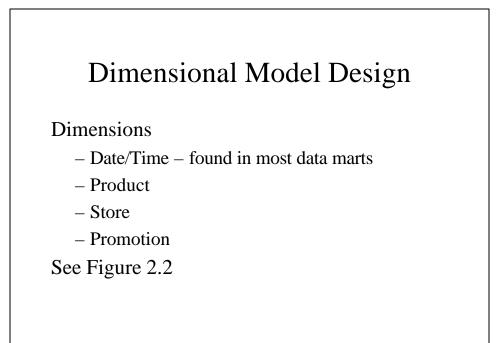
- UPCs & SKUs
- Data Collection pts POS, Vendor deliveries
- Logistics of ordering, stocking, and selling products while maximizing profit (sales price less costs)
- Pricing, Volumes and Promotion



Dimensional Model Design

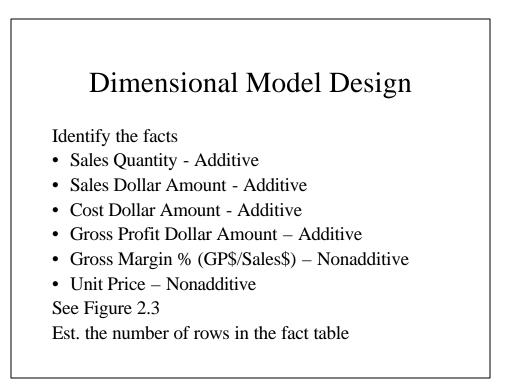
Declare the grain

- Individual line item (implies store and product) on a POS transaction (implies daily)
 - Broadest users
 - Slicing and dicing



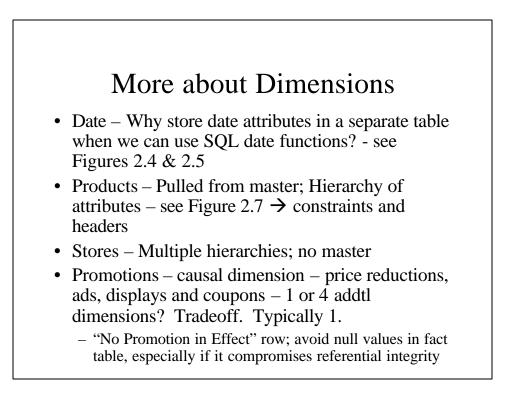
Conformed Dimensions

- EDW is comprised of many data marts
- In order for DMs to use the same dimensions (ie. product) in joins with fact tables, the dimension must be the same – consistent keys, consistent attribute names, consistent attribute definitions, and consistent attribute definitions.



Primer on SQL queries Assignment 3 example Constraints

SELECT time.Date, products.Brand, products.Description, stores.Region, stores.Store_Location_Desc, sales_facts.Actual_sales_price, sales_facts.Quantity_sold, sales_facts.Sales_revenue FROM time, products, sales_facts, stores WHERE sales_facts.Product_key=products.Product_key AND products.Brand LIKE '%Chef Ronaldo%' AND sales_facts.Time_key=time.Time_key AND (sales_facts.Time_key BETWEEN '0' AND '9999') AND sales_facts.Store_key=stores.Store_Key AND stores.Region LIKE '%%' ORDER BY stores.Region, products.Brand



Factless Fact Tables

What products were on promotion but not sell?

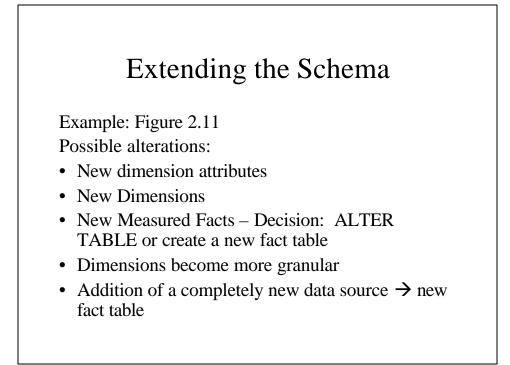
- Don't want to overpopulate our fact table
- Promotion Coverage fact table 2nd table to track what didn't happen – same key, different grain (day or week)
- Factless because there are no measurements; it merely captures the relationships
- Chapter 12



- POS transaction number → Grouping key (DD) → looks like a dimension key but doesn't carry the attribute info of a dimension (empty); hence, degenerate
- Typically part of the primary key of the fact table
- See Figure 2.10
- Other examples: Order numbers, invoice numbers, bill of lading numbers

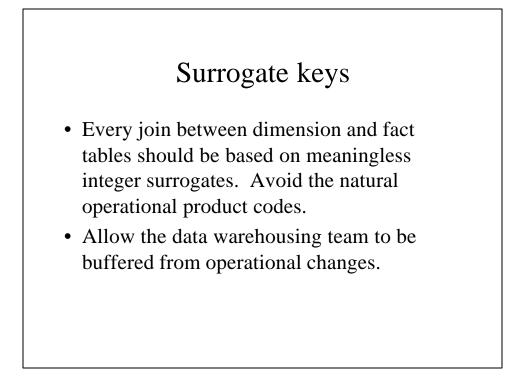
Reporting

- What are the weekly sales \$ volume by promotion for the snacks category during Jan 2002 for stores in the Boston district?
- See Figure 2.10
- High quality dimension attributes are crucial because they are the source of query constraints and result set labels.
- Cross tabular reports are better.



Resist Temptations to ...

- Snowflake Dimensions save space and/or maintenance time at the expense of <u>ease of use</u> and <u>performance</u> – see Figure 2.12
 - Dimension tables are relatively small compared to fact tables.
- Denormalize the fact table –ie. including fields from the Date dimension in Fact tables.
 - 25 or more dimensions may require us to find ways to combine correlated dimensions into a single dimension. Typical fact tables have 15 or less.



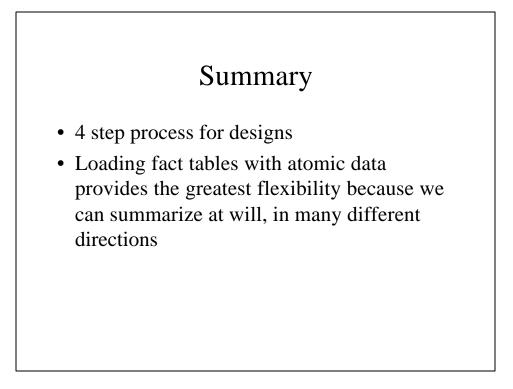
Market Basket Analysis

Which products were sold in the same market basket together? Combinations – Market Basket Analysis, Affinity Grouping

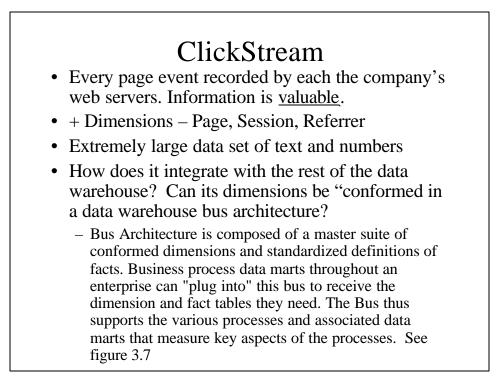
Best done by Data Mining tools and/or OLAP. Why? Constraints and groupings need to be done in the fact table.

Can be done by creating an new fact table – see figure 2.14

Semiadditive measures – don't double count!!!



E-Commerce



Clickstream

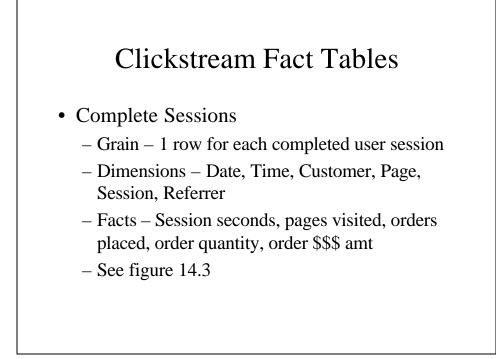
- Many log files, many websites (ISPs, Ad Referrers, Search Engines, Web watcher services, ect.) are involved in 1 session for a user → How do we integrate these sources? How do we sync them in time?
- XML has the potential to make web pages far more expressive

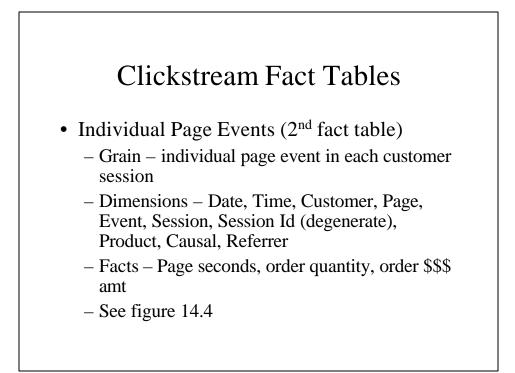
Challenges

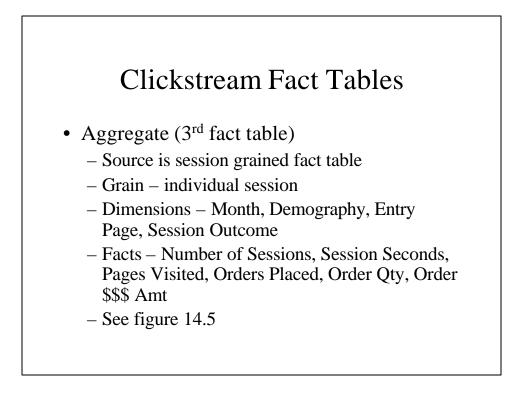
- Identifying the Visitor Origin
- Identifying the Session
- Identifying the Visitor
- Proxy Servers
- Browser caches

+ Dimensions

- Page
- Event
- Session
- Referral
- Existing dimensions Date, Time, ect.







Integrating into the EDW

• See figure 14.6

Web Profitability Data Mart

- Allecate activity costs and infrastructure costs down to each sales transaction
- View of profitability over all our sales channels, not just the web
- New data mart
 - Grain individual product sold on a sales ticket to a customer at a point in time (like Retail example)
 - See figure 14.7

Conclusion

Clickstream

- Identifying the Visitor Origin
- Identifying the Session
- Identifying the Visitor
- Proxy Servers
- Design and Integration