



More Warehouse Design Issues

- **Data transformation** and **data cleansing**
 - E.g. correct mistakes in addresses (misspellings, zip code errors)
 - **Merge** address lists from different sources and **purge** duplicates
- *How to propagate updates*
 - Warehouse schema may be a (materialized) view of schema from data sources
 - View maintenance
- *What data to summarize*
 - Raw data may be too large to store on-line
 - Aggregate values (totals/subtotals) often suffice
 - Queries on raw data can often be transformed by query optimizer to use aggregate values