NATIONAL ENGINEERING CENTER
University of the Philippines
Diliman, Quezon City

4.0 Designing Fact Tables

Eugene Rex L. Jalao, Ph.D.

Associate Professor

Department Industrial Engineering and Operations Research

University of the Philippines Diliman

Module 2 of the Business Intelligence and Analytics Track of UP NEC and the UP Center of Business Intelligence

Outline for This Training

1. Introduction to Data Warehousing
2. DW Lifecycle and Project Management
   - Case Study on DW PM
3. Dimensional Modeling
4. **Designing Fact Tables**
5. Designing Dimension Tables
   - Case Study on Dimension Modeling
6. Extraction Transformation and Loading
   - Case Study on ETL Planning
7. Transformation and Loading Methodologies
   - Case Study on ETL

Outline for this Session

- Fact Table Facts
- Granularity
- Fact Table Examples
- Fact Table Contents
- Fact Table Measures and Additivity

Fact Table Facts

- A fact is a performance measure
  - Sales of Product X
- Fact value not known in advance; only when an event measurement occurs
  - Actual Sales
- Facts are numeric
  - In PhP
- The most useful facts are numeric and additive
  - At least interval type of attributes

Fact Table Traits

- Are usually the largest tables
- Are usually appended to
- Can grow quickly
- A single fact table can contain either detail or summarized data
- Their measures are typically though not necessarily additive
- Are primarily joined to dimension tables through foreign keys

Outline for this Session

- Fact Table Facts
- **Granularity**
- Fact Table Examples
- Fact Table Contents
- Fact Table Measures and Additivity

Fact Table Granularity

- The fact table's grain is the business definition of the measurement event that produces the fact table
  - Example: Each time a customer submits an order online a customer order event ultimately becomes a row in the customer order fact table.
- Declaring the grain means a fact table row represents the blank in this statement: "A fact row is created when _____ occurs."

Determining the Grain of a Fact Table

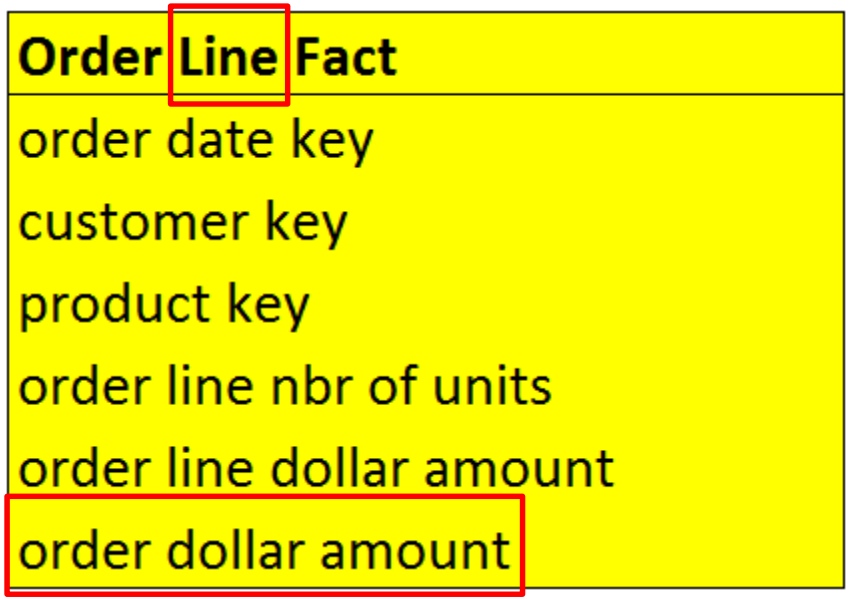
- In business terms
  - What is the meaning of an individual row in the fact table
- In data modeling terms
  - What is the unique logical identifier
  - What are the identifying dimension keys
- In ETL terms
  - What is the rule for populating the table

Grain of a Fact Table Example

- Granularity statement
  - "One row for each product sold by store by day"

Daily Product Sales by Store Summary Fact
day key - lpk1
store key - lpk2
product key - lpk3
quantity
dollar amount

All Measures Must be of the Same Grain



Outline for this Session

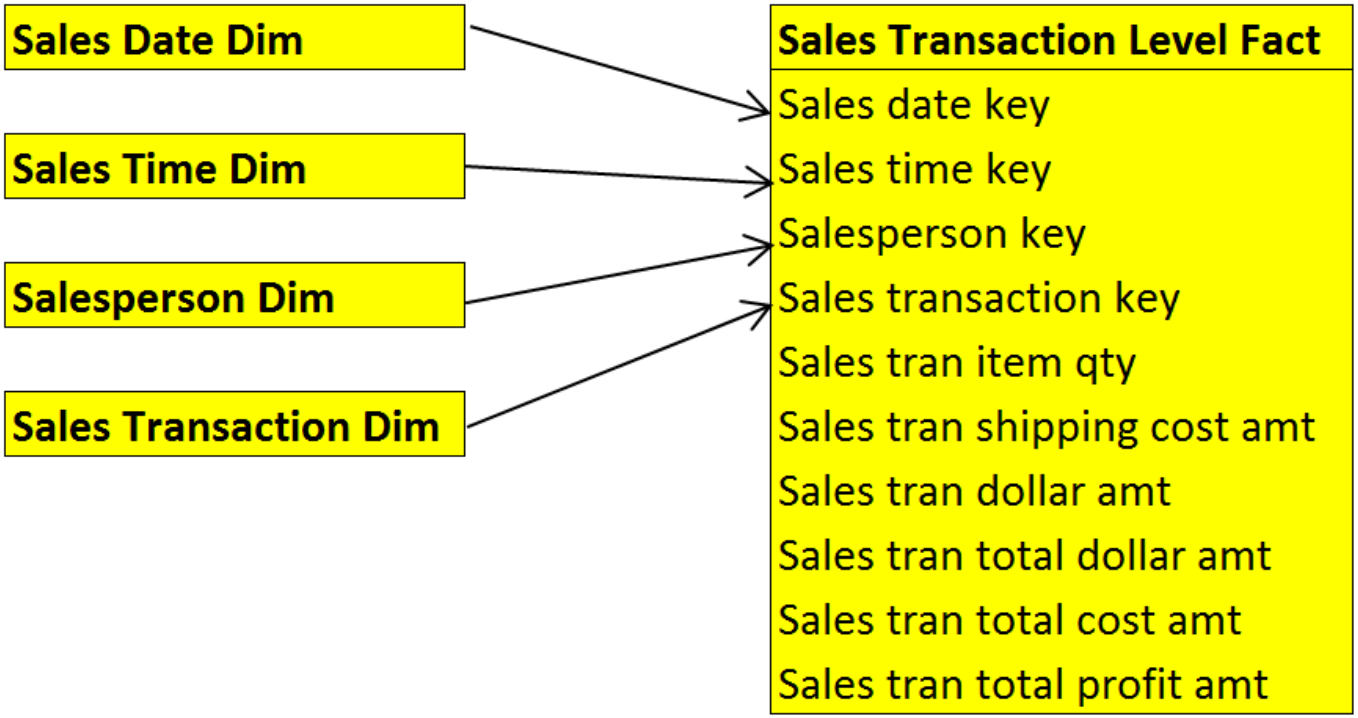
- Fact Table Facts
- Granularity
- **Fact Table Examples**
- Fact Table Contents
- Fact Table Measures and Additivity

Fact Table Examples

- Detail
- Analytical
- Summaries

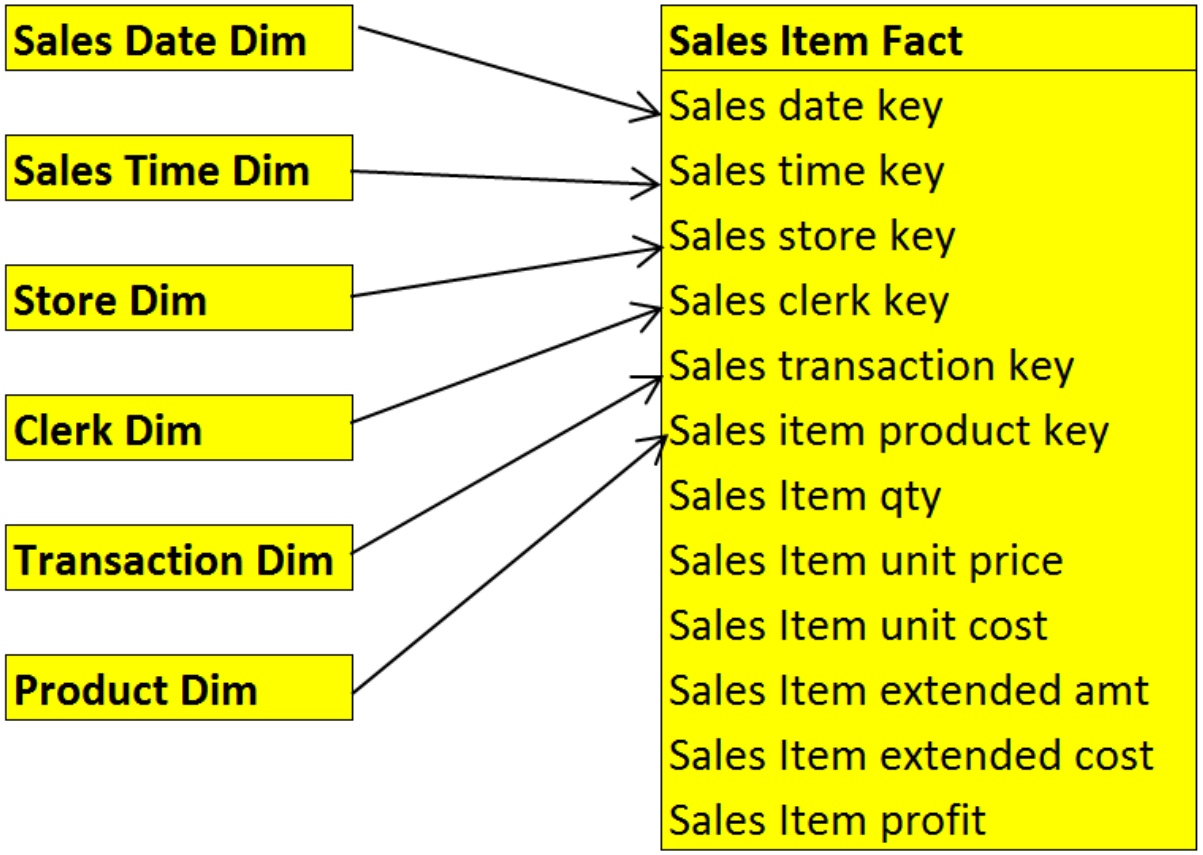
Transaction Level Fact Table

- One row per sales order.



Transaction Item Fact Table

- One row per sales order item.



Detail Fact Table – Granularity Statement

- Granularity statement
  - "One row for each item in a transaction"
- Notice that the standard dimensions are not part of the granularity statement

Sales Item Fact
day key
transaction key - lpk1
item nbr - lpk2
store key
product key
promotion mix key
distribution mgr key
time
quantity
dollar amount

Granularity Enforcement

- ETL Population Rules
  - Transaction Key: generated from the transaction ID as part of the ETL process.
  - Item Number: even if it did not exist in the source transaction it can be generated during the ETL process

Sales Item Fact
day key
transaction key - lpk1
item nbr - lpk2
store key
product key
promotion mix key
distribution mgr key
time
quantity
dollar amount

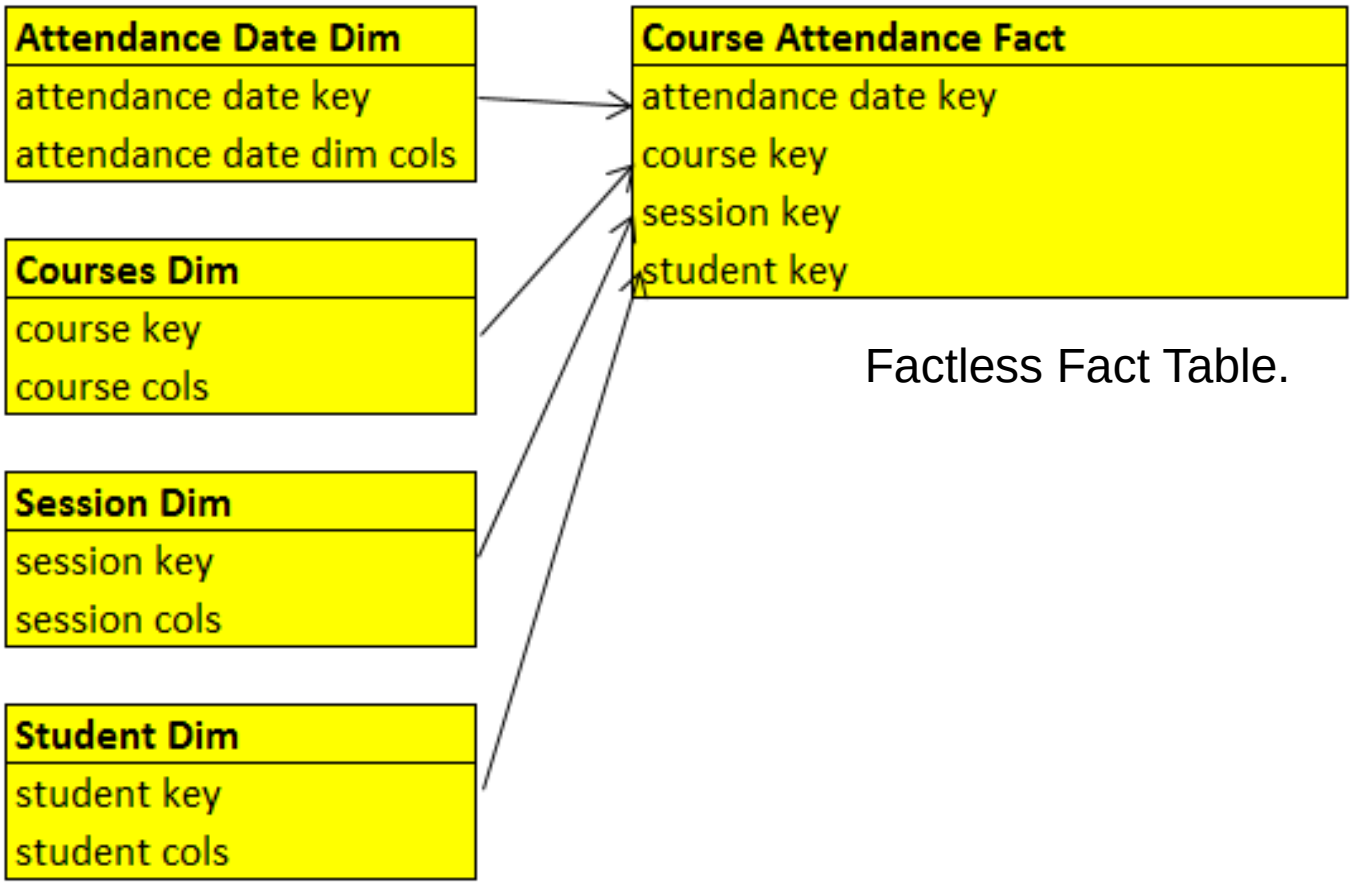
Detail Fact Table - Dimensionality

- What is Dimensionality?
  - Keys that are foreign keys in the Fact Table connected to primary keys of their respective Dimensions
- Sales Item Fact Dimensionality:
  - Day
  - Store
  - Product
  - Promotion Mix
  - Distribution Mgr

Sales Item Fact
day key
transaction key - lpk1
item nbr - lpk2
store key
product key
promotion mix key
distribution mgr key
time
quantity
dollar amount

Event Fact Table

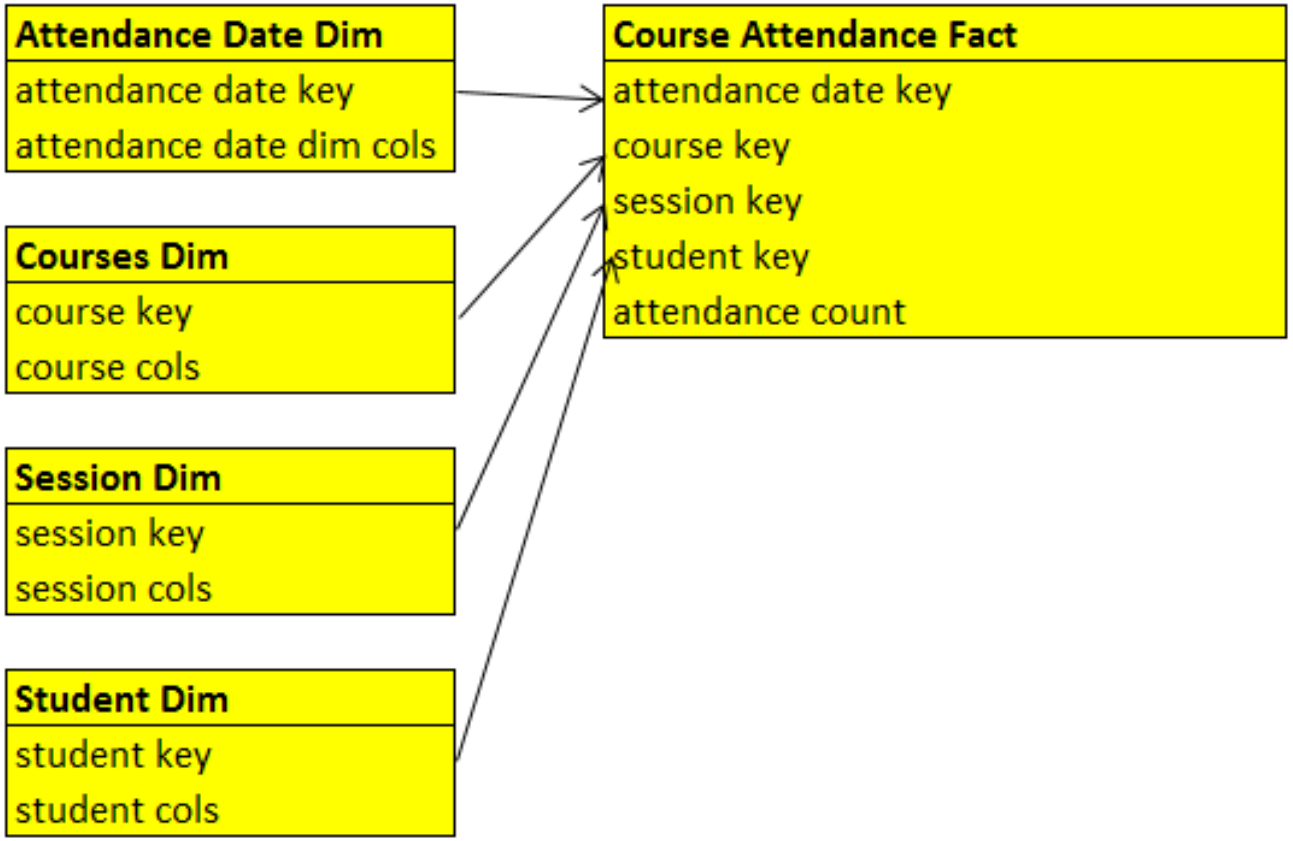
- One row for each student who attends each class



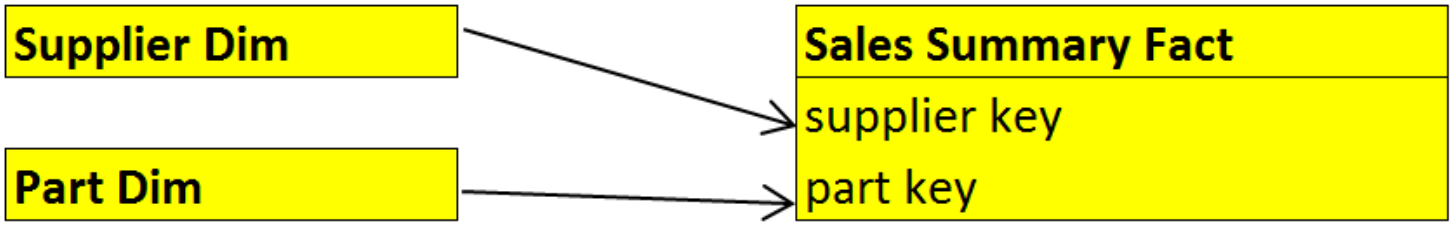
Factless Fact Table.

Enhanced Event Fact Table

- One row for each student who attends each class



Dimension Relationships Fact



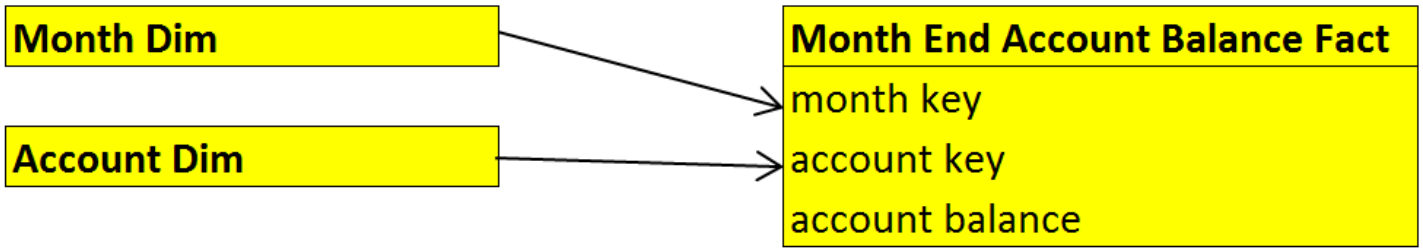
- Resolves many to many relationships between several pairs of dimension tables

Fact Table Examples

- Detail
- Analytical
- Summaries

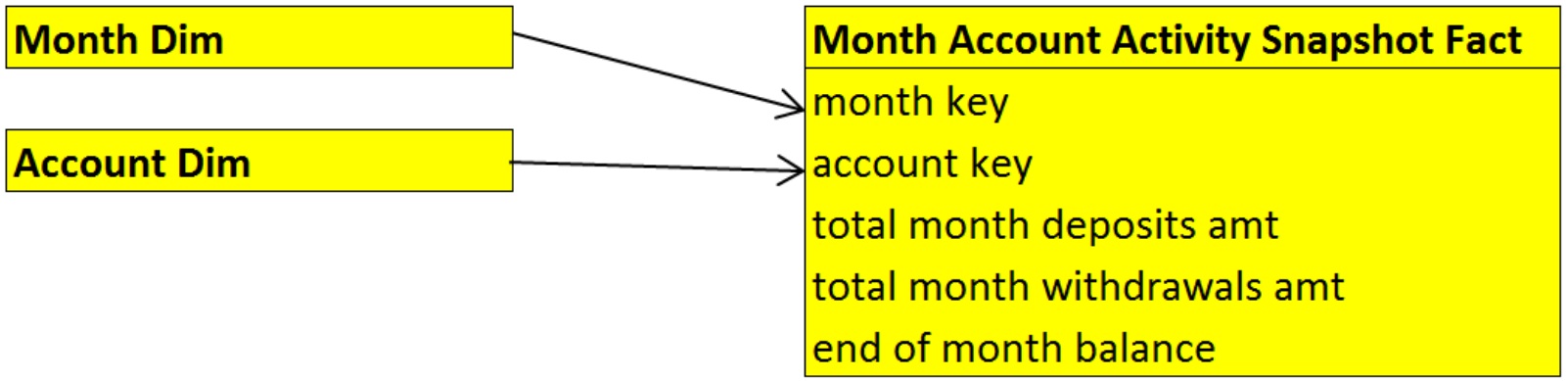
Snapshot Fact Table (Point in time values)

- One row per account per month



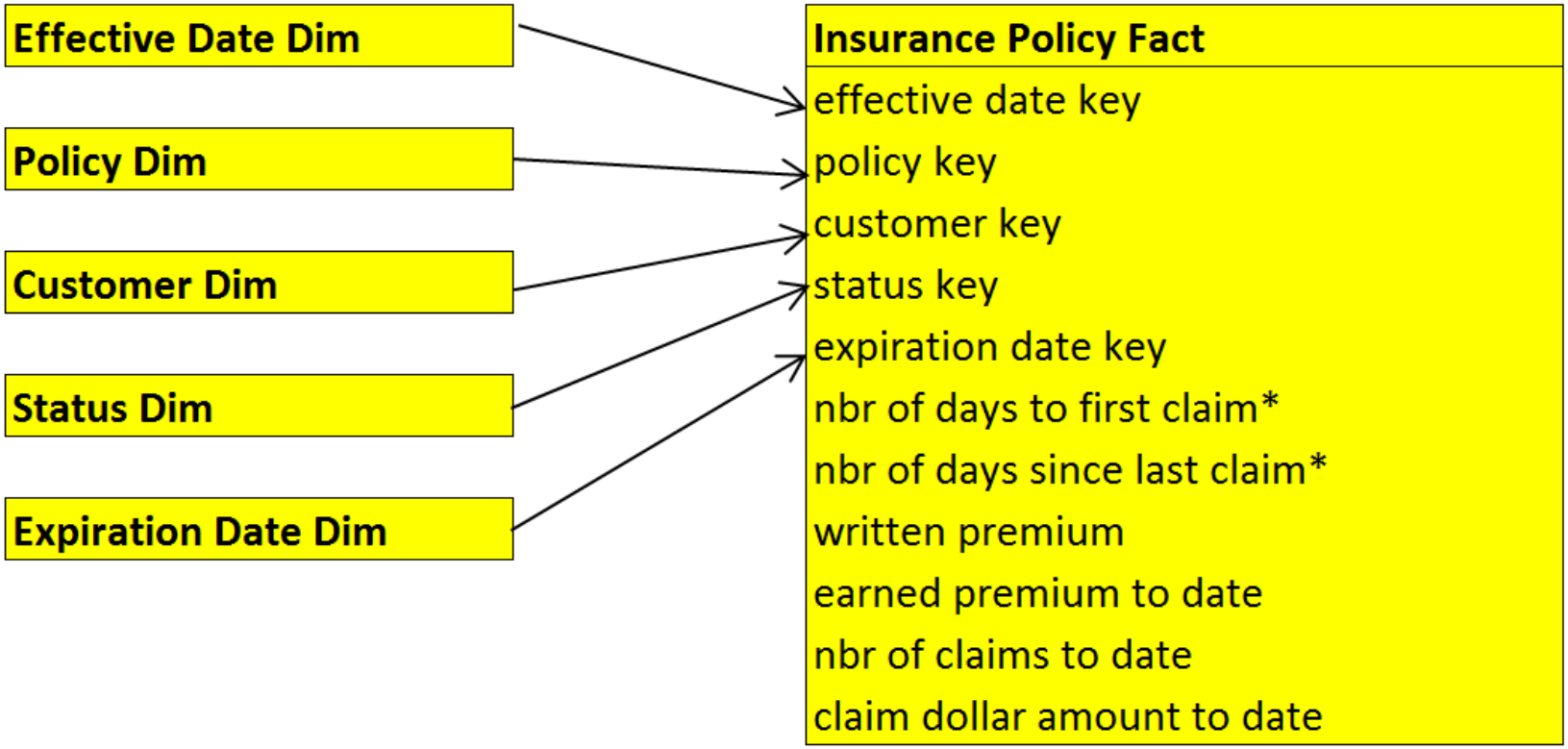
Periodic Snapshot Fact Table

- One row per account per month



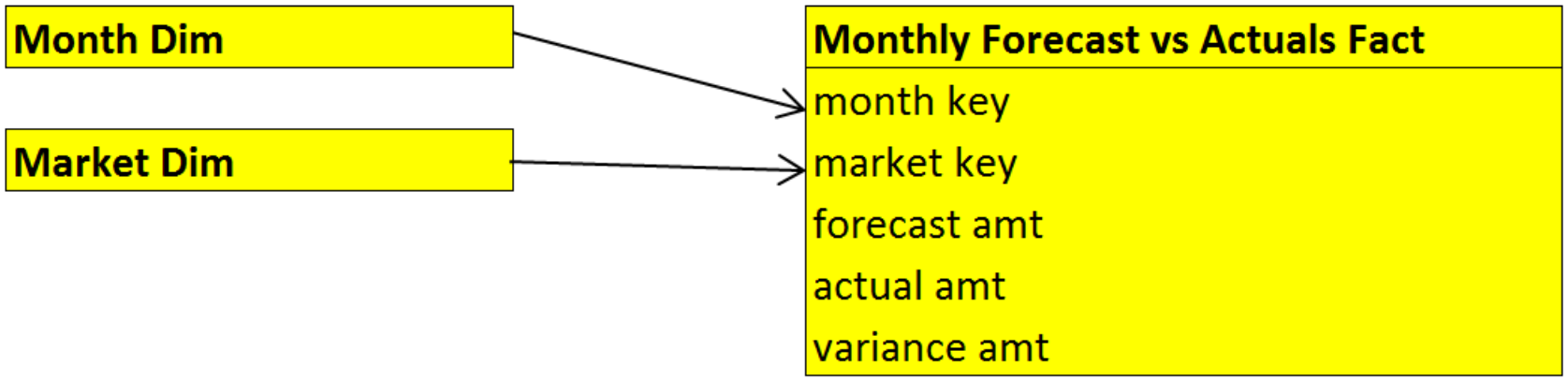
Accumulating Snapshot (Status) Fact Table

- One row per policy



Aggregate Fact Table

- One row per market per month

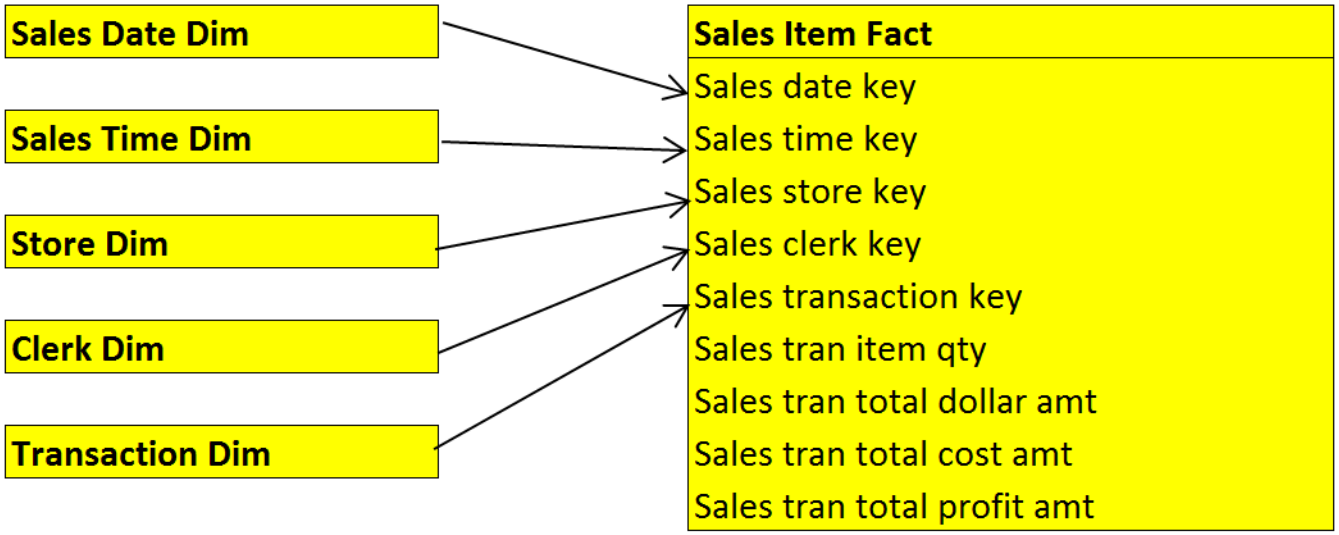


Fact Table Examples

- Detail
- Analytical
- Summaries

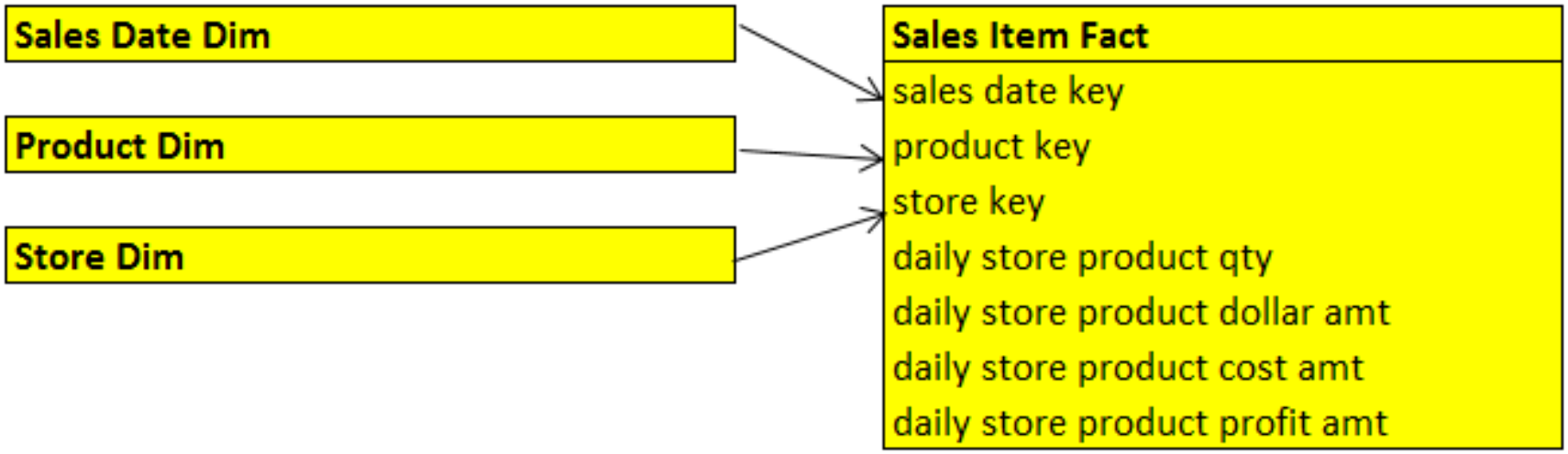
Transaction Summary Fact Table

- One row for each sales transaction



Daily Product Sales by Store Summary Fact Table

- One row per product per day per store



Summary Fact Table – Granularity Statement

- Granularity statement
  - "One row for each product sold by store by day"
- A perfect cube since the lpk is made up of all of the dimension keys

Daily Product Sales by Store Summary Fact
day key - lpk1
store key - lpk2
product key - lpk3
quantity
dollar amount

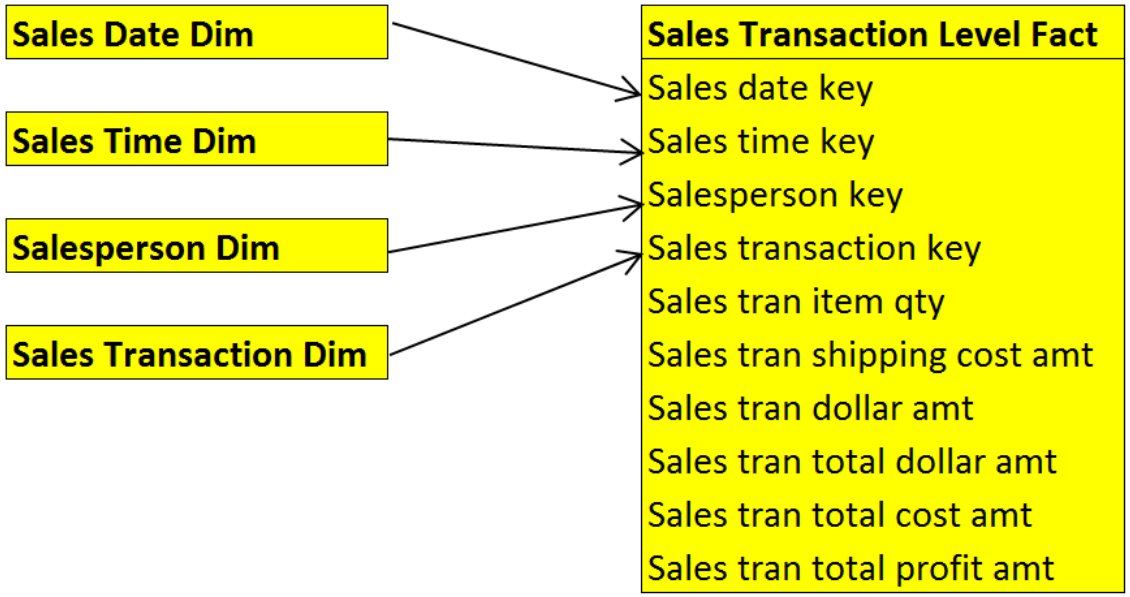
Summary Fact Table - Dimensionality

- Dimensions Match Granularity
- Dimensions identify the grain or granularity of this table.
  - Day
  - Store
  - Product
- Identifying Dimensions

Daily Product Sales by Store Summary Fact
day key - lpk1
store key - lpk2
product key - lpk3
quantity
dollar amount

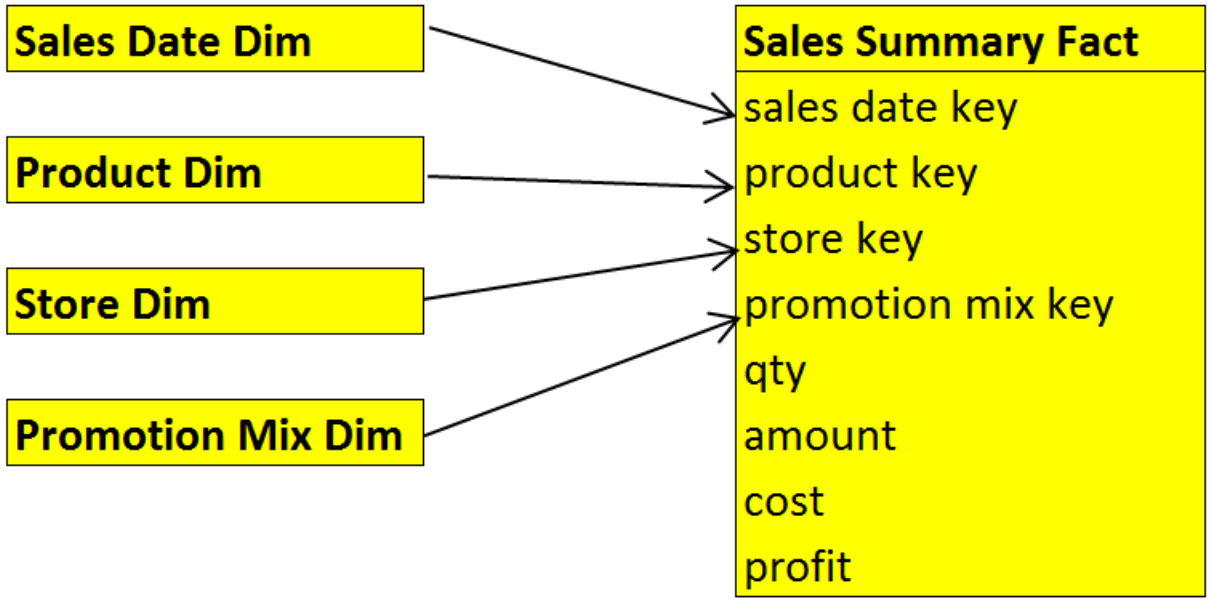
Non-Identifying or Tagging Dimensions

- Tagging or non-identifying dimensions can be added to a fact table without changing the granularity
- Dimensionality does not match the granularity
- Sometimes the grain of a fact table is not made up of all the dimensions in the fact table



Query That Can't Be Satisfied From the Regular Sales Fact

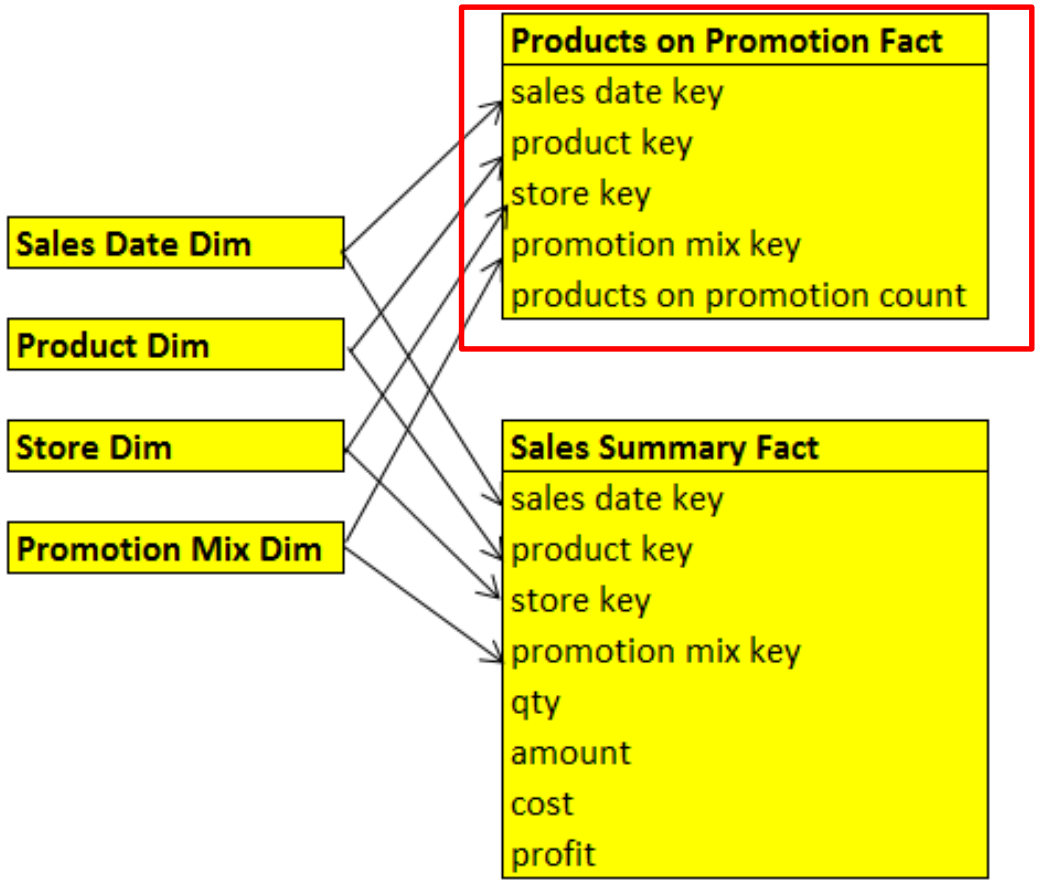
- How many products which were on sale yesterday did not sell?



- Need a new dimension

Another Detail Level Fact Table
Coverage Fact Table

- Need new table for query:
- One row for every product on promotion in every store on every day, regardless of whether or not the product sells in that store on that day

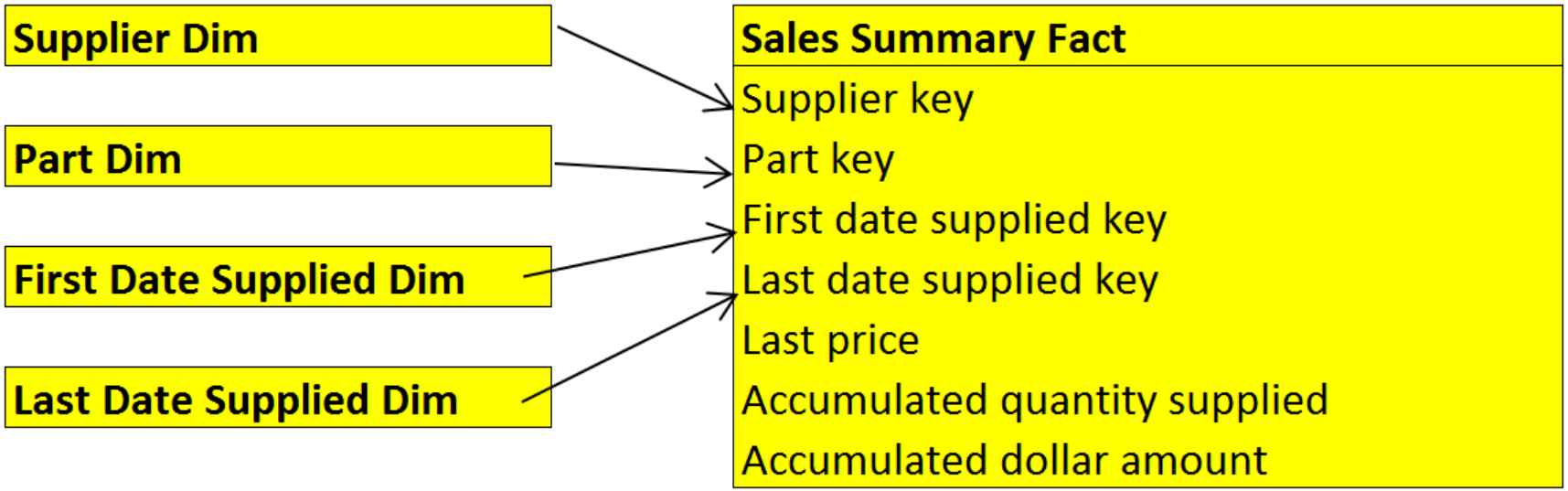


Coverage Query

- How many products which were on sale yesterday did not sell? Combine two queries as follows:

```
SELECT product_name
FROM products_on_promotion cf, store_dim s,
     product_dim p, date_dim d
WHERE sale_date = '6/1/2002' AND d.day key =
     cf.day key AND p.product key =
     cf.product_key AND product_key NOT IN
     (SELECT product_key FROM sales_summary
     _fact WHERE sales_date = '6/1/2002')
```

Extended to Accumulating Snapshot (Status Fact)



Classifications Table

TABLE	TYPE	CLASS
Sales Item Fact	Transaction Item	Detail
Sales Transaction Level Fact	Transaction Level	Hybrid (detail + summary)
Sales Transaction Summary Fact	Transaction Summary	Summary
Daily Product Sales by Store Summary Fact	Summary	Summary
Course Attendance Fact	Event	Detail
	Enhanced Event	Detail
Month End Account Balance Fact	Snapshot	Analytical
Month End Account Activity Summary and Balance Fact	Periodic Snapshot	Hybrid (analytical + summary)
Insurance Policy Fact	Accumulating Snapshot	Analytical
Daily Store Products on Promotion Fact	Coverage – just another detail fact table	Detail
Product Supplier Fact	Dimension Relationships	Detail

Outline for this Session

- Fact Table Facts
- Granularity
- Fact Table Examples
- **Fact Table Contents**
- Fact Table Measures and Additivity

Fact Tables Content

- Only dimension keys and measures

Daily Product Sales by Store Summary Fact
sales date key
store key
product key
distribution mgr key
qty
dollars

- Exceptions:
  - degenerate dimensions
  - line item numbers

Fact Storage

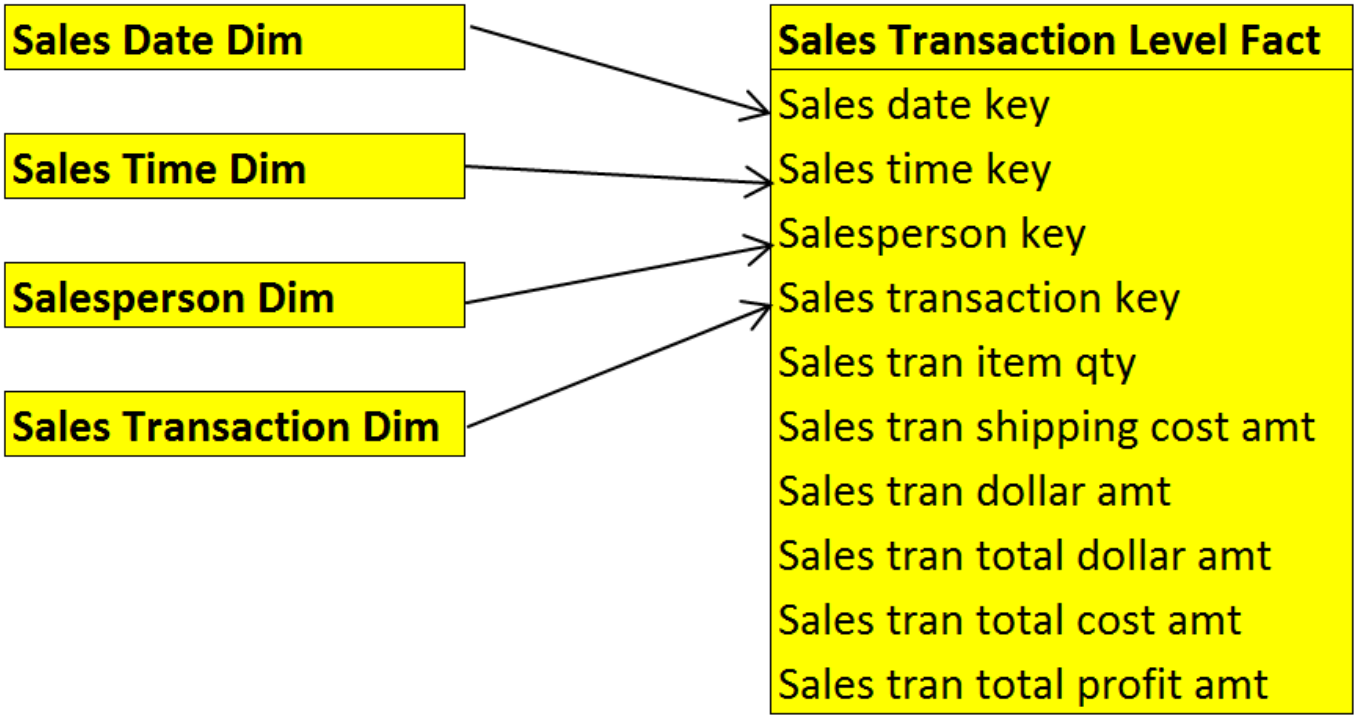
- All attributes are stored as Integers
- Usually stored in a 4 byte sized Integer = 32 Bits
  - 32 Ones and Zeros
- A 32-Bit attribute can handle:
  - Minimum: 0
  - Maximum: 4,294,967,295

Warehouse Translation Example

- Source System POS Record
  - (250,000 rows per day)
  - Transaction id
  - Timestamp
    - SKU – (1-n)
    - SKU (1-n) Quantity
    - SKU (1-n) Price
    - SKU (1-n) Extended Amount
  - Sub-Total
  - Tax
  - Total
- Assuming an average of 15 items per transaction
- Item Level Fact Table (3,750,000 rows per day)
  - Transaction id
  - Sales date key
  - Sales time period key
  - Product key
  - Quantity
  - Unit Price
  - Extended Amount

Transaction Level Fact Table Size

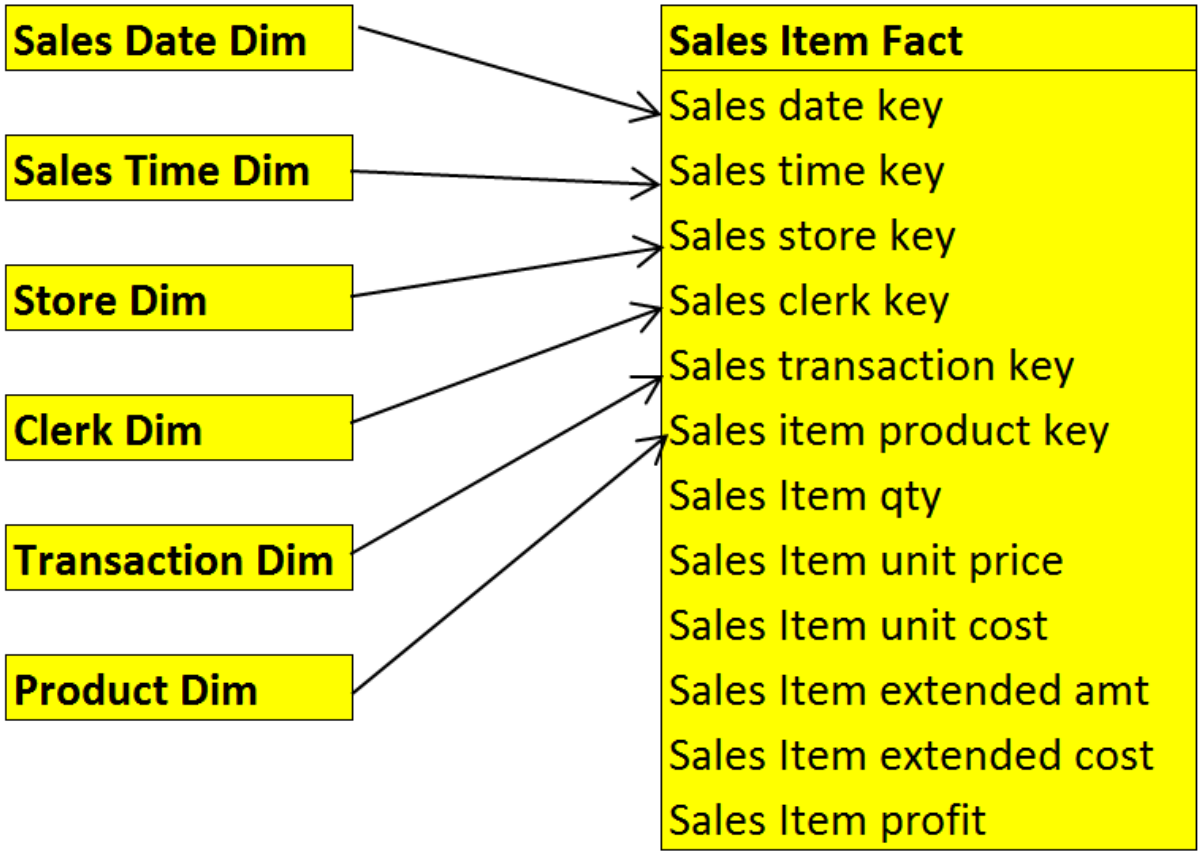
- One row per order.



- 250,000 rows per day
- Assume 40 byte Avg row size
- 10 MB per day
- 3.65 GB per year

Transaction Item Fact Table Size

- One row per sales item per order.



- 3,750,000 rows per day
- 1.4 billion rows per year
- Assuming 50 bytes per row
- 188 MB per day
- 68 GB per year

No Indicators or Flags …

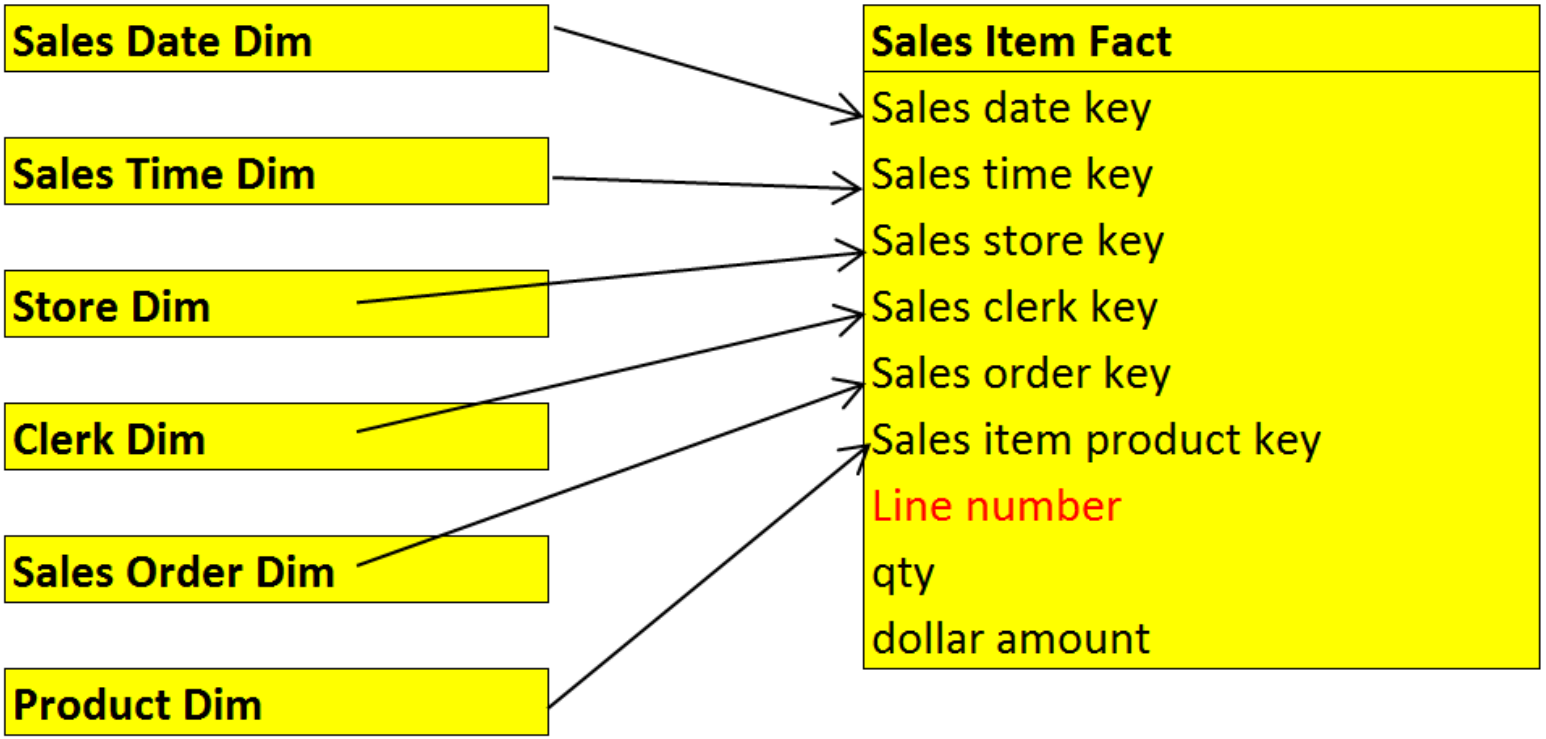
Daily Product Sales by Store Summary Fact
sales date key
store key
product key
distribution mgr key
on promotion flag
daily product sales qty
daily product sales dollars

No Dates or Timestamps …

Daily Product Sales by Store Summary Fact
sales date
sales date key
store key
product key
distribution mgr key
audit key
update date time
daily product sales qty
daily product sales dollars

Exceptions

- Line numbers for line level fact tables



Outline for this Session

- Fact Table Facts
- Granularity
- Fact Table Examples
- Fact Table Contents
- **Fact Table Measures and Additivity**

About Measures and Facts

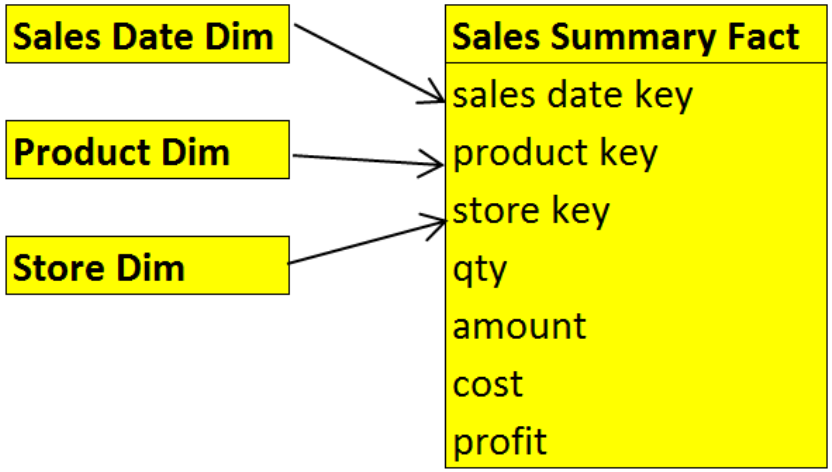
- Measures can be base facts from a source system
- Measures can be derived or calculated from base facts
- Measures are commonly called facts
- A measure may be used in multiple fact tables
- Measures sometimes called metrics
  - Example: Key Performance Indicators or KPIs are metrics or measures; often used in dashboards

Three Types of Facts

- Additive - can sum by any/all dimensions.
  - Examples: Quantity, Cost
- Semi-additive - can add some dimensions but not all
  - Typically additive in all dimensions except date/time
  - Examples: Quantity On Hand, Account Balance,
- Non-additive - cannot be summed; must be calculated from other facts.
  - Example: a ratio (sum of numerator / sum of denominator)
  - But can apply aggregate functions such as Average, Max, Min

Additive Measures

- Can be summed across all dimensions and all combinations of dimensions



- day, sum(amt)
- store, sum(amt)
- product, sum(amt)
- day, store, sum(amt)
- day, product, sum(amt)
- store, product, sum(amt)
- day, store, product, sum(amt)

Semi-Additive Measures

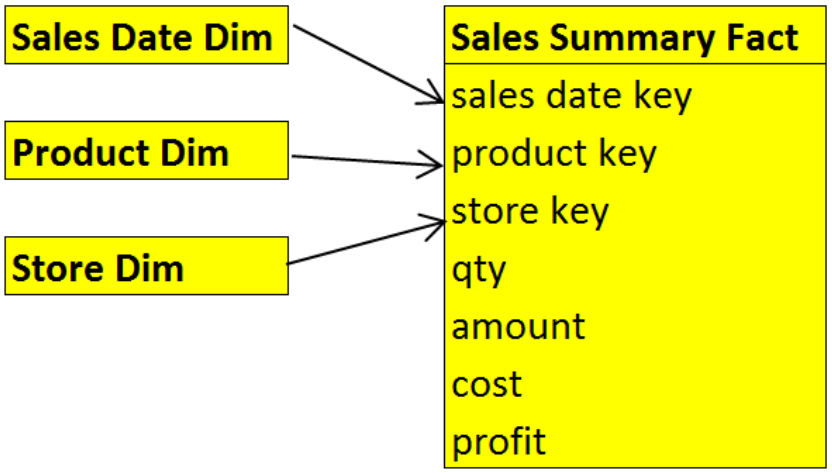
- Can be summed across some, but not all dimensions

Inventory Level Fact
date key
warehouse key
product key
quantity on hand

- The sum of product qty on hand across all or a sub-set of warehouses at a point in time makes sense
- The sum of product quantity on hand across multiple points in time does not make sense

Non-Additive Measures

- Can not be summed across all dimensions, but can be aggregated other ways (avg, min, max)



- Unit price cannot be summed across any dimensions, but can be averaged or checked for the min or max values encountered

Outline for this Session

- Fact Table Facts
- Granularity
- Fact Table Examples
- Fact Table Contents
- Fact Table Measures and Additivity

References

- Kimball, Ralph, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker, *The Data Warehouse Life Cycle Toolkit, Second Edition*, Wiley, 2008, ISBN 978-0-470-14977-5
- Schmitz, Michael D. UCI Irvine Data Warehousing Notes (2014), High Performance Data Warehousing
- Simon, Alan. CIS 391 PPT Slides
- Jeltema ,Bernie, UCI Irvine Data Warehousing Notes (2014), Strategic Frameworks, Inc.