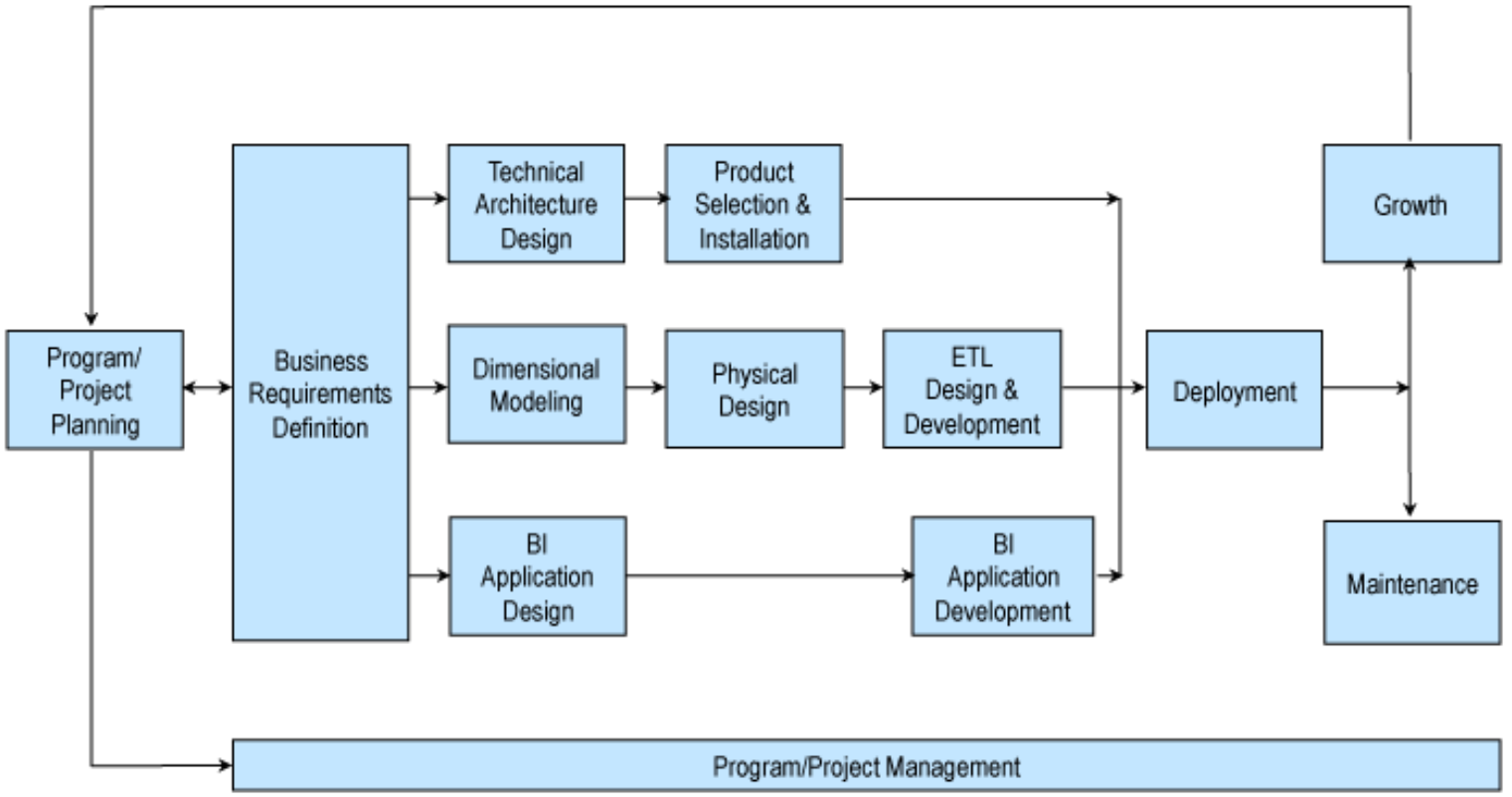
DBW624 – Lecture 2
Business Requirements and Business Process Dimension Model

Business Requirements

- Kimball Lifecycle
- http://www.kimballgroup.com/2009/08/04/design-tip-115-kimball-lifecycle-in-a-nutshell/



The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Requirements

- Business driven vs IT driven
- Must start with business value
    - Business sponsorship
        - Visionary
        - Resourceful
        - Reasonable
    - Enterprise-level business requirements
        - Each business process generally has it’s own system
        - Examples : order entry system or call tracking system
        - Conduct business and IT interviews
        - Conduct reviews of enterprise requirements
        - Well defined success criteria
        - Data Auditing / Data Profiling
        - Interview summary
        - Business requirements category to supporting business process
            - Figure 1-5

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Requirements

- Figure 1-5 – Business Requirements and Supporting Business Processes

Business Requirement Category	Inferred or Requested Analysis	Supporting Business Process	Comments
Sales Planning	Reseller historical orders analysis	Orders	By customer, by territory, by sales region (from state)
	Sales forecast	Orders	Forecast is a business process that uses orders data as an input
Sales Performance	Order by current territory	Orders	
	Orders by original territory	Orders	
	Sales rep performance report	Orders, Orders forecast	Orders and forecast by sales rep
Sales Reporting	Resellers ranked by orders in a given territory	Orders	
	Churned customer list	Orders	Customers who have not ordered in X months
Price Lists	Current price list	Orders	This is a connectivity issue, not a data warehouse issue
Special Offers	Relevant customers by territory based on order history	Orders	
	Inventory status (out of stock)	Inventory	
Customer (Reseller) Satisfaction	Customer Satisfaction Dashboard	Multiple	This is a compound requirement based on several underlying business processes
	Calls by complaint type, product and customer attributes	Call tracking	
	Order metrics of satisfaction	Orders	Eg: due date vs ship date
	Returns by reseller by return reason	Returns	

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Summary of Business Processes Derived from Business Requirements

(Fig 1-5)

Table 1.1: A subset of business processes derived from the requirements interviews

Letter	Business Process	Supported Business Analyses
A	Orders	Orders reporting and analysis, orders forecasting, advertising effectiveness, customer satisfaction, production forecasting, product profitability, customer profitability
B	Orders forecast	Sales performance, business planning, production forecasting
C	Call tracking	Call center performance, customer satisfaction, product quality, call center resource planning, customer profitability, product profitability
D	Returns	Customer satisfaction, product quality, customer profitability, product profitability, net sales

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Requirements

- Enterprise-level business requirements (continued)
  - Business matrix
    - Figure 1-6
    - Figure 1-7

- Prioritize business requirements
  - Business value impact vs feasibility
    - Figure 1-8

- Planning the project
  - Break overall business requirements into individual projects
  - A project is a row in Figure 1-7

- Project level business requirements
  - Same process as used for business requirements

If the business requirements portion goes wrong, everything following will be wrong. Spend the time here until you are comfortable and sponsors are in agreement and there's a success criteria

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Requirements

- Figure 1-6 – Enterprise Business Matrix

Business Process	Dimensions						
	Date	Product	Vendor	Shipper	Dist Ctr	Store	Promo
Purchase Orders	X	X	X		X		
Distribution Center Deliverables	X	X	X	X	X		
Distribution Center Inventory	X	X			X		
Store Deliveries	X	X		X	X	X	
Store Inventory	X	X				X	
Store Sales	X	X				X	X

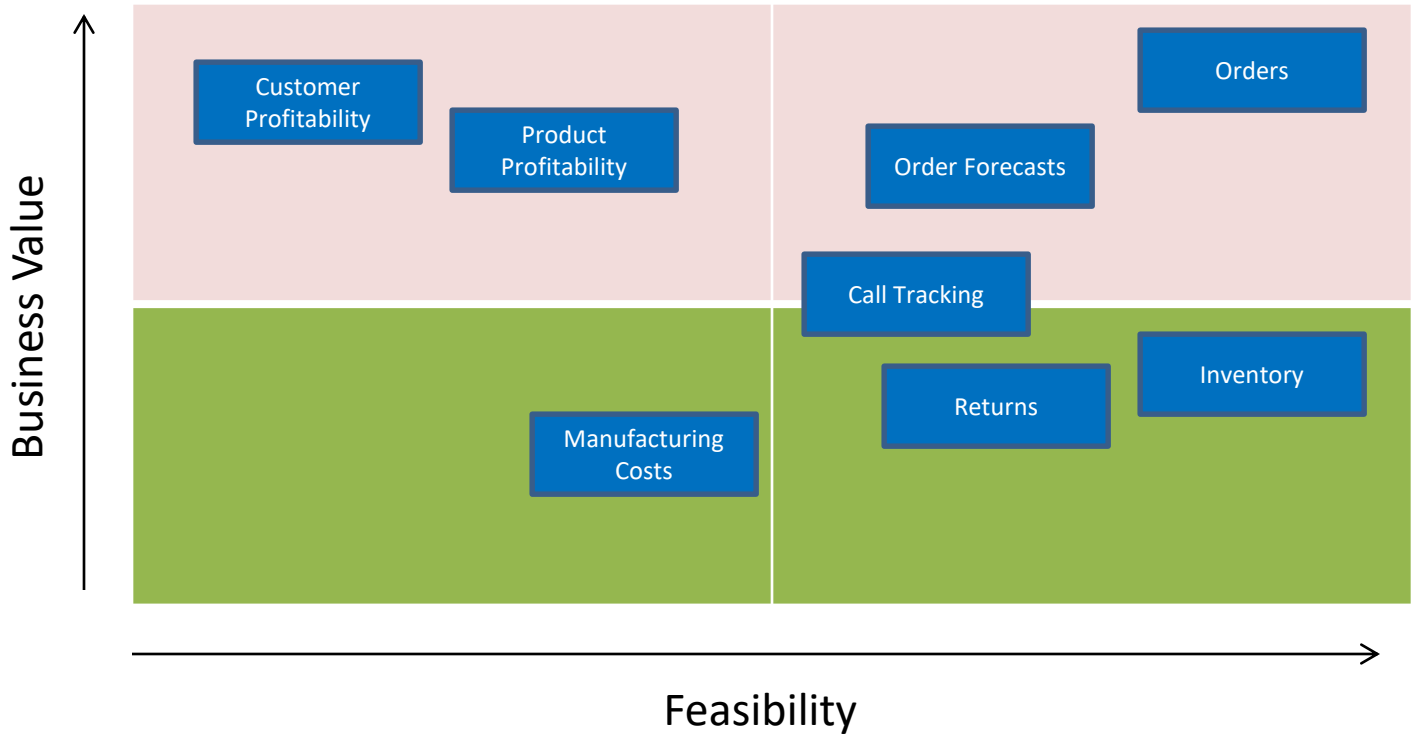
- Figure 1-7 – Enterprise Business Matrix (from Figure 1-5)

Business Process	Dimensions										
	Date	Product	Employee	Customer (Reseller)	Customer (Internet)	Sales Territory	Currency	Channel	Promotion	Call Reason	Facility
Orders Forecasting	X	X	X	X	X	X	X	X			
Orders	X	X	X	X	X	X	X	X	X		
Inventory	X	X	X								X
Call tracking	X	X	X	X	X	X				X	
Returns	X	X		X	X	X	X		X		X

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Requirements – How to Prioritize

- Figure 1-8 – Prioritization Grid



The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

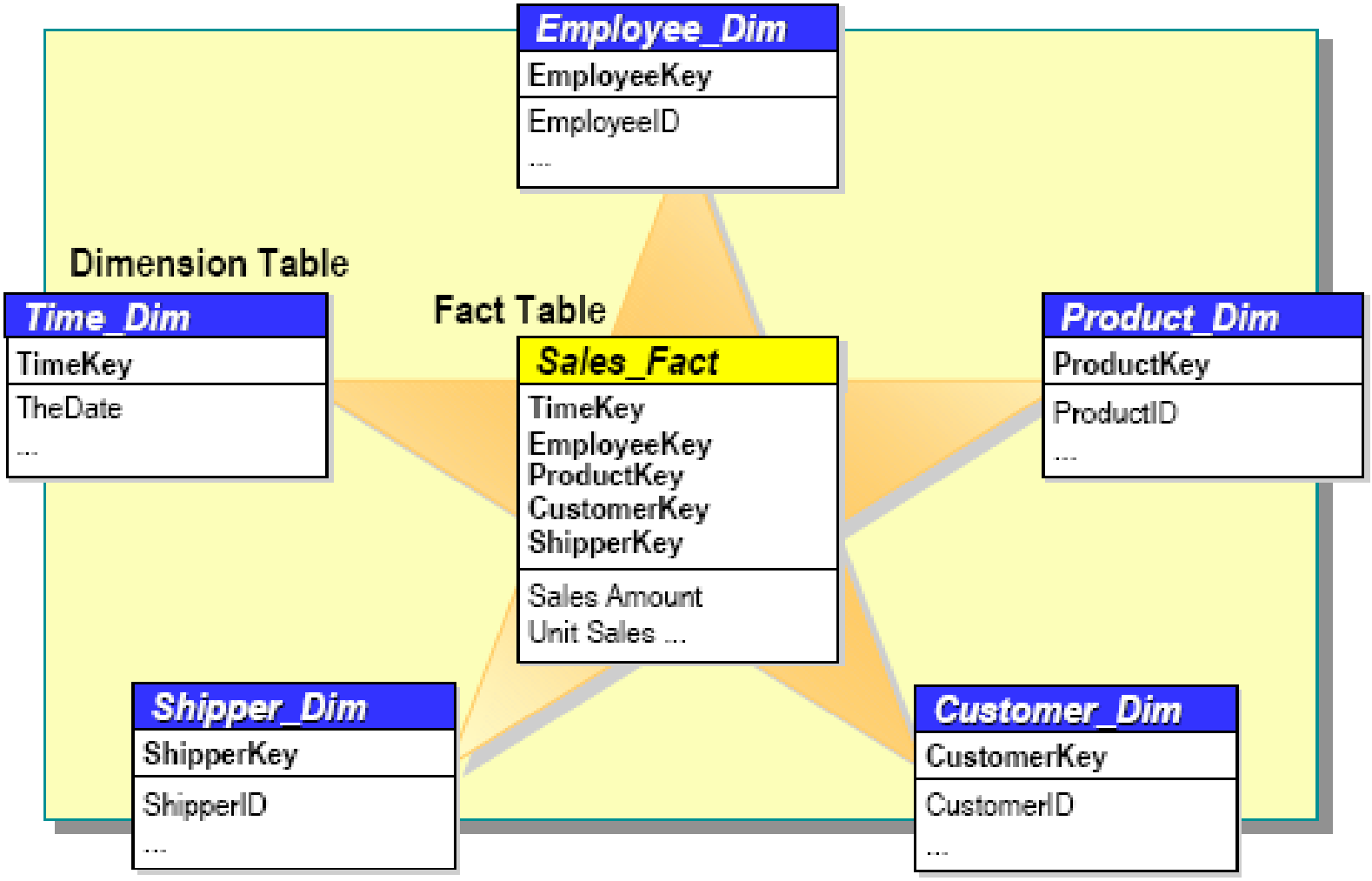
- Goal : Make sure the users get access to the data they need

- Dimensional Model
  - Heart of the DW/BI system
  - Target of the ETL system
  - Structure of the physical database
  - Model behind the query and reporting applications

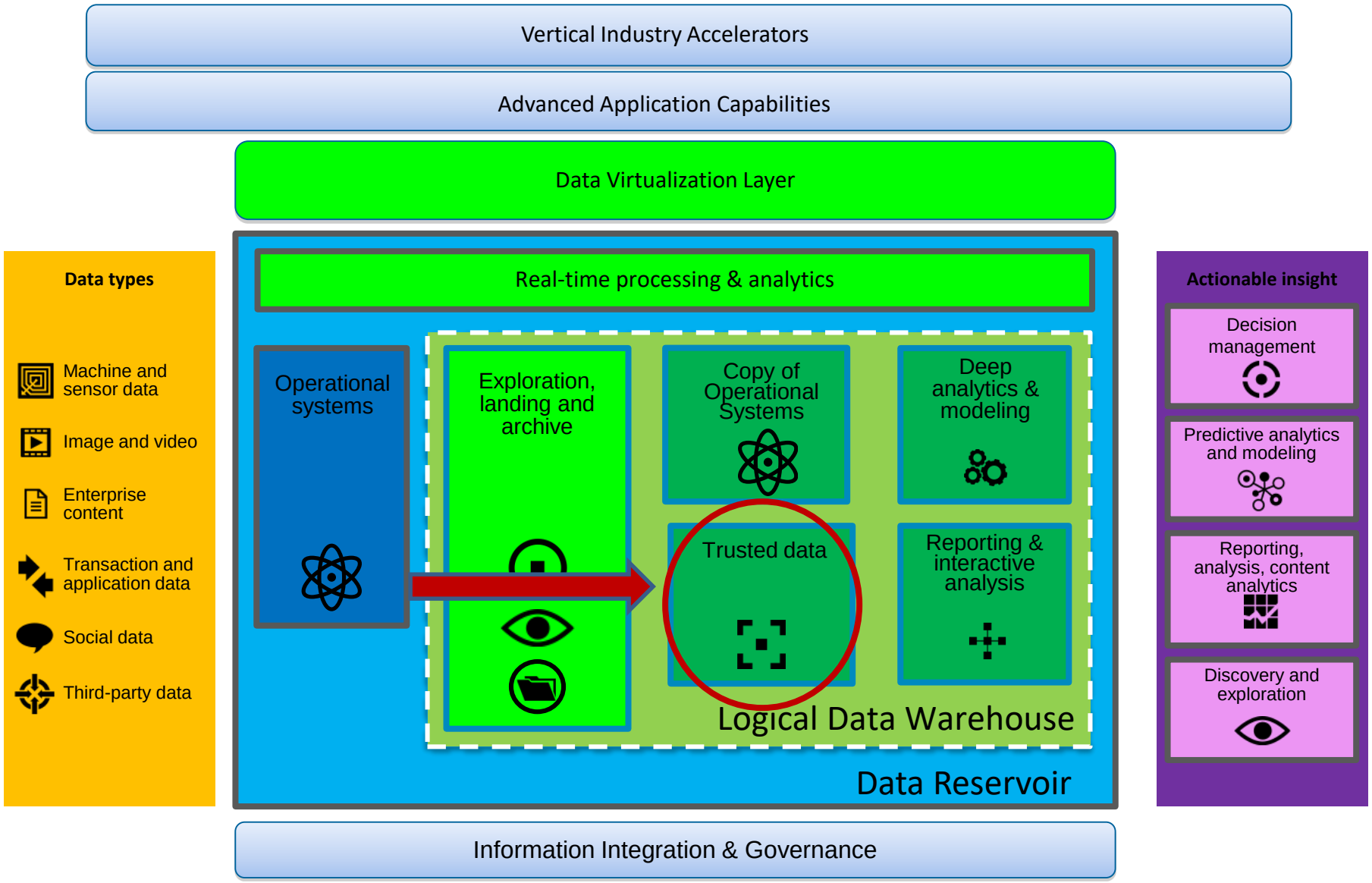
- Dimensional Model – Success criteria
  - Simple
  - High performance
  - Relevant / Accurate

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

The Star Schema



Analytics Platform: Location of Dimension Model



Business Process Dimension Model - Concepts

- Facts
  - Fact table contains the measurements associated with a specific business process (ex : admitting a patient, customer support request)
  - Usually the primary key of a fact table contains multiple parts which map to dimension tables
  - Example : admitting a patient
    - Patient ID
    - Name
    - Date
    - ….
  - Atomic fact tables are important
  - Single grain – very granular detail
  - Derived facts/columns (ie: sales tax, profit, currency exchange)
  - Three fact table types
    - Transaction
    - Periodic snapshot
    - Accumulating snapshot
  - Fact tables are NOT normalized – repeated information is fine (like city)
  - Fact tables are just tables in a database

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Grain of a Fact Table

- Fact table consists of measurements of a business process
- The grain is the description of the measurement event in the physical world that gives rise to a measurement
- The grain of the dimensional model is the finest level of detail that is implied when the fact and dimension tables are joined
- Examples
  - When the grocery store scanner measures the quantity and the charged price of a product being purchased, the grain is literally the beep of the scanner
  - The grain of a SALES fact table might be stated as "Sales volume by Day by Product by Store“
    - Each record in this fact table is therefore uniquely defined by a day, product and store.

- Fact Tables are usually very large – 90+% of the data in the star schema

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Fact Table – Usual Columns

Surrogate Key
Degenerate Dimension
Foreign Key 1
Foreign Key 2
…
Foreign Key n
Attribute 1
Attribute 2
…
Attribute m
Calculation/Measure 1
Calculation/Measure 2
…
Calculation/Measure p

No real business meaning – provide uniqueness and mapping to source

Core to the star schema – foreign keys map to primary keys of dimension tables

Information from transaction record or information from dimension tables used in frequent analytic queries

Information that is calculated based on the attributes to avoid repetitive calculations for each analytic query

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

- Dimensions
    - Tables in a database
    - Dimension tables are NOT normalized
    - Dimension table are designed for:
        - Simplification
        - Performance
    - Spotting dimensions
        - “by” …. “group by”, “order by”
    - Ability to re-use dimensions is key to enterprise data warehouse concept
    - Conformed dimension
    - Drill across
    - Example : Customer purchases with customer support calls
    - Figure 2-2 … putting dimensions and facts together

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Conformed Dimensions

- When two business processes use the exact same product dimension with the same keys, they use a *conformed* dimension
- Conformed dimensions are the cornerstone of the *enterprise-enabled* DW/BI system.

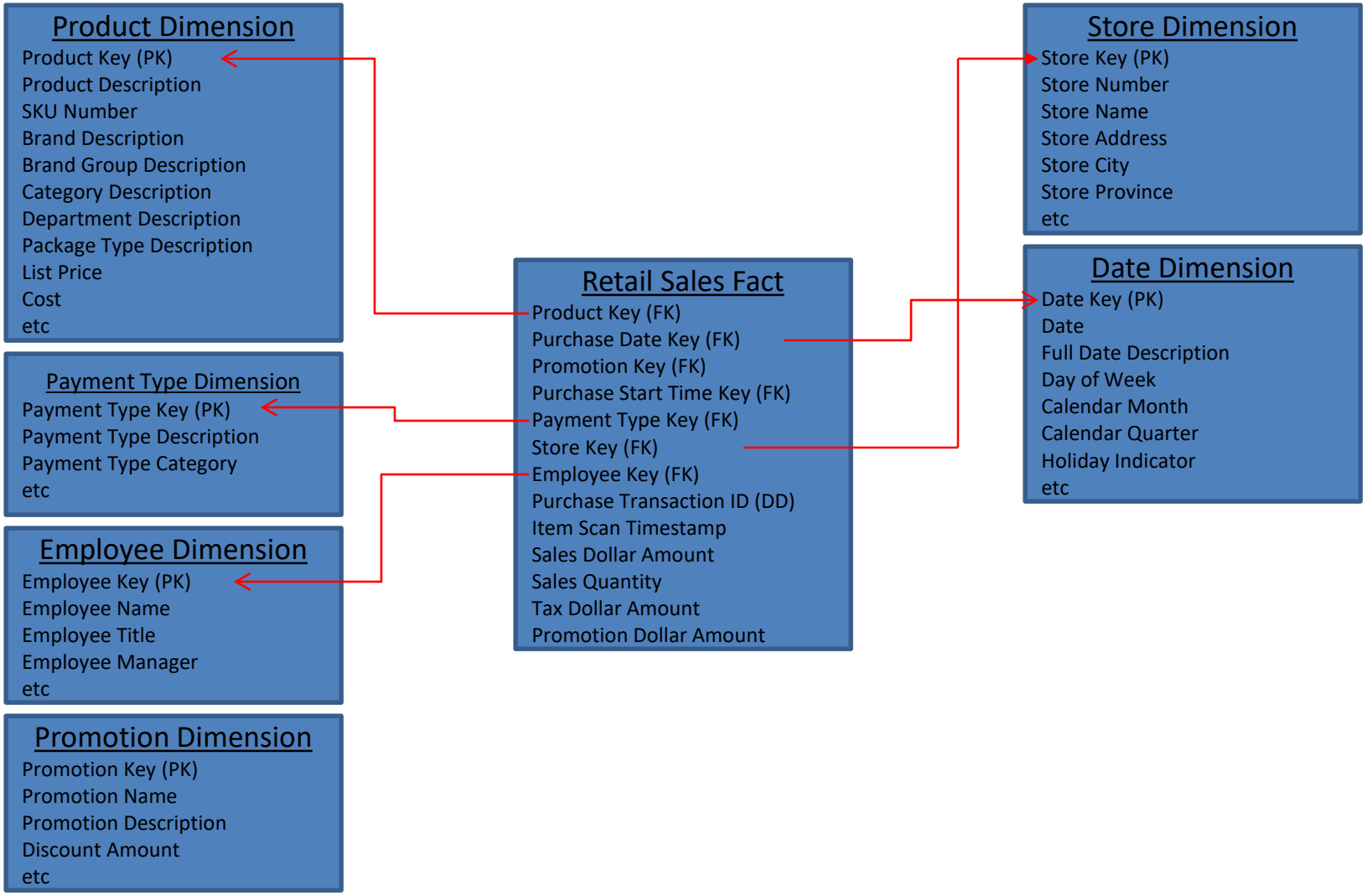
Example: Retail Grocery store

- A dimension such as product will be used in both the retail sales and the store inventory dimensional models
- Because they are exactly the same products, both models must use the same dimension with the same keys to reliably support true, cross-business process analysis.

- This kind of analysis involving data from more than one business process is called *drill across*

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Figure 2-2 – Basic Dimensional Model



The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

- Size of Fact tables vs Dimension tables
- Fact tables are usually >90% of the size of the star schema

- Example: McDonalds sales warehouse
  - In 120 countries around the world
  - 36,899 restaurants around the world
  - 375,000 employees worldwide
  - 145 items on their menu (products)
  - 1-3 marketing specials at any one time
  - 68,000,000 customers each day worldwide
- Assume: Each customer buys an average of 3 items per visit
- Assume: Want to keep 5 years of data for historical analytics and trend analysis

- Stores Dimension Table: <37,000 rows
- Employee Dimension Table: <400,000 rows
- Product Dimension Table: <200 rows
- Date Dimension Table: ~365 x 5 rows, so, <2000 rows
- Marketing Promotion Dimension Table: < 100 rows
- Fact Table: 68,000,000 x 3 (items) x 365 (days) x 5 (years) = 372,300,000,000 rows

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

- Surrogate Keys
  - Separate from keys in the transaction system
  - Generally meaningless
  - Required for :
    - Protect against changes in transaction (source) systems
    - Allows for integration of data from different sources
    - Allows for manipulation of data in dimensions (add rows)
    - Track changes in the dimension table
    - Performance

- Slowly Changing Dimensions
  - Employee – years of service
  - Employee – department number
  - You want to be thinking about which information may change
  - You also want to think about how important currency is

- Dates
  - Fundamental dimension
  - Perfect example of where you need surrogate keys

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Surrogate Key Example

The following table shows an extract of the product reference data

Product_code	Product_name
PRX 002 39 061	Trailchef water bag
PRX 003 40 067	Trailchef canteen
PRY 003 47 086	Trailchef kitchen set

Same table with the use of surrogate keys for product reference data

Surrogate_id	Product_code	Product_name
1	PRX 002 39 061	Trailchef water bag
2	PRX 003 40 067	Trailchef canteen
3	PRY 003 47 086	Trailchef kitchen set

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

- Degenerate Dimensions
  - Included in fact table, but, not in any dimension table
  - Example : transaction ID
  - Can be good to tie back to the transaction system

- Snowflaking
  - Act of connecting look-up tables to fields in the dimension table
  - Discouraged – opposite of denormalization
  - Example : address map from postal code to city to province, etc

- Many-to-many Relationships
  - Desire : One to many
    - One row in dimension table maps to many rows in fact table
  - Complex : Many to many
  - Example : multiple sales people doing same sale
    - Need to create a group table

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model - Concepts

- Hierarchies
  - This is the way in which we drill down on information
  - Variable depth hierarchies
    - Example : bill of material
    - SQL can handle this (recursion), however, very complex

- Aggregates
  - Creating of summary tables
  - Great for performance
  - Easily accessible through SQL
  - Particularly important for ad-hoc queries
  - Examples:
    - Revenue of month to date
    - Profit of month to date
    - Products sold so far this month
    - Summaries of previous month/quarter/year, etc.

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Business Process Dimension Model – Modeling

- Steps :
  - Graphical model (high level model)
  - Detailed modeling
    - Definitions
    - Sources
    - Relationships
    - Data quality issues
    - Transformations required
  - Review & validation

- Considerations :
  - Data architecture strategy
  - Naming conventions
  - What data do you keep around

- Develop the detailed dimensional model
- Testing and refining the model – this is an iterative process
- Reviewing the validating the model

The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

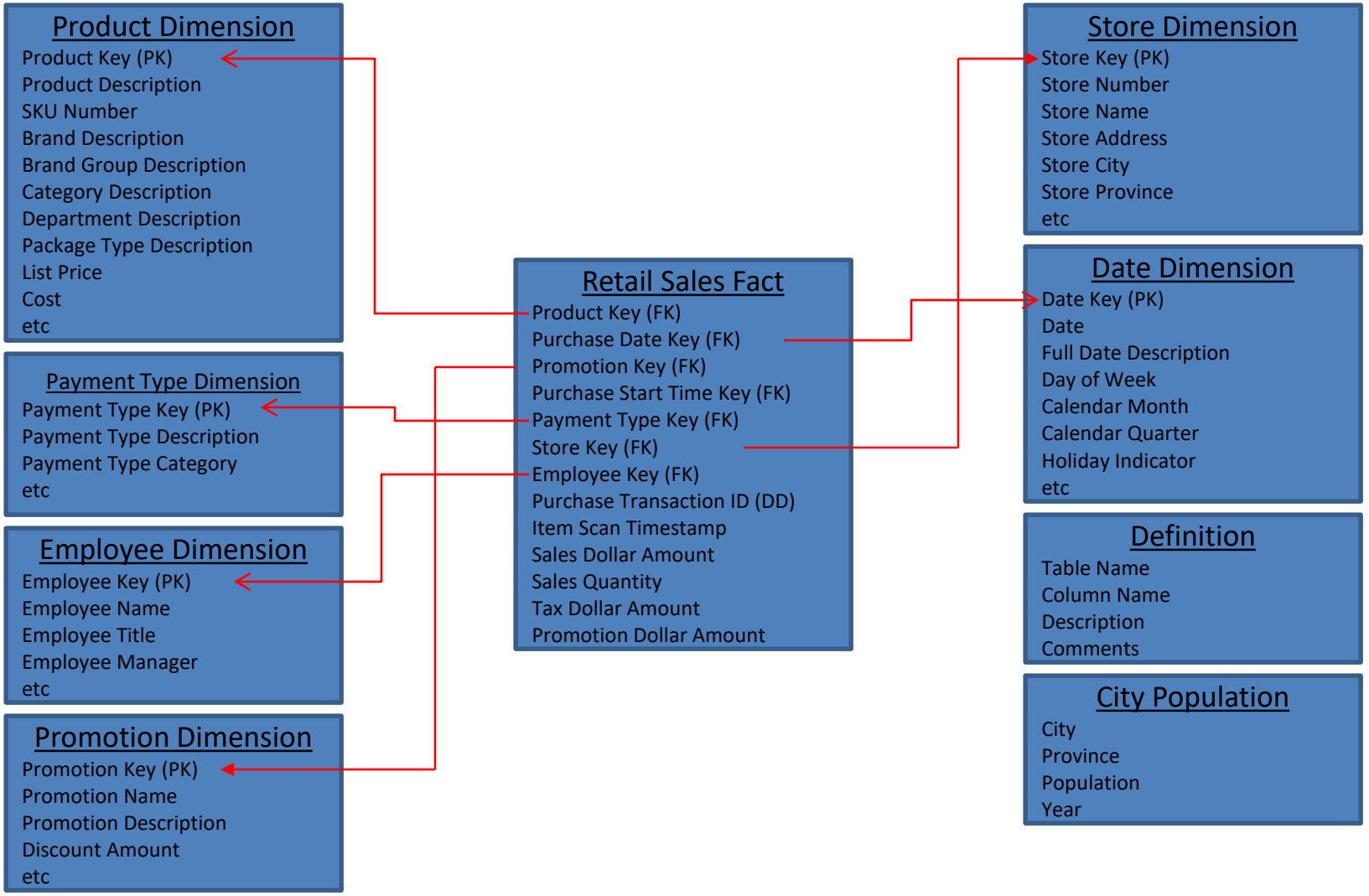
Business Process Dimension Model – Other Table Types

- Definition Table
  - Used to keep track of definitions and important information about your warehouse design and data model
  - Think of this as equivalent to having comments in a program
  - Objective is to make sure if someone else comes along, they can figure out what was used, how and why
  - Columns would include:
    - TABLE_NAME
    - COLUMN_NAME
    - DESCRIPTION
    - COMMENTS
    - (maybe others)

- Reference Tables
  - Used to store information important to the business, but, not directly mapped to the star schema with primary / foreign keys
  - Examples:
    - Currency exchange rates
    - Mapping of postal codes or zip codes

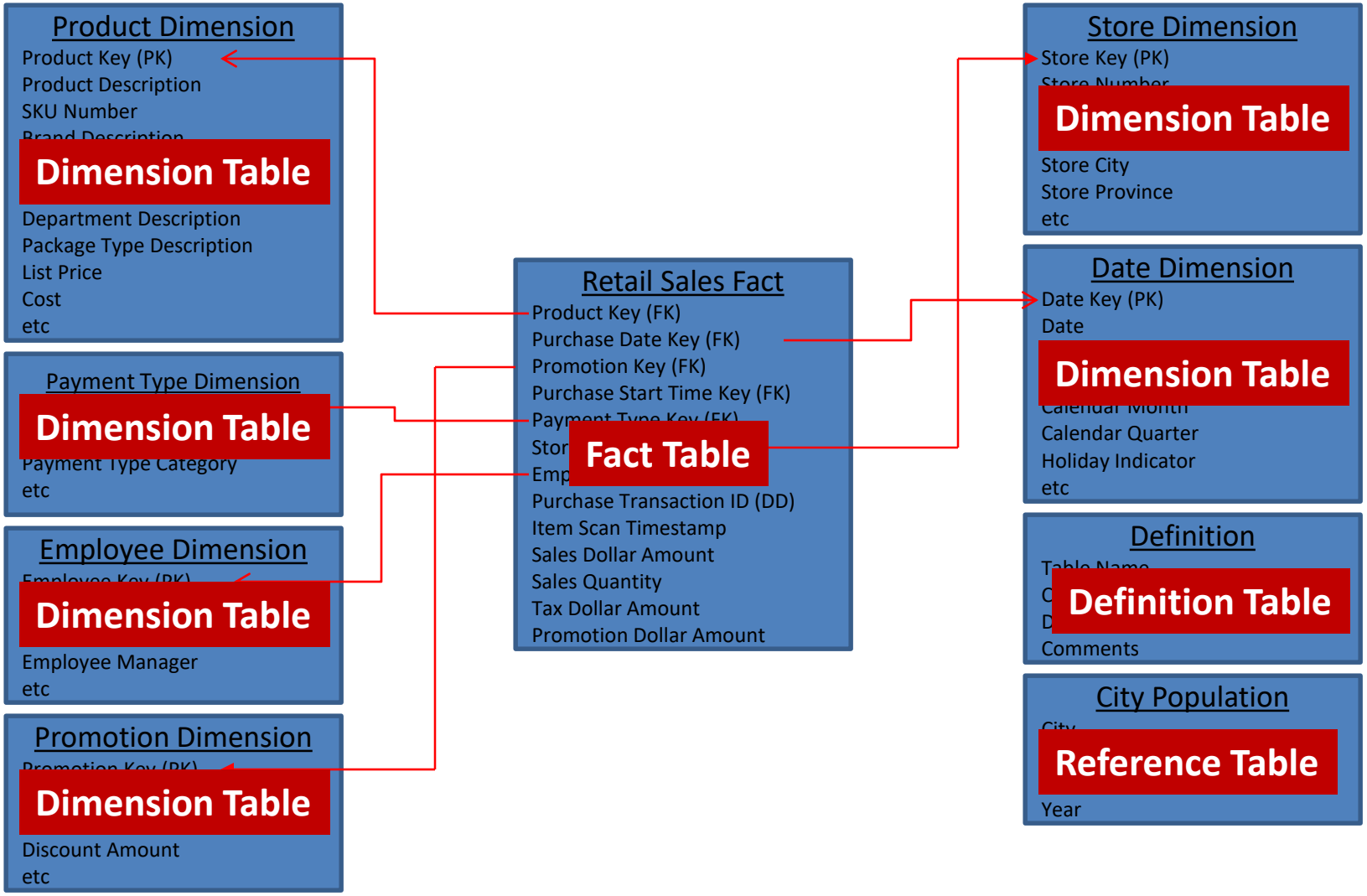
The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Dimensional Model Including Definition and Reference Tables



The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball

Dimensional Model Including Definition and Reference Tables



The lecture notes are based off of the course text which is : The Microsoft Data Warehouse Toolkit Second Edition
Author : Joy Mundy and Warren Thornthwaite with Ralph Kimball