Guest Speaker Session

Spot the BI in Retail

Session Debrief

Retail Mix

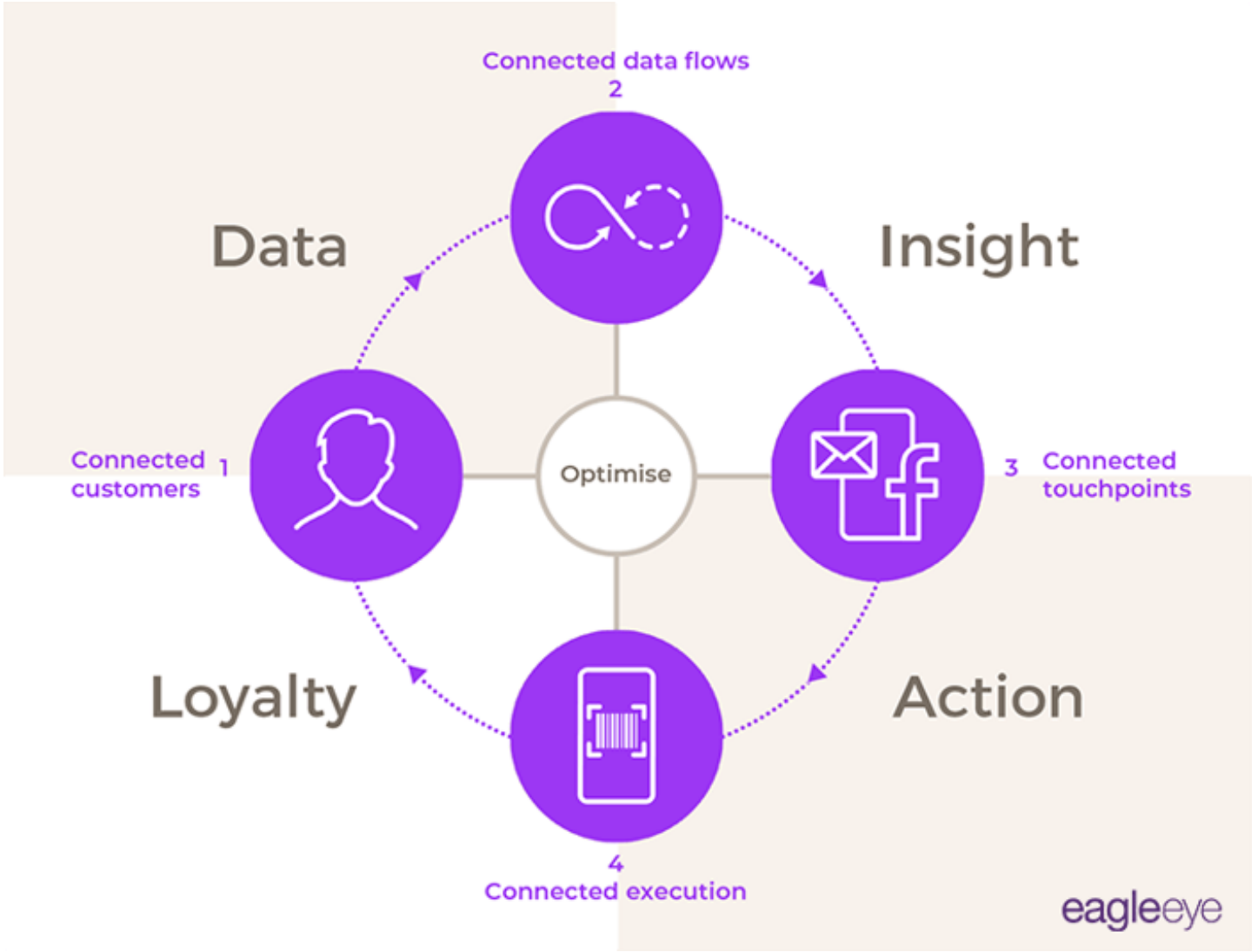

The 6 P's of Retail Marketing
Product
What your business is selling
Price
How much your products will cost
Place
Where you will sell your products
Promotion
How people will find out about your product
People
Who will buy and sell your products
Presentation
What your products look like

Retail Strategy

Walmart Retail Strategy Mix

- Location: Free standing stores
- Assortment: Large # of Categories, few items in each
- Pricing: Low, EDLP (Every day low pricing)
- Communication: TV, Newspaper, social media, etc.
- Store Design&Display: Basic, Special display for products
- Customer Service: Limited

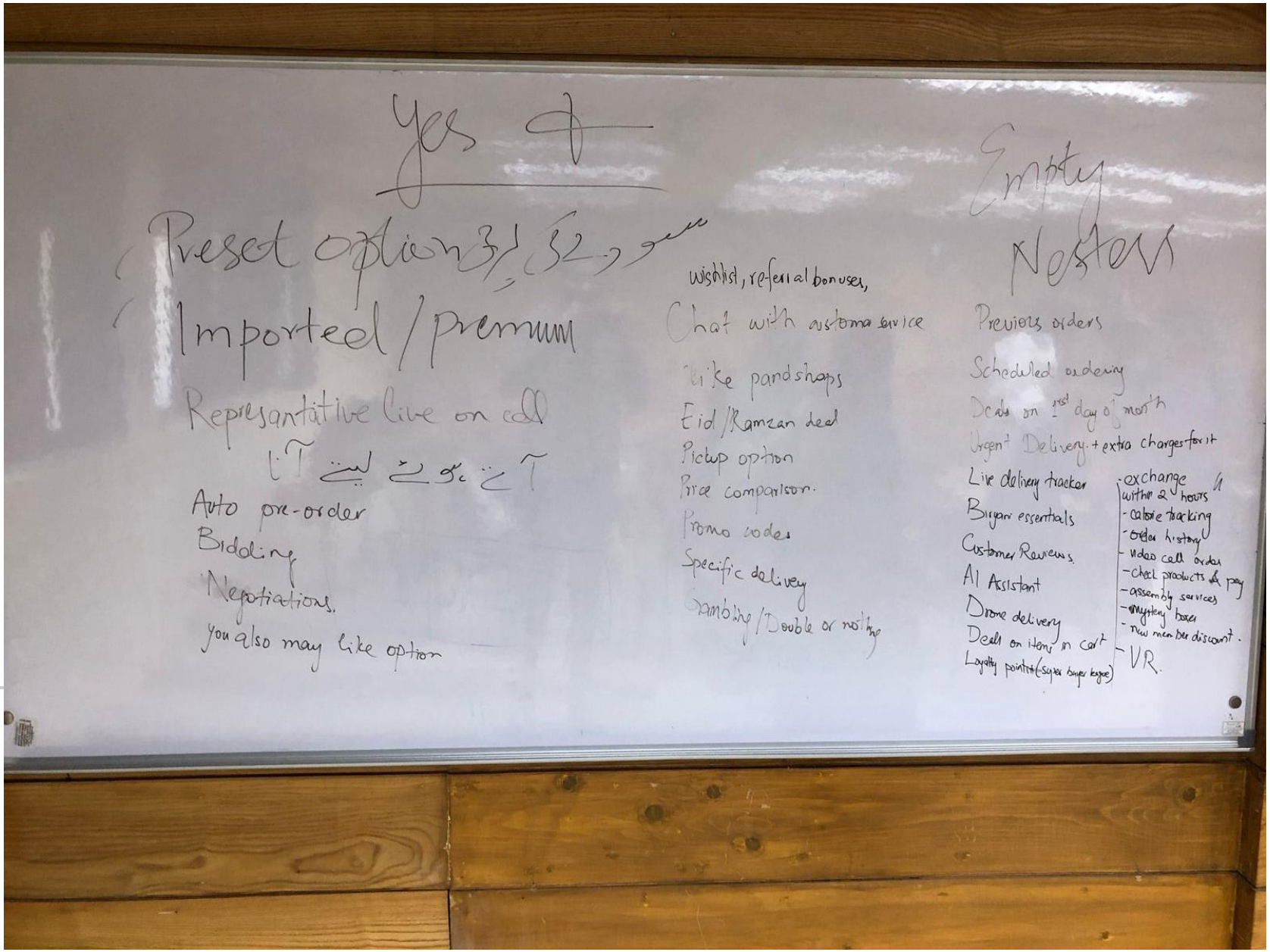
DIAL


Connected data flows
2
Data
Insight
Connected
customers
1
Optimise
3
Connected
touchpoints
Loyalty
Action
4
Connected execution
eagleeye

Empathy for Consumers

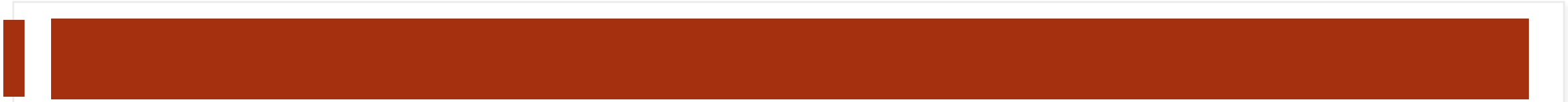
Brainstorming the ideal Imtiaz App

(YES AND... BI_02)



What other things you learned that were new for you?

Guest speaker session or Assignment



DSS Concepts in Action Today

- **Key Features of DSS:**
  - Data Integration: Consolidating structured and unstructured data.
  - Analytical Models: Using AI/ML for predictions.
  - User Interaction: Dashboards and decision-friendly interfaces.

- *"An online store recommends products based on your past behavior. What drives this capability?"*

Connecting BI Features to DSS Concepts

- Which BI features did you spot?
- How do these features support users' and businesses' decisions?

Trace these BI features back to DSS principles

- Data Integration:
  - Which BI features rely on gathering data from multiple sources?
- Analytical Models:
  - How do features like product recommendations showcase predictions?
- User Interface:
  - Which features simplify data visualization for decision-making?

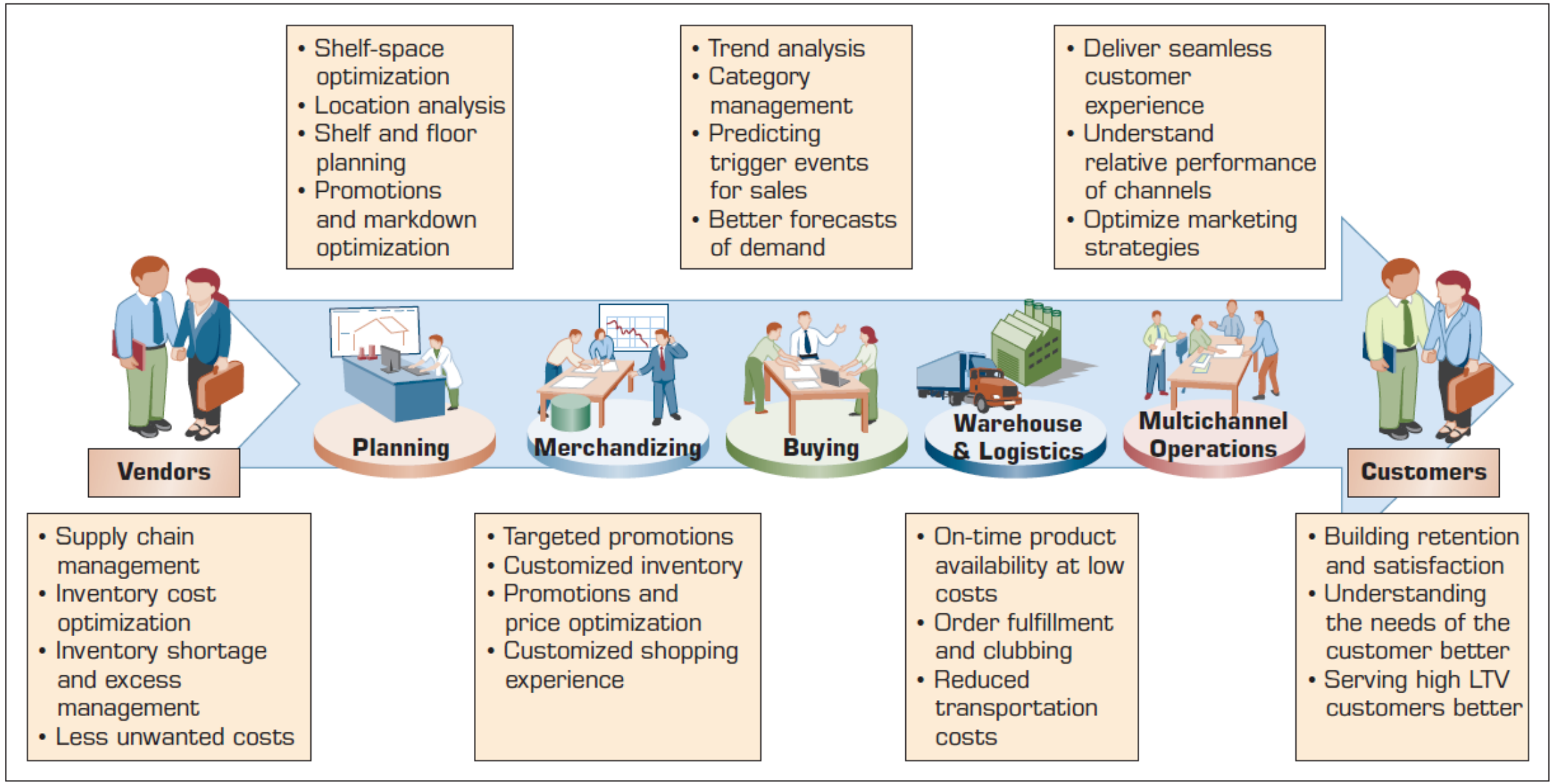


FIGURE 1.12 Example of Analytics Applications in a Retail Value Chain. Contributed by Abhishek Rathi, CEO, vCreaTek.com

Why DSS Still Matters in BI Today

- DSS provides the conceptual backbone for BI systems.
- Modern BI combines DSS principles with advanced analytics (AI, ML) and real-time data.
- BI is about empowering decisions at all levels—individual, organizational, and strategic.

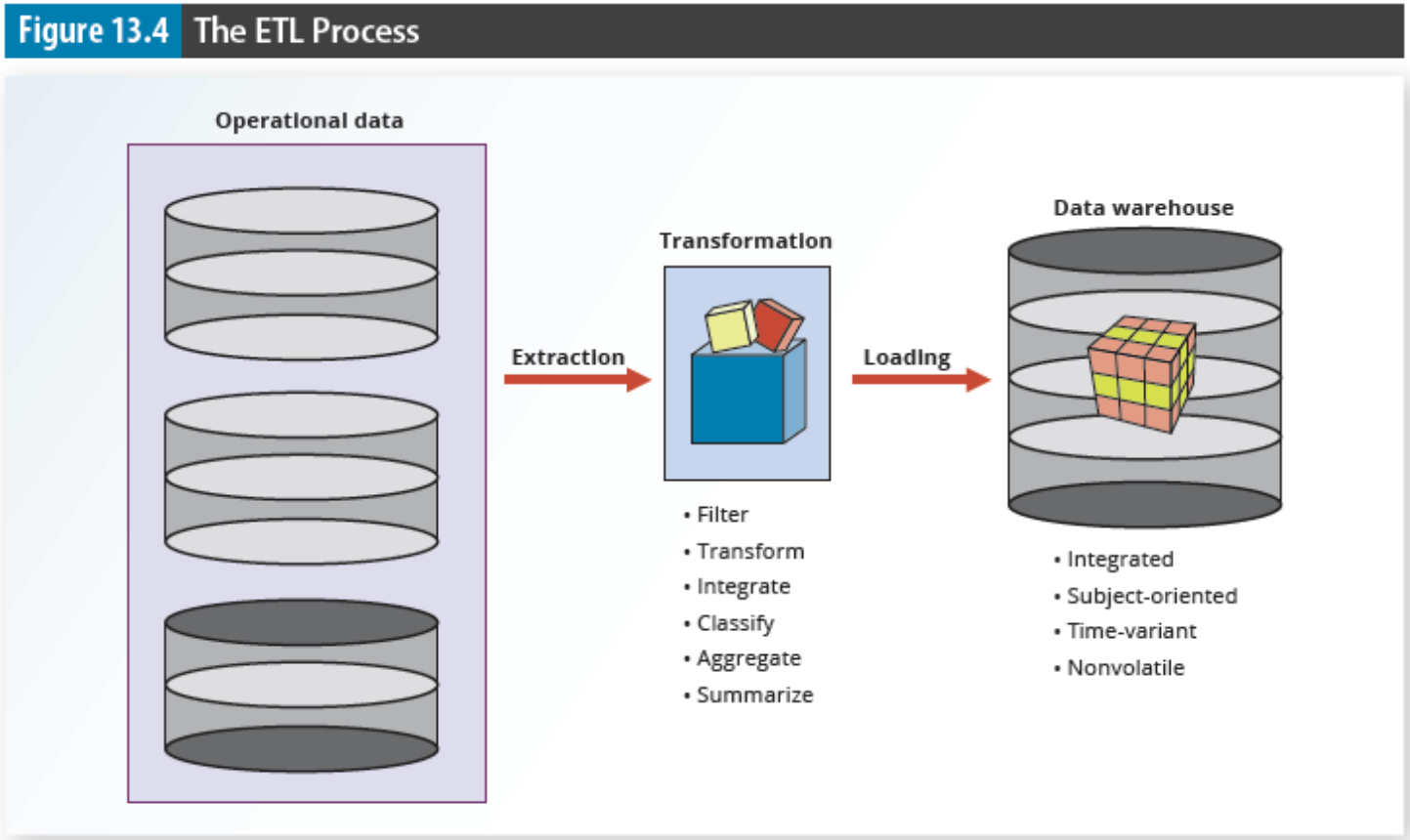
DWH Theory

CS 459 Business Intelligence

Data Warehouse – Expert Definition

- **Ralph Kimball:** "a copy of transaction data specifically structured for query and analysis"
- **Bill Inmon:** "A data warehouse is a:
  - Subject oriented
  - Integrated
  - Non-volatile
  - Time variant

collection of data in support of management's decisions.



Inmon's DWH

Subject-Oriented

Data collected relates to a particular subject (e.g sales, customer, etc.)

Integrated

Data has been standardized regardless of how it is stored in the source systems.

Data Warehouse

Non-Volatile

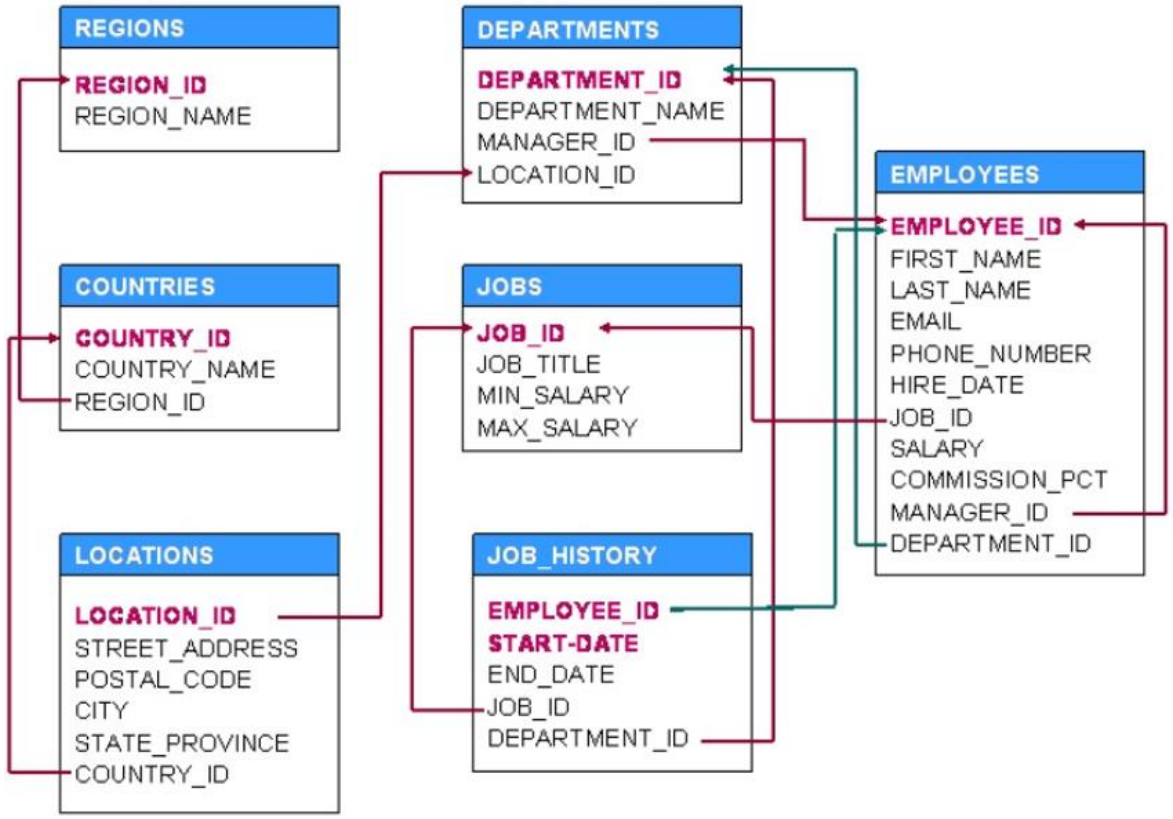
Data in the DW is hardly ever over-written or deleted - once committed, the data is static, read-only, and retained for future reporting

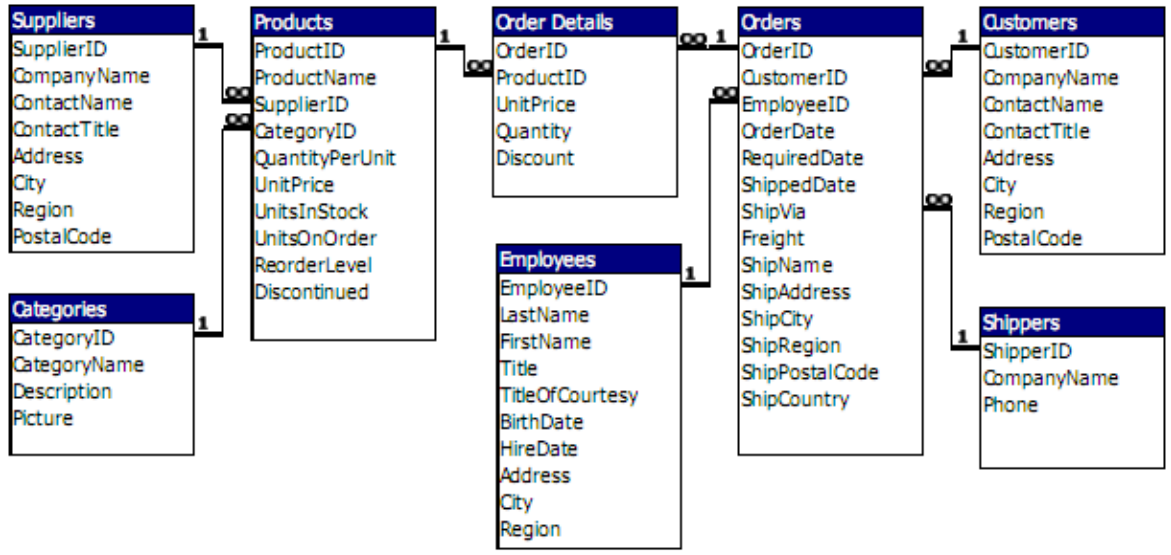
Time Variant

The data collected changes with time. Newer snapshots record the updates.

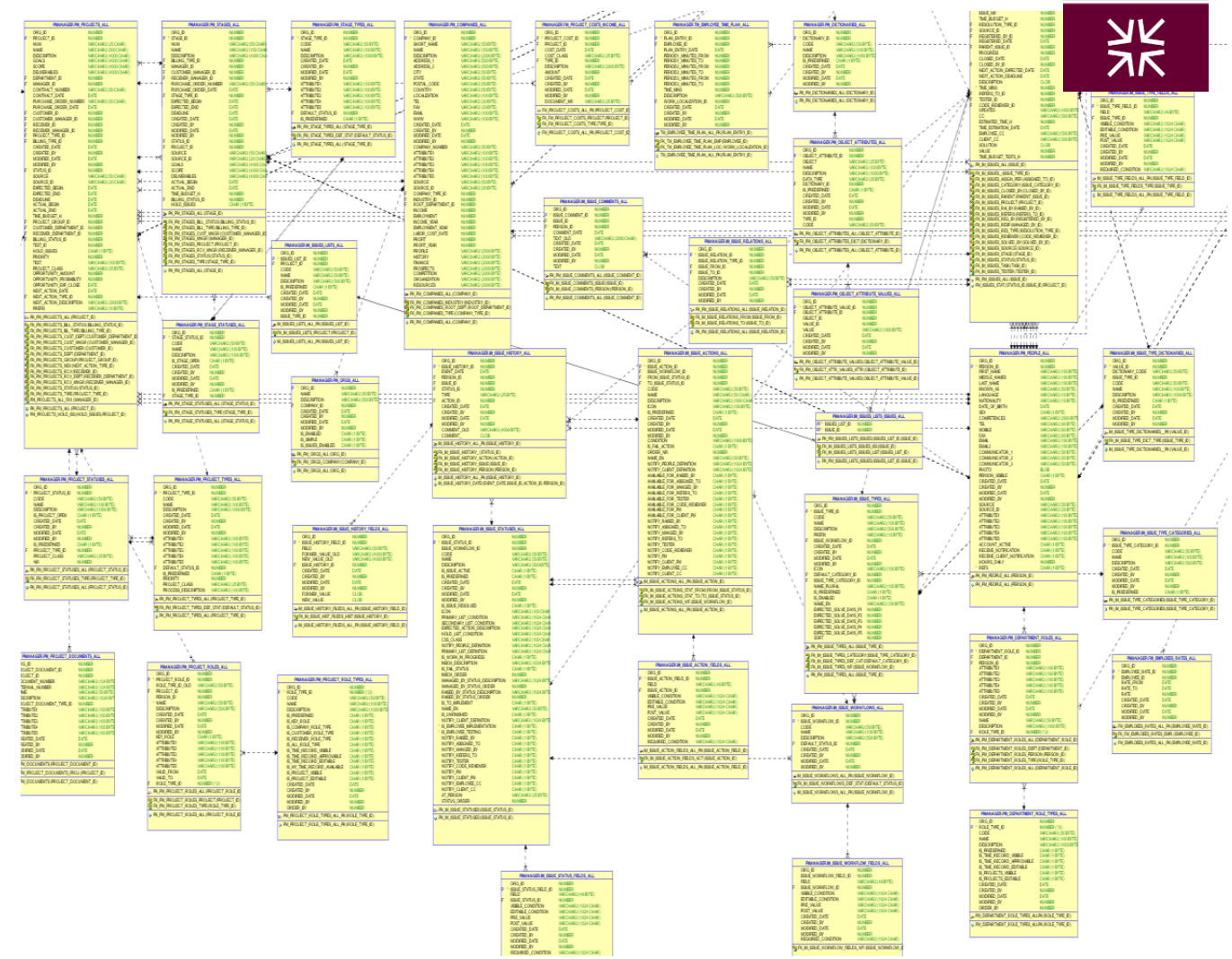
Defining Data Warehouse

Naïve Databases





Database Spaghetti



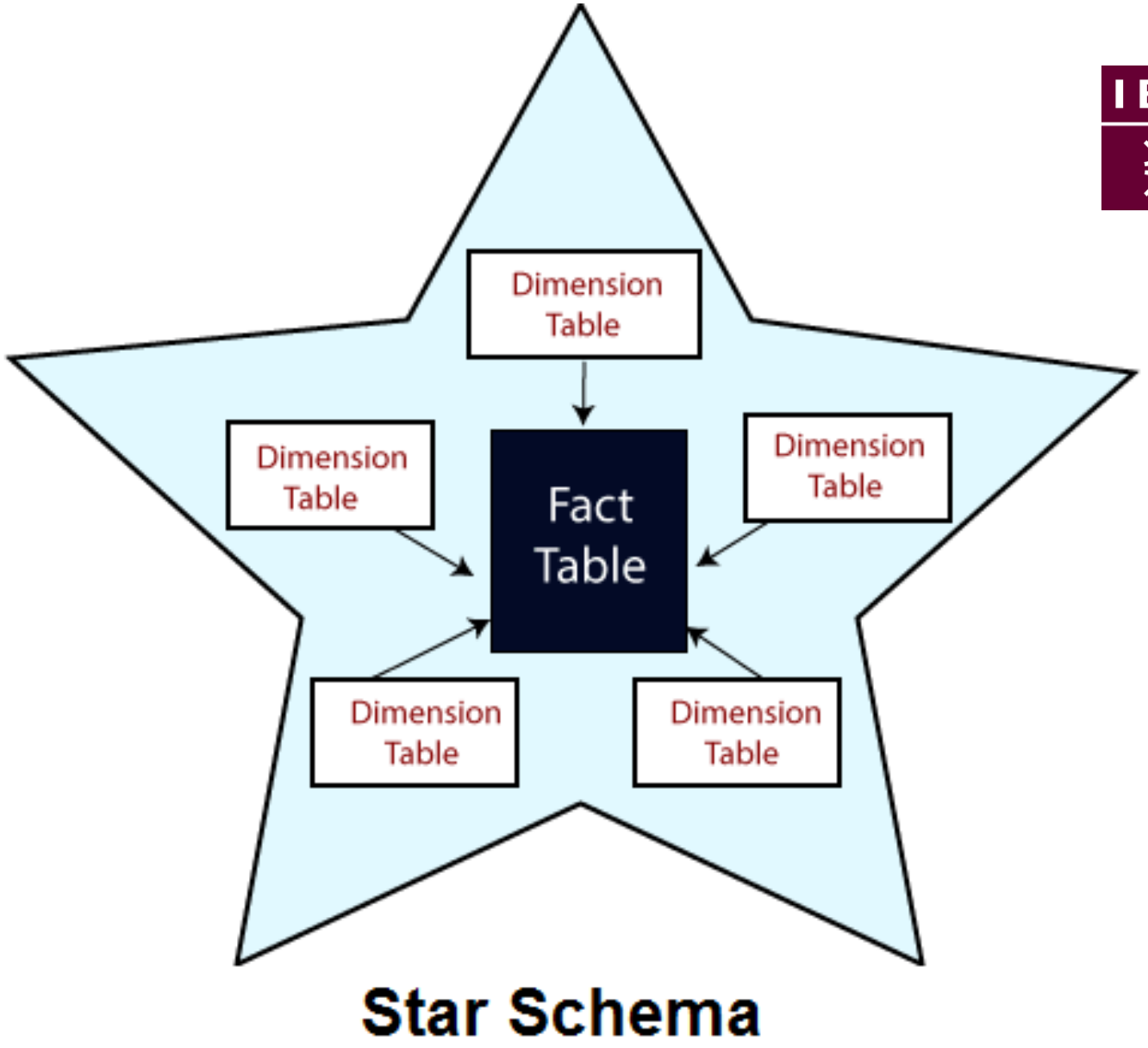
Database Spaghetti

- Normalization decomposes the structure into greater tables.
- Random changes in database leads to further problems.
- Growing business leads to exponential growth of tables.

Problems:

- Multi-table joins → long, time-consuming, complex
- In the olden days, hardware availability was an issue: small size yet too expensive to acquire.

Introducing Star Schema



Star Schema *Analyzing Facts across Dimensions*

- The **fact table** stores two types of information: numeric values and dimension attribute values. Using a sales database as an example:
- **Numeric Value cells (facts)**
  - Are unique to each row or data point and do not correlate or relate to data stored in other rows.
  - These might be facts about a transaction, such as an orderID, total amount, net profit, order quantity or a **business measure**.
- **The Dimension Attribute Values**
  - Store the **foreign key value** for a row in a related dimensional table.
  - Many rows in the fact table will reference this type of information. So, for example, it might store the sales employee ID, a date value, a product ID or a branch office ID.

Star Schema →*Analyzing Facts across Dimensions*

- **Dimension tables** store supporting information to the fact table.
- The dimension tables contain the textual context associated with a business process measurement event.
- They describe the "who, what, where, when, how, and why" associated with the event.
- Each star schema database has at least one dimension table, but will often have many. Each dimension table will relate to a column in the fact table with a dimension value, and will store additional information about that value.

Examples

- The employee dimension table may use the employee ID as a key value and can contain information such as the employee's name, gender, address or phone number.
- A product dimension table may store information such as the product name, manufacture cost, color or first date on market.

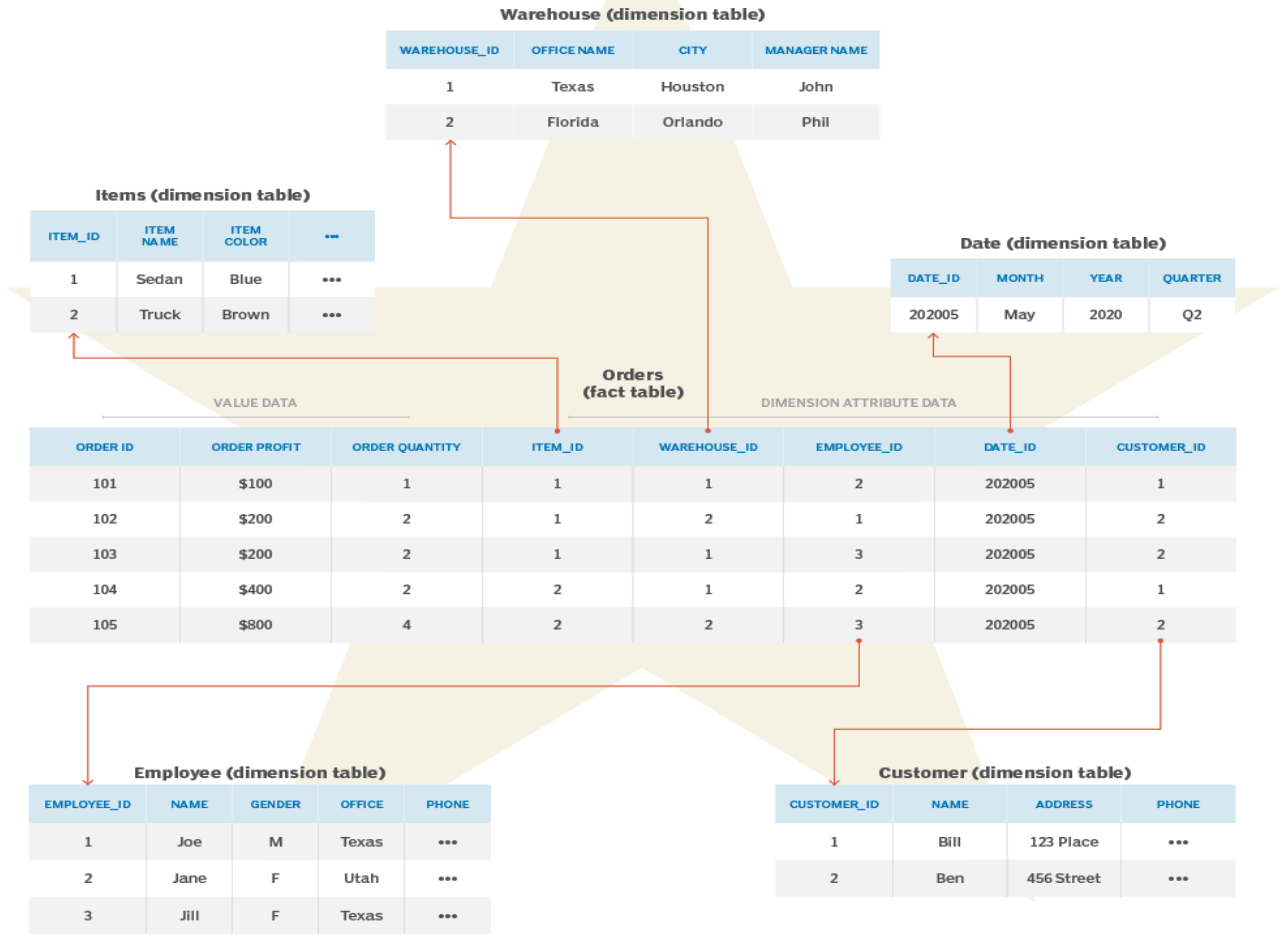
Construct an Orders ERD

- **Required tables**
  - Orders
  - Employees
  - Customer
  - Item
  - Warehouse

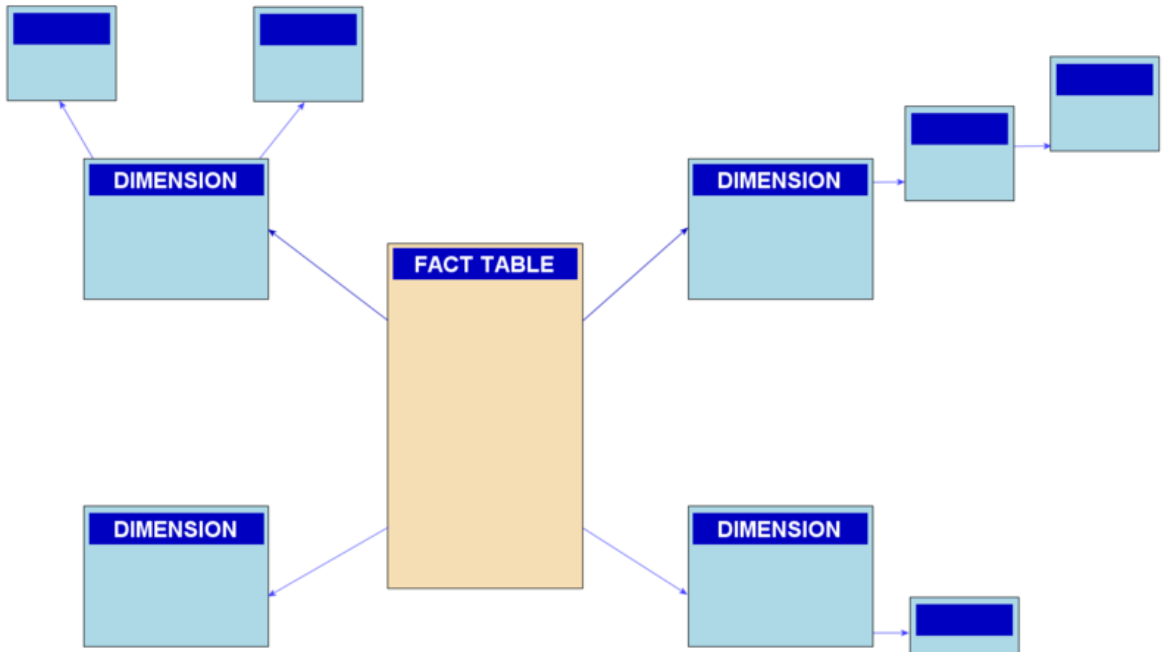
- **Relations:**
  - One order is placed by a single customer and completed by a single employee.
  - Employees can be assigned many orders.
  - Customers can place multiple orders.
  - A single order may have multiple items and items can be ordered in multiple orders.
  - The order is fulfilled by a single warehouse

ERD to Star Schema

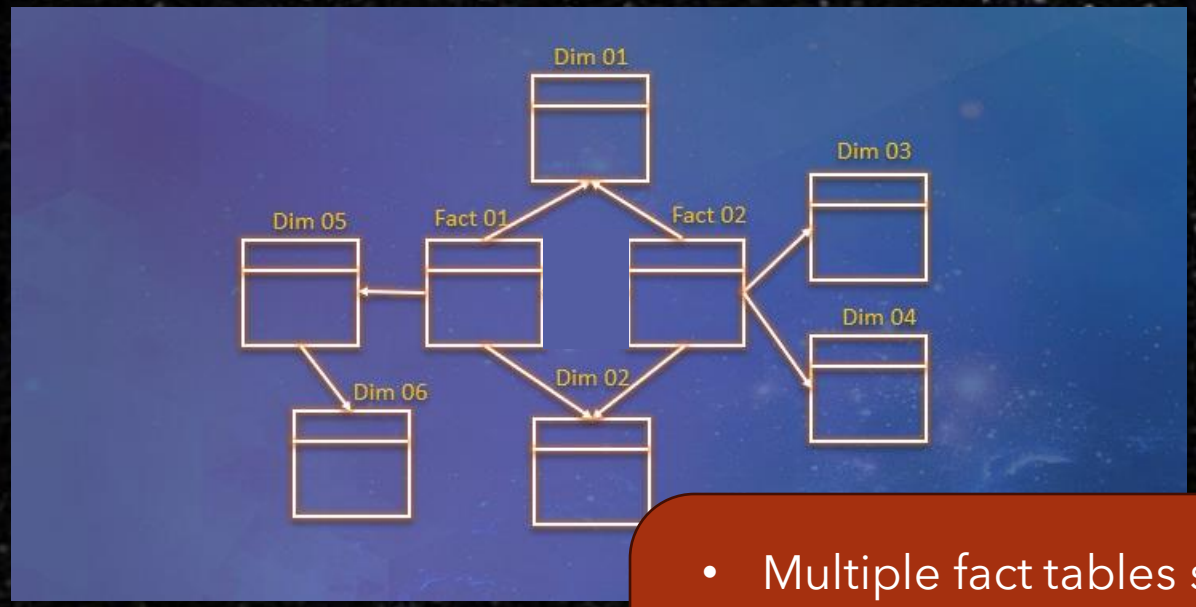
Star schema

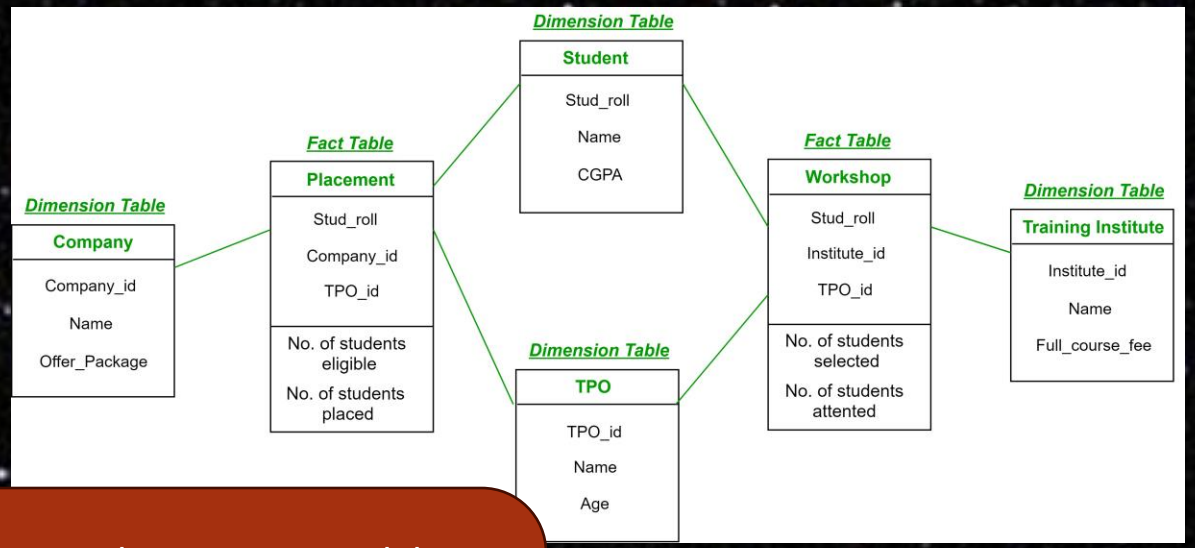

Warehouse (dimension table)
WAREHOUSE_ID | OFFICE NAME | CITY | MANAGER NAME
1 | Texas | Houston | John
2 | Florida | Orlando | Phil
Items (dimension table)
ITEM_ID | ITEM NAME | ITEM COLOR | •••
1 | Sedan | Blue | •••
2 | Truck | Brown | •••
Date (dimension table)
DATE_ID | MONTH | YEAR | QUARTER
202005 | May | 2020 | Q2
Orders (fact table)
VALUE DATA
DIMENSION ATTRIBUTE DATA
ORDER ID | ORDER PROFIT | ORDER QUANTITY | ITEM_ID | WAREHOUSE_ID | EMPLOYEE_ID | DATE_ID | CUSTOMER_ID
101 | $100 | 1 | 1 | 1 | 2 | 202005 | 1
102 | $200 | 2 | 1 | 2 | 1 | 202005 | 2
103 | $200 | 2 | 1 | 1 | 3 | 202005 | 2
104 | $400 | 2 | 2 | 1 | 2 | 202005 | 1
105 | $800 | 4 | 2 | 2 | 3 | 202005 | 2
Employee (dimension table)
EMPLOYEE_ID | NAME | GENDER | OFFICE | PHONE
1 | Joe | M | Texas | •••
2 | Jane | F | Utah | •••
3 | Jill | F | Texas | •••
Customer (dimension table)
CUSTOMER_ID | NAME | ADDRESS | PHONE
1 | Bill | 123 Place | •••
2 | Ben | 456 Street | •••
©2021 TECHTARGET. ALL RIGHTS RESERVED TechTarget

Snowflake Schema



FACT Constellation A.k.a GALAXY schema





- Multiple fact tables share dimension tables
- Can be viewed as a collection of star schemas therefore called the fact constellation or galaxy schema.
- Uses more joins to connect the tables

Key differences

	Star Schema	Snowflake Schema	Galaxy Schema
Elements	Single Fact Table connected to multiple dimension tables with no sub-dimension tables	Single Fact Table connects to multiple dimension tables that connects to multiple sub-dimension tables	Multiple Fact Tables connects to multiple dimension tables.
Normalization	Denormalized	Normalized	Normalized
Number of Dimensions	Multiple dimension tables map to a single Fact Table	Multiple dimension tables map to multiple dimension tables	Multiple dimension tables map to multiple Fact Tables
Data Redundancy	High	Low	Low

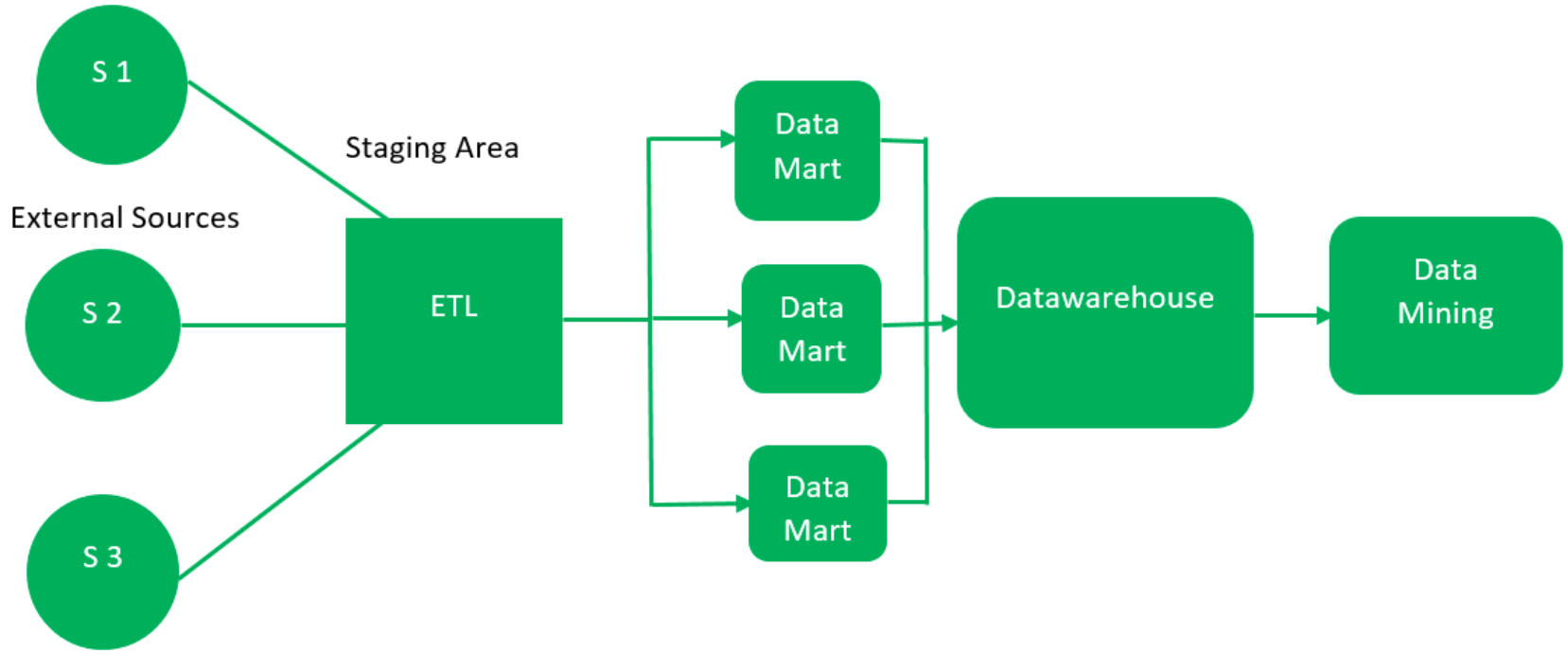
Key differences

	Star Schema	Snowflake Schema	Galaxy Schema
Performance	Fewer foreign keys resulting in increased performance	Decreased performance compared to Star Schema from higher number of foreign keys	Decreased performance compared to Star and Snowflake. Used for complex data aggregation.
Complexity	Simple, designed to be easy to understand	More complicated compared to Star Schema – can be more challenging to understand	Most complicated to understand. Reserved for highly complex data structures
Storage Usage	Higher disk space due to data redundancy	Lower disk space due to limited data redundancy	Low disk space usage compared to the level of sophistication due to the limited data redundancy
Design Limitations	One Fact Table only, no sub-dimensions	One Fact Table only, multiple sub-dimensions are permitted	Multiple Fact Tables permitted, only first level dimensions are permitted

Kimball's approach

Bottom - Up Kimball Approach

- Extract – Transform and Load into Data Marts accessing a single business area.
- Data Marts are integrated into the DW.



4-step Dimensional Design Process

- Step 1: Select the Business Process
- Step 2: Declare the Grain
- Step 3: Identify the Dimensions
- Step 4: Identify the Facts

Step 1: Select the Business Process

- A business process is a low-level activity performed by an organization, such as taking orders, invoicing, receiving payments, handling service calls, registering students, performing a medical procedure, or processing claims.

Step 2: Declare the Grain

- Declaring the grain means specifying exactly what an individual fact table row represents.
- The grain conveys the level of detail associated with the fact table measurements.
- It provides the answer to the question, "How do you describe a single row in the fact table?"
- The grain is determined by the physical realities of the operational system that captures the business process's events.

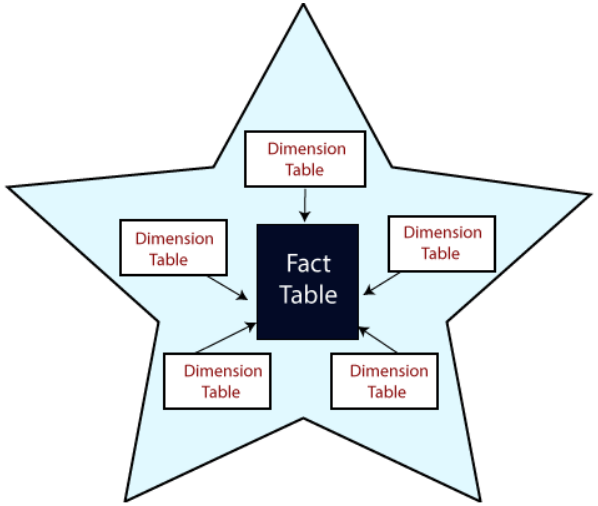
Examples of Grain

Example grain declarations include:

- One row per scan of an individual product on a customer's sales transaction
- One row per line item on a bill from a doctor
- One row per individual boarding pass scanned at an airport gate
- One row per daily snapshot of the inventory levels for each item in a warehouse
- One row per bank account each month

Step 3: Identify the Dimensions

- Dimensions fall out of the question, "*How do business people describe the data resulting from the business process measurement events?*"
- You need to decorate fact tables with a robust set of dimensions representing all possible descriptions that take on single values in the context of each measurement.
- If you are clear about the grain, the dimensions typically can easily be identified as they represent the "*who, what, where, when, why, and how*" associated with the event.
- Examples of common dimensions include **date***, **product**, **customer**, **employee**, and **facility**.



Step 4: Identify the Facts

- Facts are determined by answering the question, *"What is the process measuring?"*
- Business users are keenly interested in analyzing these performance metrics.
- All candidate facts in a design must be true to the grain defined in step 2.
- Facts that clearly belong to a different grain must be in a separate fact table.
- Typical facts are numeric **additive** figures, such as quantity ordered or dollar cost amount.

4-step Dimensional Design Process

- Step 1: Select the Business Process
- Step 2: Declare the Grain
- Step 3: Identify the Dimensions
- Step 4: Identify the Facts

Need to consider both the business users' requirements and the realities of the source data

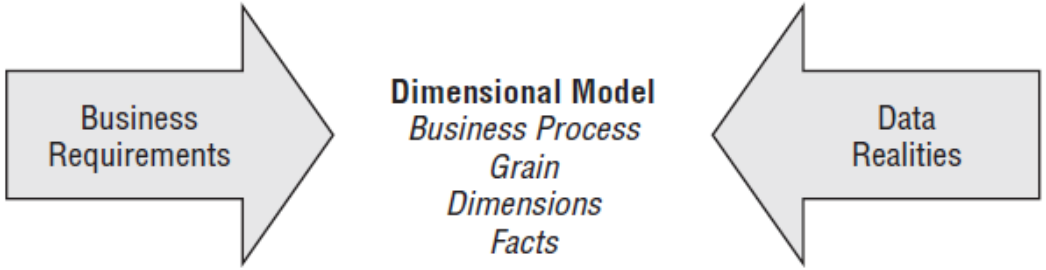


Figure 3-1: Key input to the four-step dimensional design process.

Case Study: Retail

Business Process

- Management wants to better understand customer purchases as captured by the POS system.
- The business process you're modeling is POS retail sales transactions.

Allstar Grocery
123 Loon Street
Green Prairie, MN 55555
(952) 555-1212

Store: 0022
Cashier: 00245409/Alan

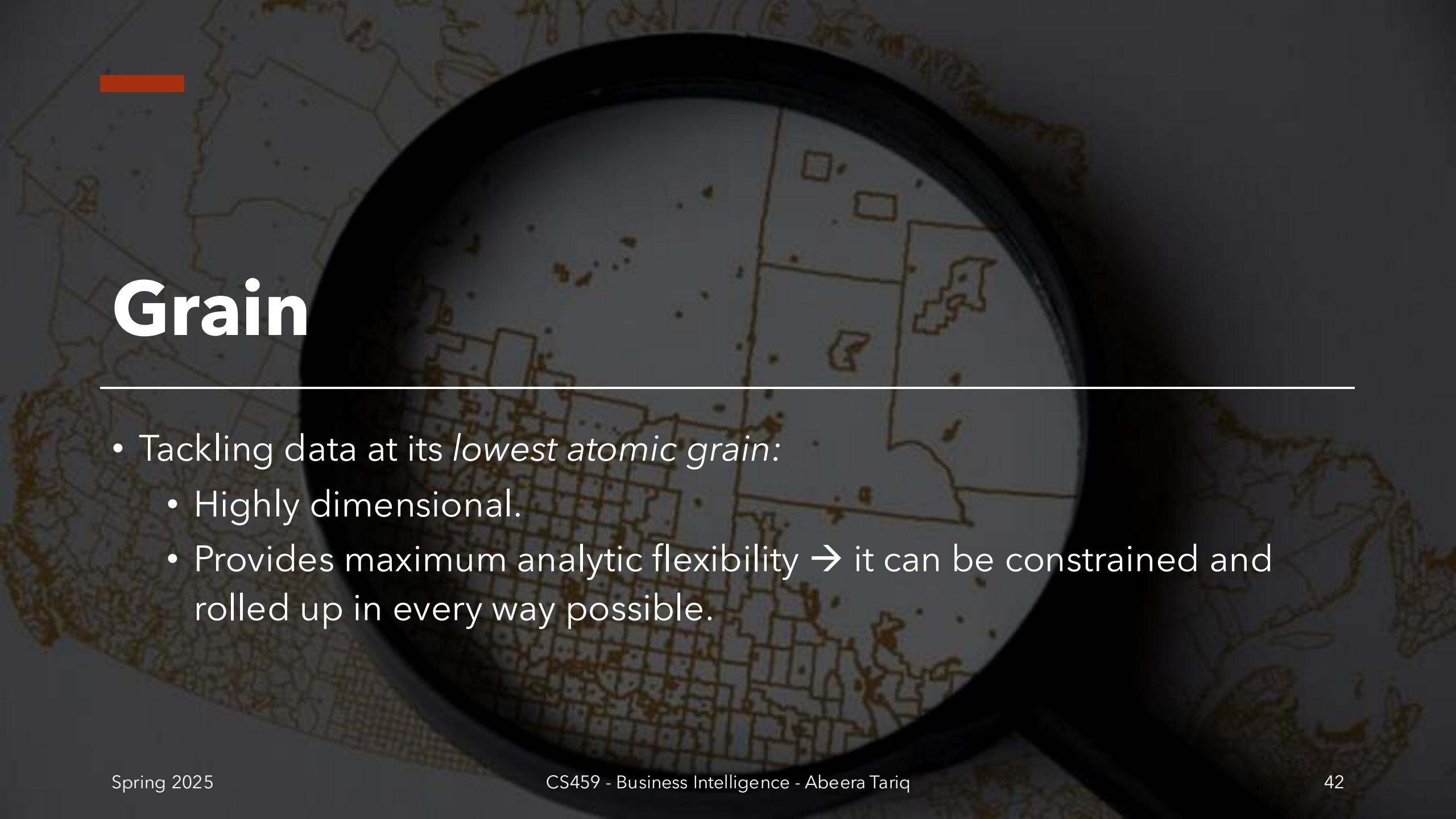
0030503347 Baked Well Multigrain Muffins	2.50
2120201195 Diet Cola 12-pack	4.99
Saved $.50 off $5.49	
0070806048 Sparkly Toothpaste	1.99
Coupon $.30 off $2.29	
2840201912 SoySoy Milk Quart	3.19
TOTAL	12.67
AMOUNT TENDERED	
CASH	12.67
ITEM COUNT: 4	

Transaction: 649 4/15/2013 10:56 AM

Thank you for shopping at Allstar

00649002204152013002454O9

Figure 3-2: Sample cash register receipt.



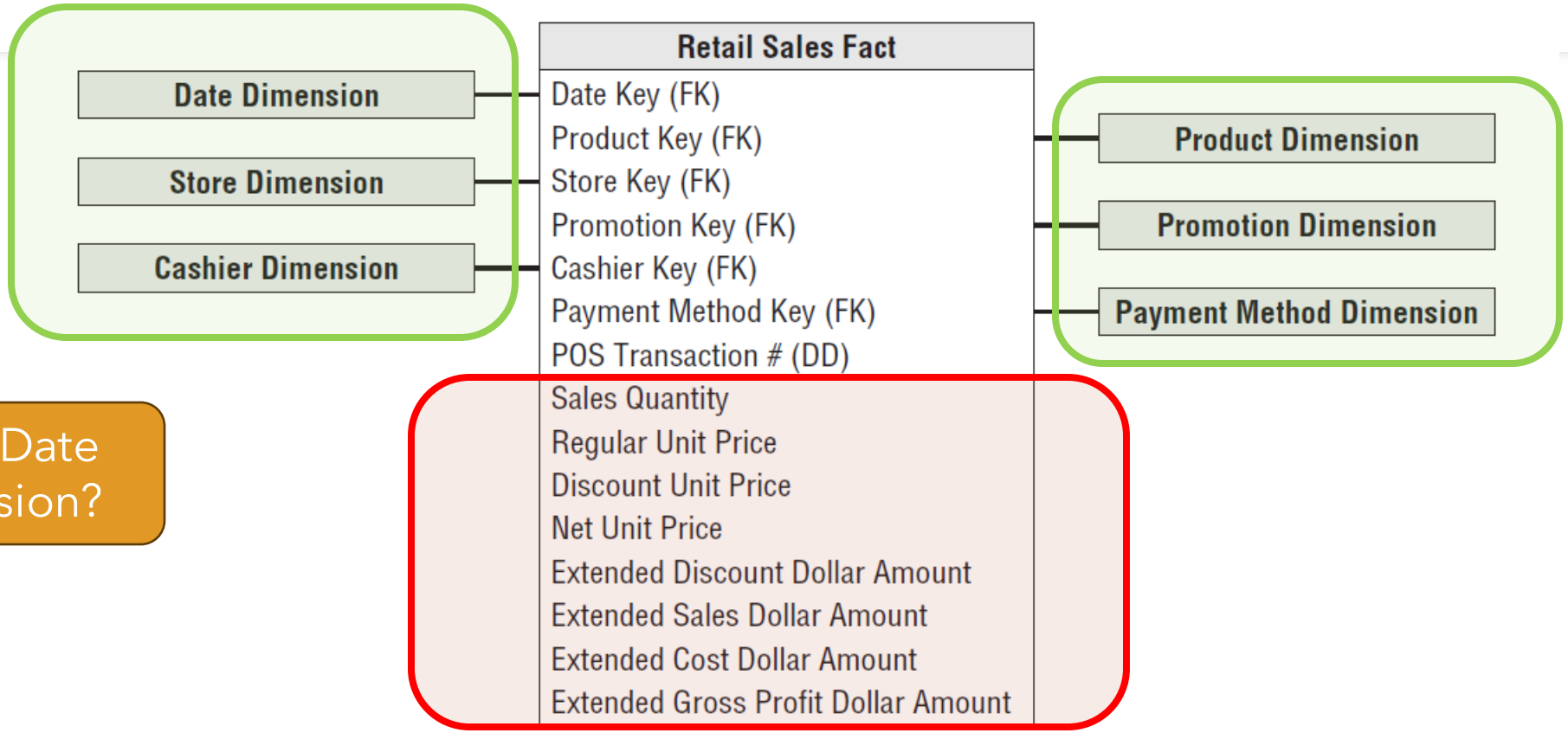
Grain

- Tackling data at its *lowest atomic grain:*
  - Highly dimensional.
  - Provides maximum analytic flexibility → it can be constrained and rolled up in every way possible.

Grain – Retail Case Study

- Most granular data is an individual product on a POS transaction, assuming the POS system rolls up all sales for a given product within a shopping cart into a **single line item**.

Facts and Dimensions

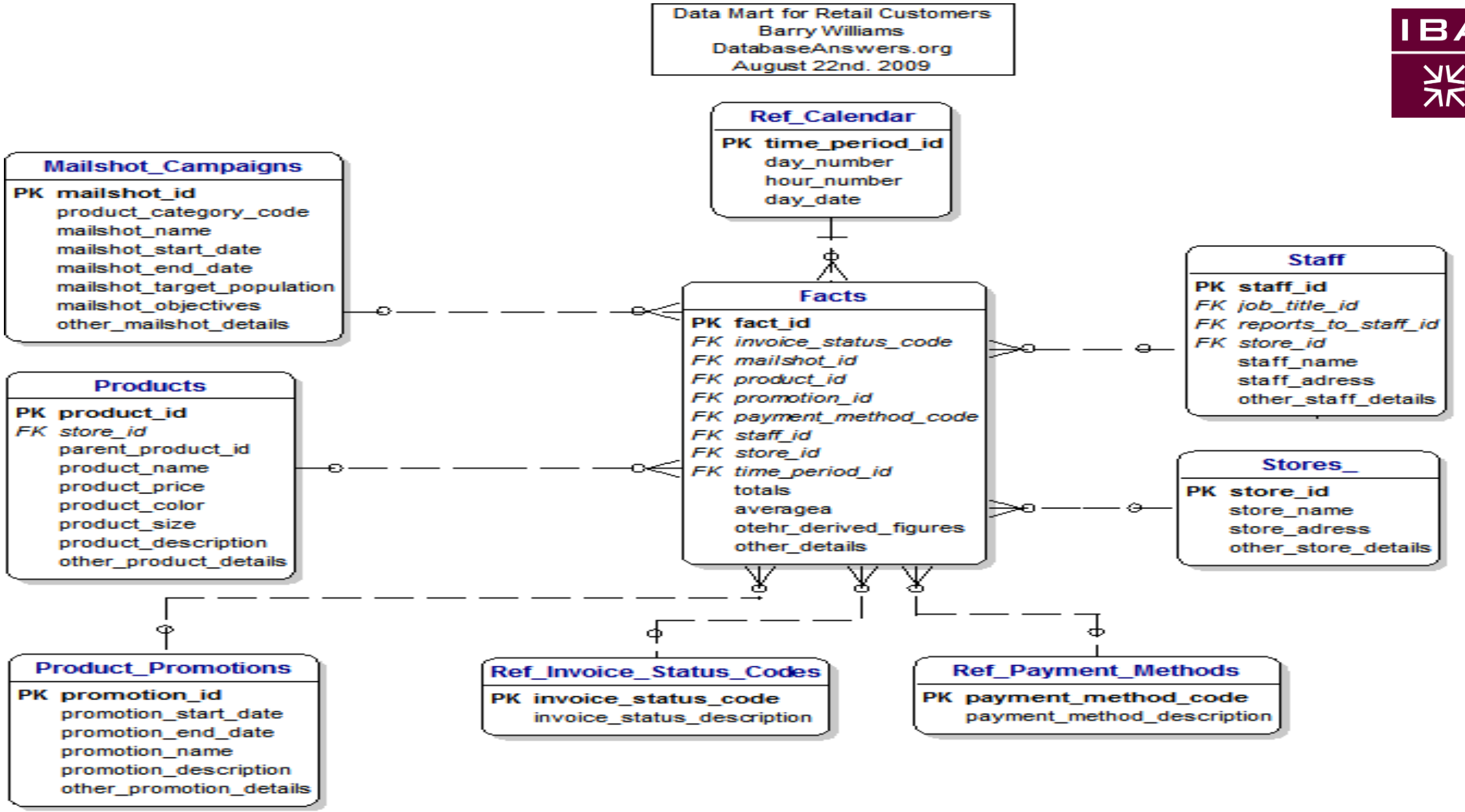


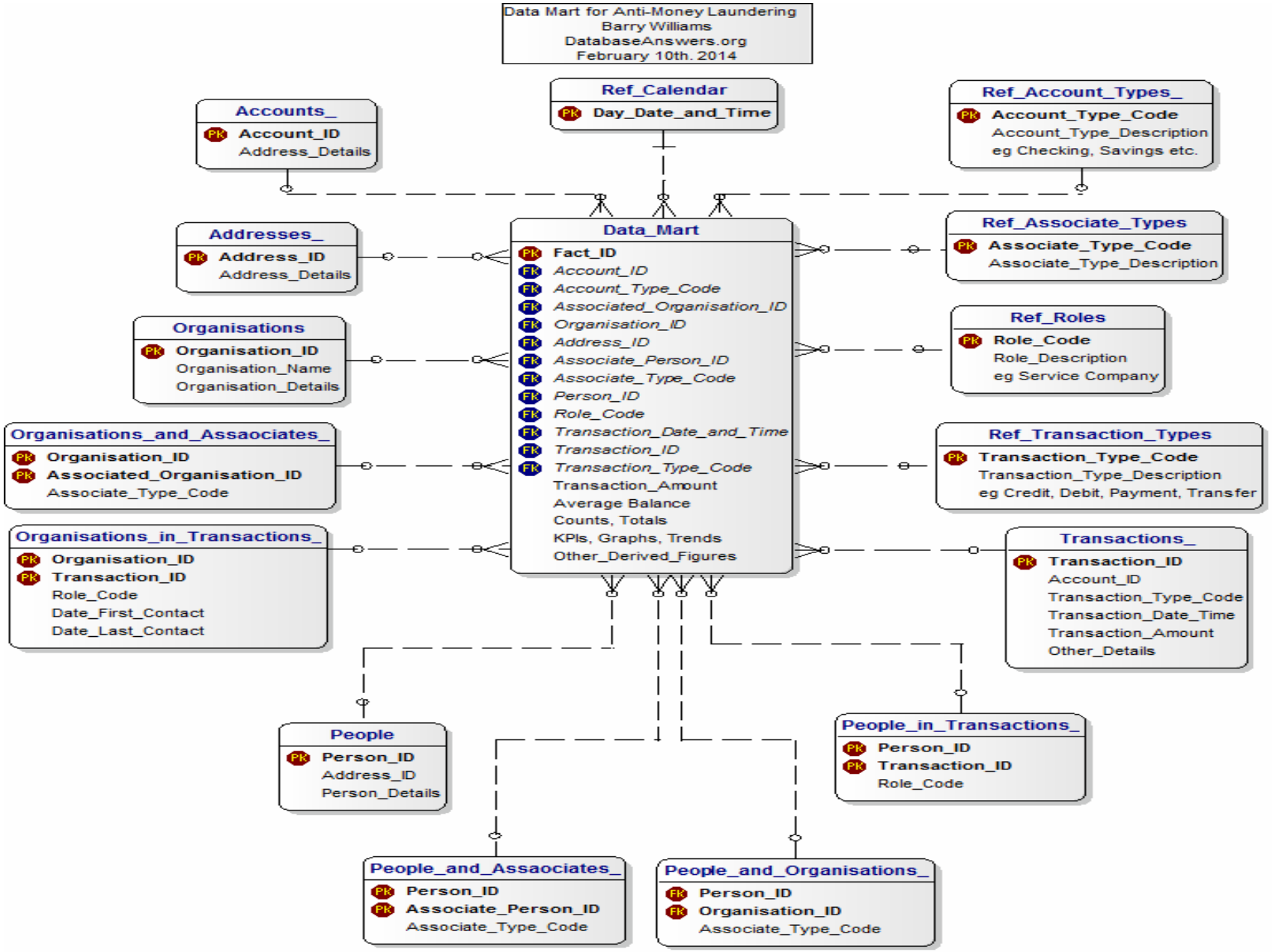
Why a Date Dimension?

Figure 3-3: Measured facts in retail sales schema.

DB to DWH Activity

- Look over your Database Projects from last semester
- Identify a business process
- Specify the grain
- List the potential facts and dimensions
- Sketch a basic star schema (neatly hand drawn or using a tool)
- Maximum teams of 3 – Connect with your DB project team members if they are taking the course (either section)

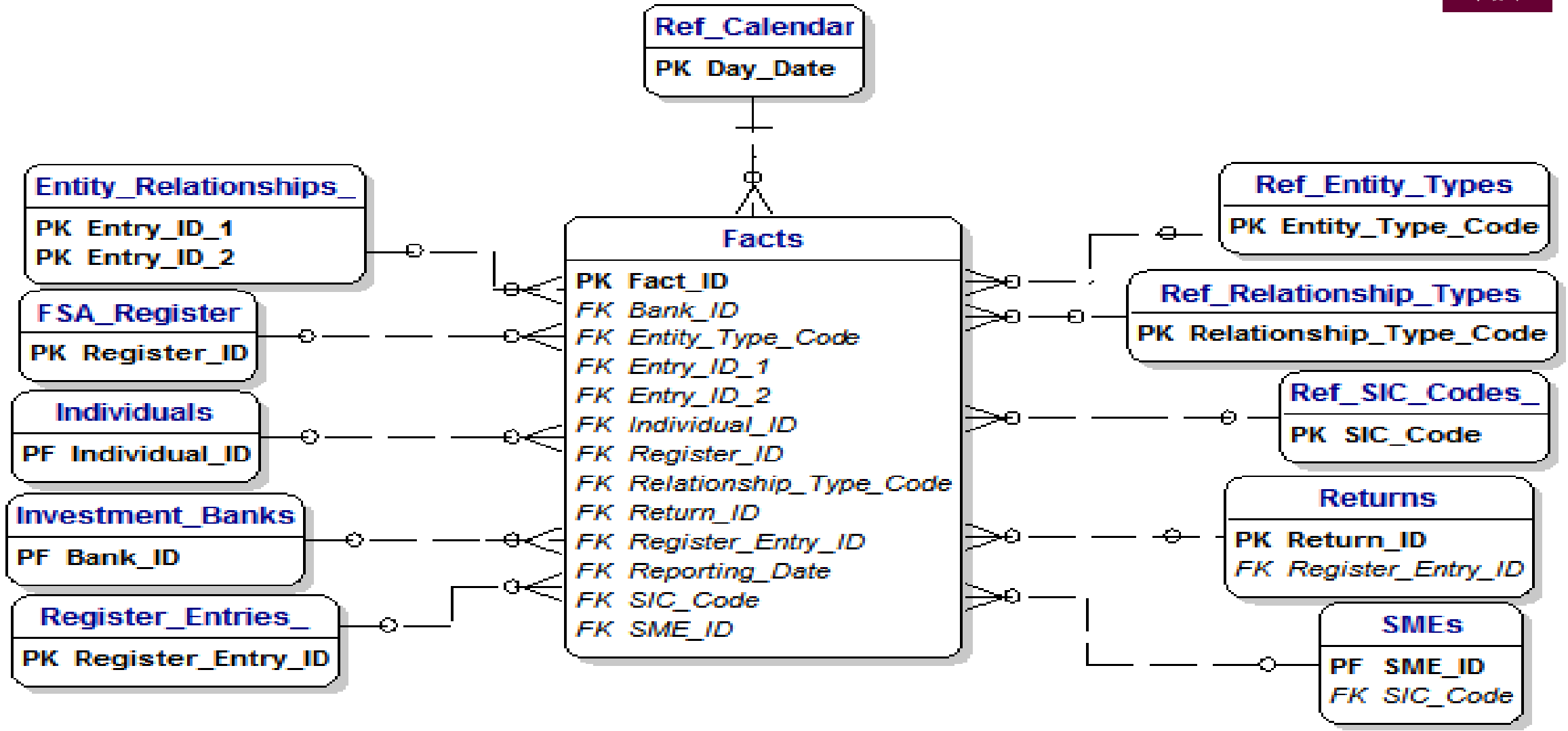

Data Mart for Retail Customers
Barry Williams
DatabaseAnswers.org
August 22nd. 2009
Ref_Calendar
PK time_period_id
day_number
hour_number
day_date
Mailshot_Campaigns
PK mailshot_id
product_category_code
mailshot_name
mailshot_start_date
mailshot_end_date
mailshot_target_population
mailshot_objectives
other_mailshot_details
Facts
PK fact_id
FK invoice_status_code
FK mailshot_id
FK product_id
FK promotion_id
FK payment_method_code
FK staff_id
FK store_id
FK time_period_id
totals
averagea
otehr_derived_figures
other_details
Staff
PK staff_id
FK job_title_id
FK reports_to_staff_id
FK store_id
staff_name
staff_adress
other_staff_details
Products
PK product_id
FK store_id
parent_product_id
product_name
product_price
product_color
product_size
product_description
other_product_details
Stores_
PK store_id
store_name
store_adress
other_store_details
Product_Promotions
PK promotion_id
promotion_start_date
promotion_end_date
promotion_name
promotion_description
other_promotion_details
Ref_Invoice_Status_Codes
PK invoice_status_code
invoice_status_description
Ref_Payment_Methods
PK payment_method_code
payment_method_description

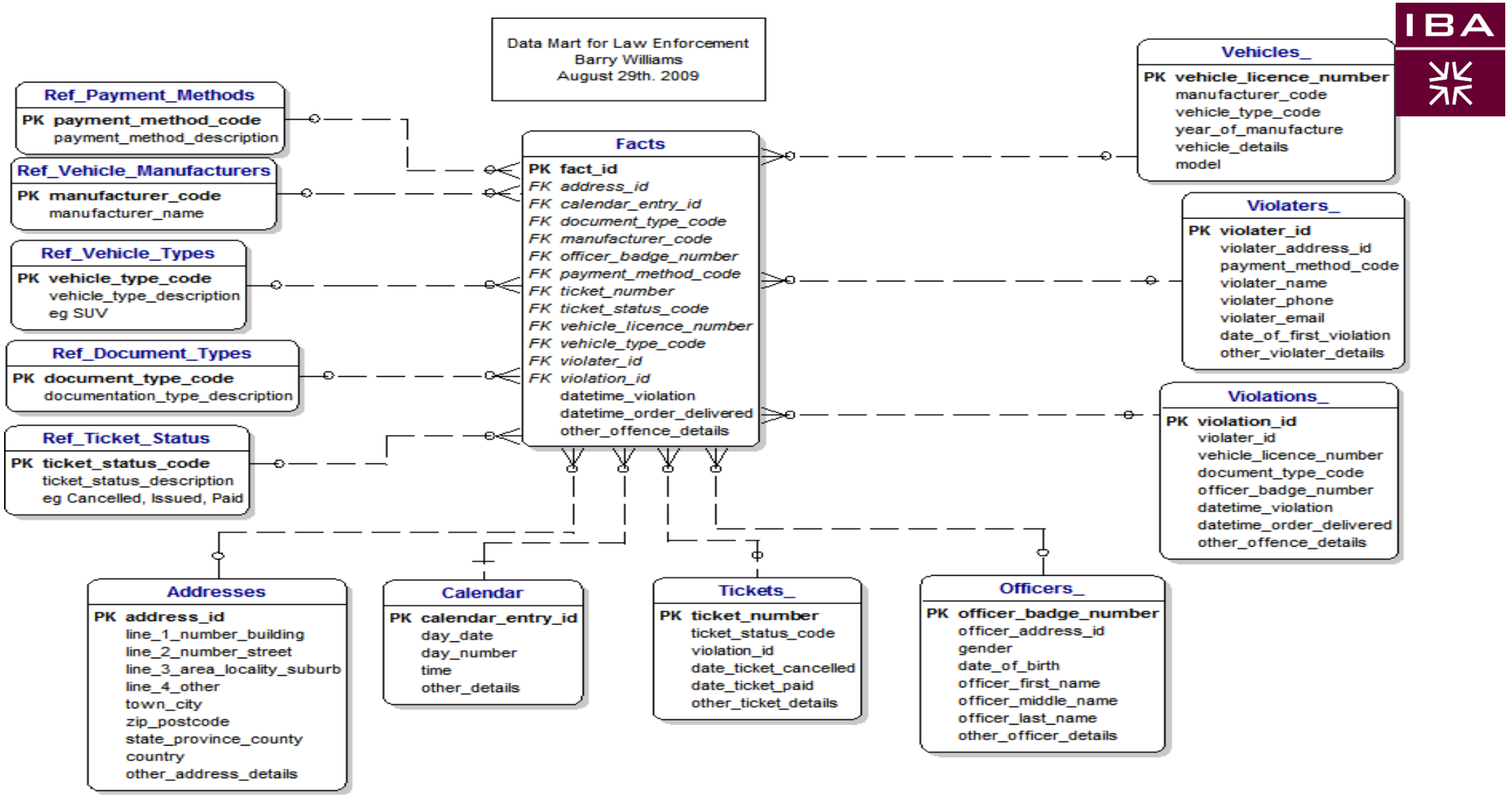


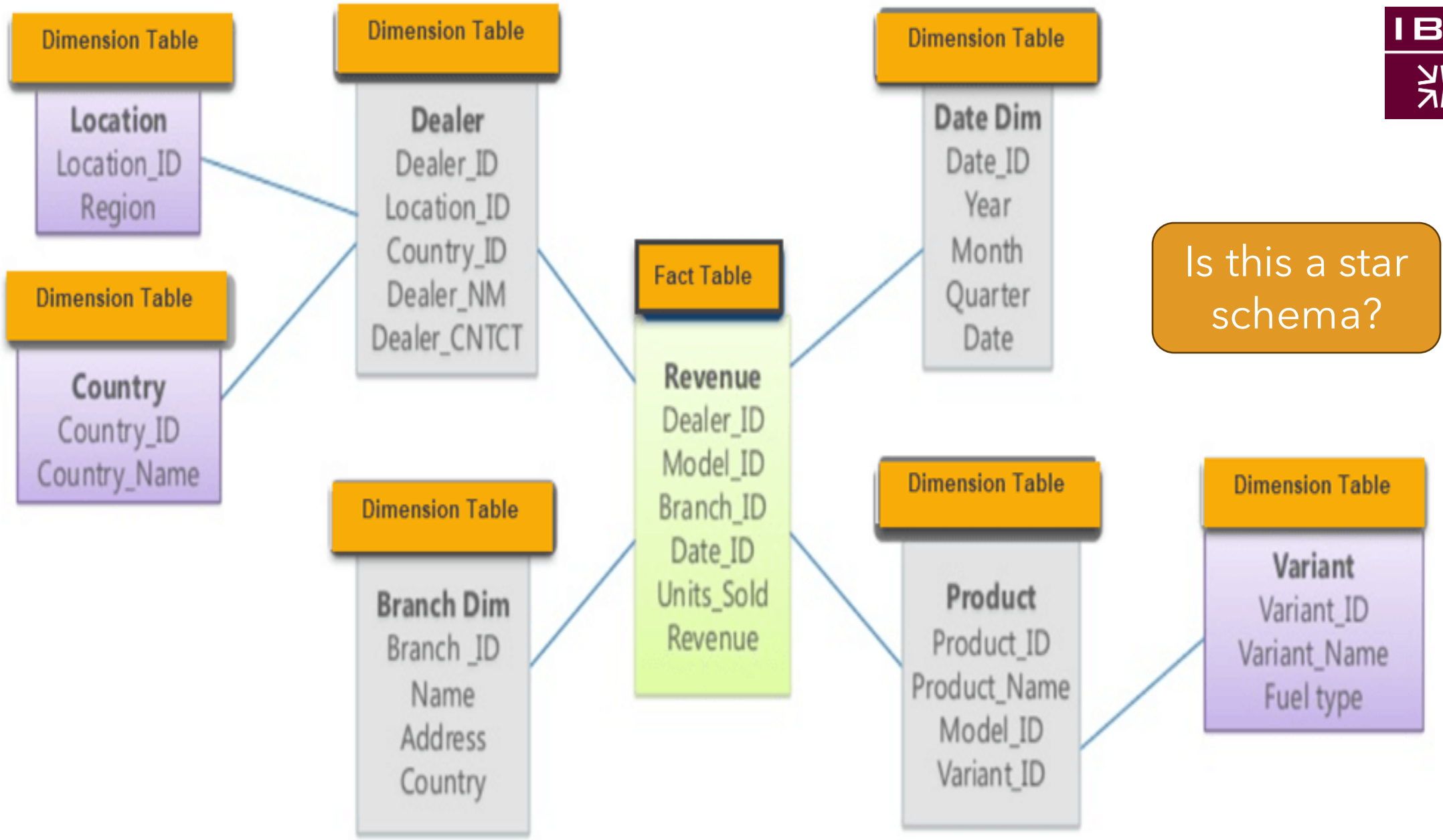
Business Analyses are required ACROSS the Non-Central Tables (dimensions)

The Central Table contains the information for all dimensions collectively, along with required KPIs

Data Mart for the Financial Services Authority
Barry Williams
Database Answers
February 12th. 2011




Data Mart for Law Enforcement
Barry Williams
August 29th. 2009
Ref_Payment_Methods
PK payment_method_code
payment_method_description
Ref_Vehicle_Manufacturers
PK manufacturer_code
manufacturer_name
Ref_Vehicle_Types
PK vehicle_type_code
vehicle_type_description
eg SUV
Ref_Document_Types
PK document_type_code
documentation_type_description
Ref_Ticket_Status
PK ticket_status_code
ticket_status_description
eg Cancelled, Issued, Paid
Facts
PK fact_id
FK address_id
FK calendar_entry_id
FK document_type_code
FK manufacturer_code
FK officer_badge_number
FK payment_method_code
FK ticket_number
FK ticket_status_code
FK vehicle_licence_number
FK vehicle_type_code
FK violater_id
FK violation_id
datetime_violation
datetime_order_delivered
other_offence_details
Vehicles_
PK vehicle_licence_number
manufacturer_code
vehicle_type_code
year_of_manufacture
vehicle_details
model
Violaters_
PK violater_id
violater_address_id
payment_method_code
violater_name
violater_phone
violater_email
date_of_first_violation
other_violater_details
Violations_
PK violation_id
violater_id
vehicle_licence_number
document_type_code
officer_badge_number
datetime_violation
datetime_order_delivered
other_offence_details
Addresses
PK address_id
line_1_number_building
line_2_number_street
line_3_area_locality_suburb
line_4_other
town_city
zip_postcode
state_province_county
country
other_address_details
Calendar
PK calendar_entry_id
day_date
day_number
time
other_details
Tickets_
PK ticket_number
ticket_status_code
violation_id
date_ticket_cancelled
date_ticket_paid
other_ticket_details
Officers_
PK officer_badge_number
officer_address_id
gender
date_of_birth
officer_first_name
officer_middle_name
officer_last_name
other_officer_details


Dimension Table
Location
Location_ID
Region
Dimension Table
Dealer
Dealer_ID
Location_ID
Country_ID
Dealer_NM
Dealer_CNTCT
Dimension Table
Date Dim
Date_ID
Year
Month
Quarter
Date
Dimension Table
Country
Country_ID
Country_Name
Fact Table
Revenue
Dealer_ID
Model_ID
Branch_ID
Date_ID
Units_Sold
Revenue
Dimension Table
Branch Dim
Branch _ID
Name
Address
Country
Dimension Table
Product
Product_ID
Product_Name
Model_ID
Variant_ID
Dimension Table
Variant
Variant_ID
Variant_Name
Fuel type
Is this a star schema?