INFO20003 Database Systems

Assignment Project Exam Help

https://powcoder.com

Dr Renata Borovica-Gajic

Add WeChat powcoder

Lecture 19

Data Warehousing

Week 10

Coverage

By the end of this class you should be able to:

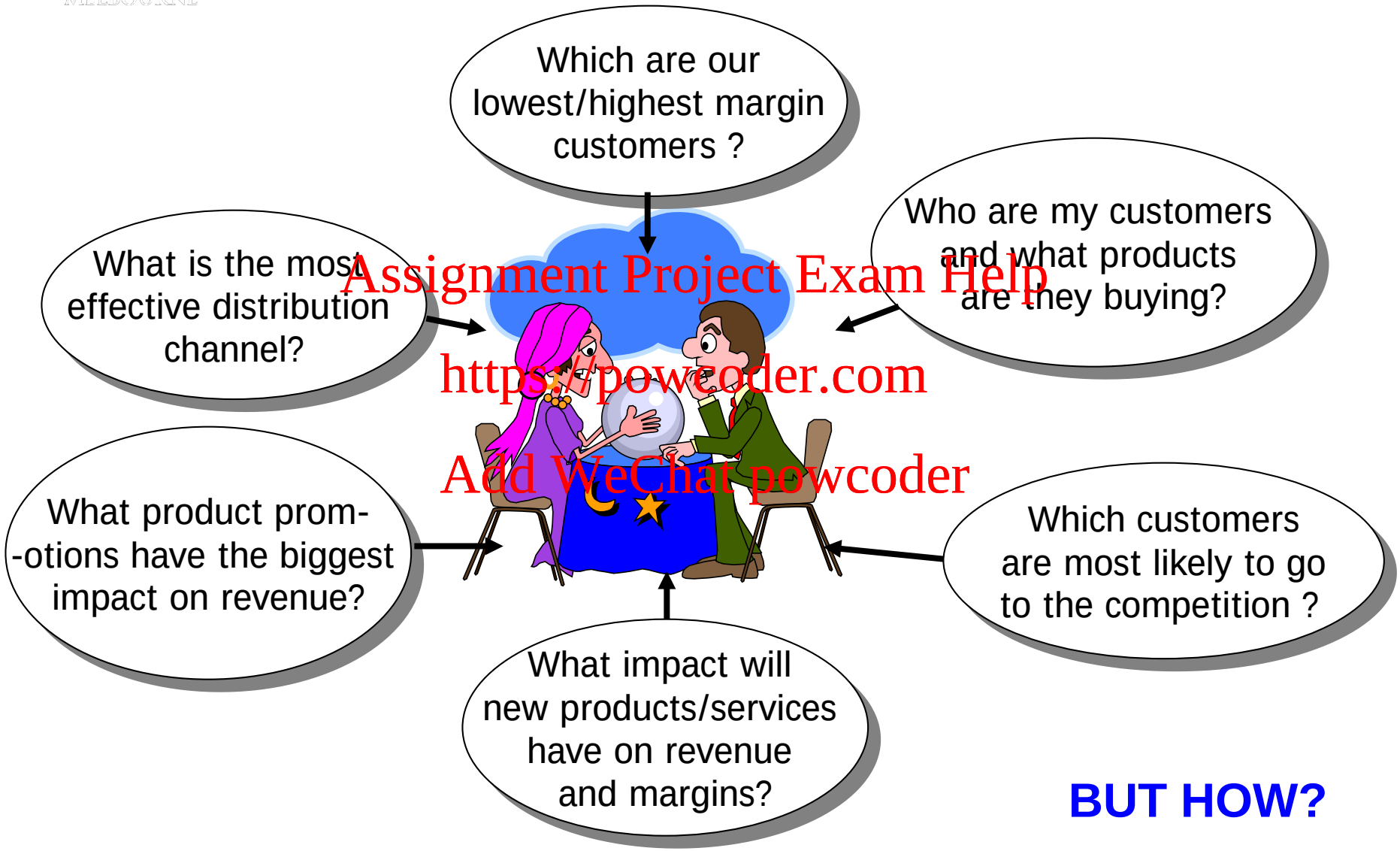
- Articulate the differences between **transactional** (operational) and **informational** (dimensional) databases
- Explain the **characteristics of a DW**
- Understand and explain the **overall architecture of a DW**
- Design **Star Schemas**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Part 1: Introduction to Data Warehousing

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Motivations: A manager wants to know….

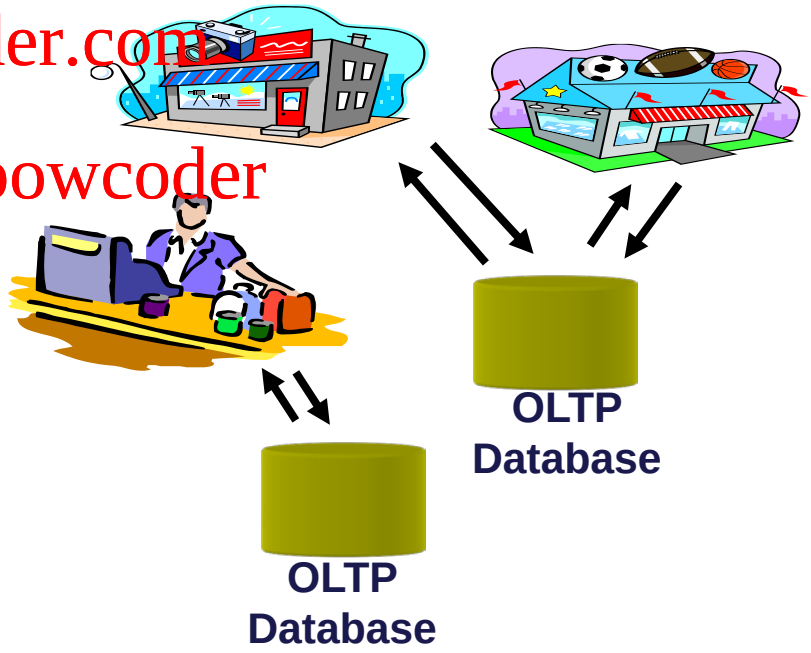


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Relational Databases for Operational Processing

- Used to run day to day business operations
- Automation of routine business processes
  - Accounting
  - Inventory
  - Purchasing
  - Sales
- Created huge efficiencies

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



© University of Melbourne

Databases are great, BUT for business…

- Too many of them
  - Everybody wanted one, or two, or more
  - Production, Marketing, Sales, Accounting …
- Everybody got what was best for them
  - IBM, Oracle, Access, Microsoft
- Eventually this re-created the problem databases were meant to solve
  - Duplicated data
  - Inaccessible data
  - Inconsistent data

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

But data is useful for analysis and decision making

What can we do about it?

- Need an integrated way of getting the ENTIRE organisational data

- Its really an Informational Database, rather than a Transactional Database
  - A single database that allows *all* of the organisations data to be stored in a form that can be used to support organisational decision processes

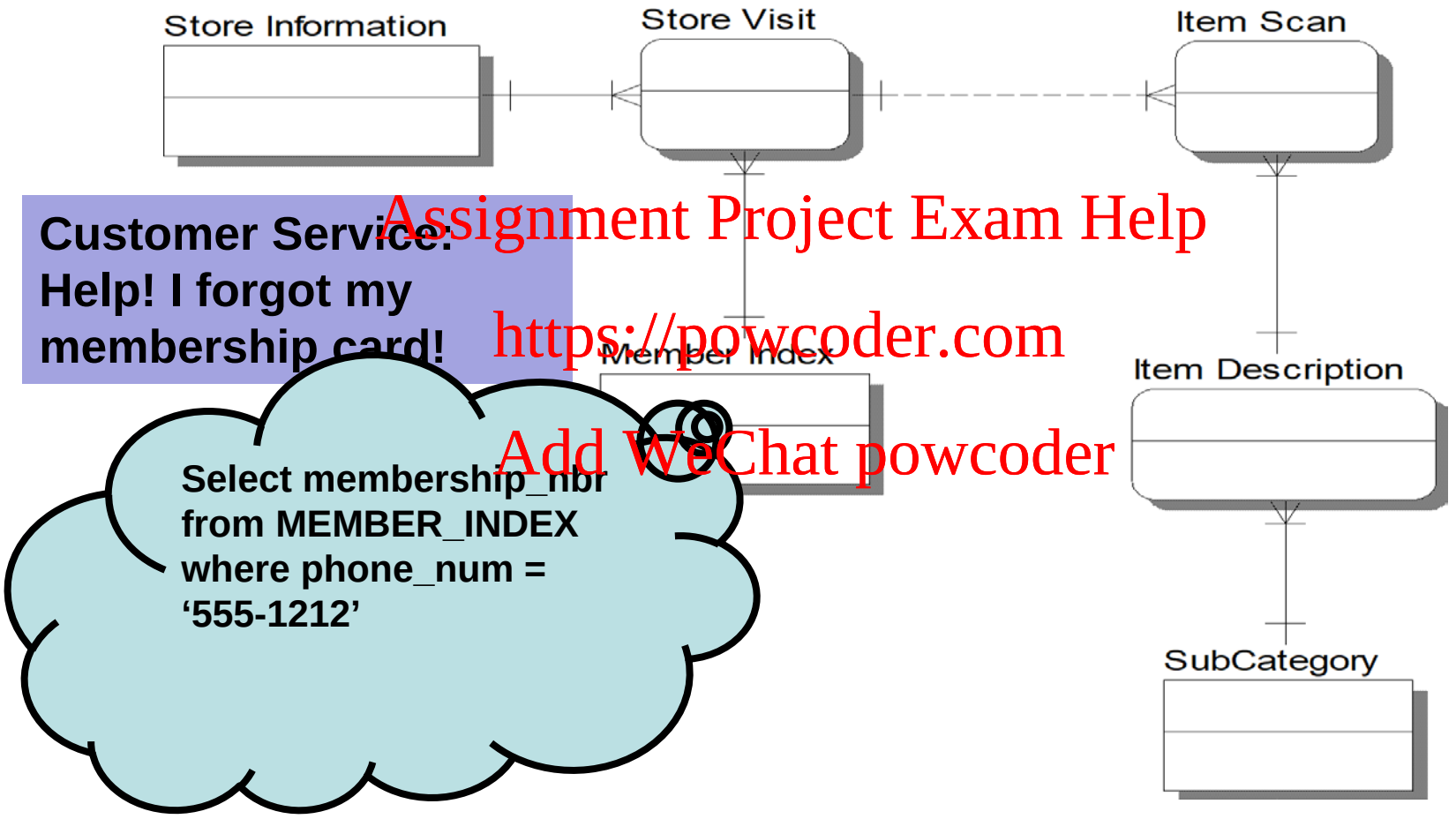
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Warehouse: An Informational Database

- **Data Warehouse:**
  - A single repository of organisational data
  - Integrates data from multiple sources
    - Extracts data from source systems, transforms, loads into the warehouse
  - Makes data available to managers/users
  - Supports analysis and decision-making
- Involve a large data store (often several Terabytes, Petabytes of data)

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

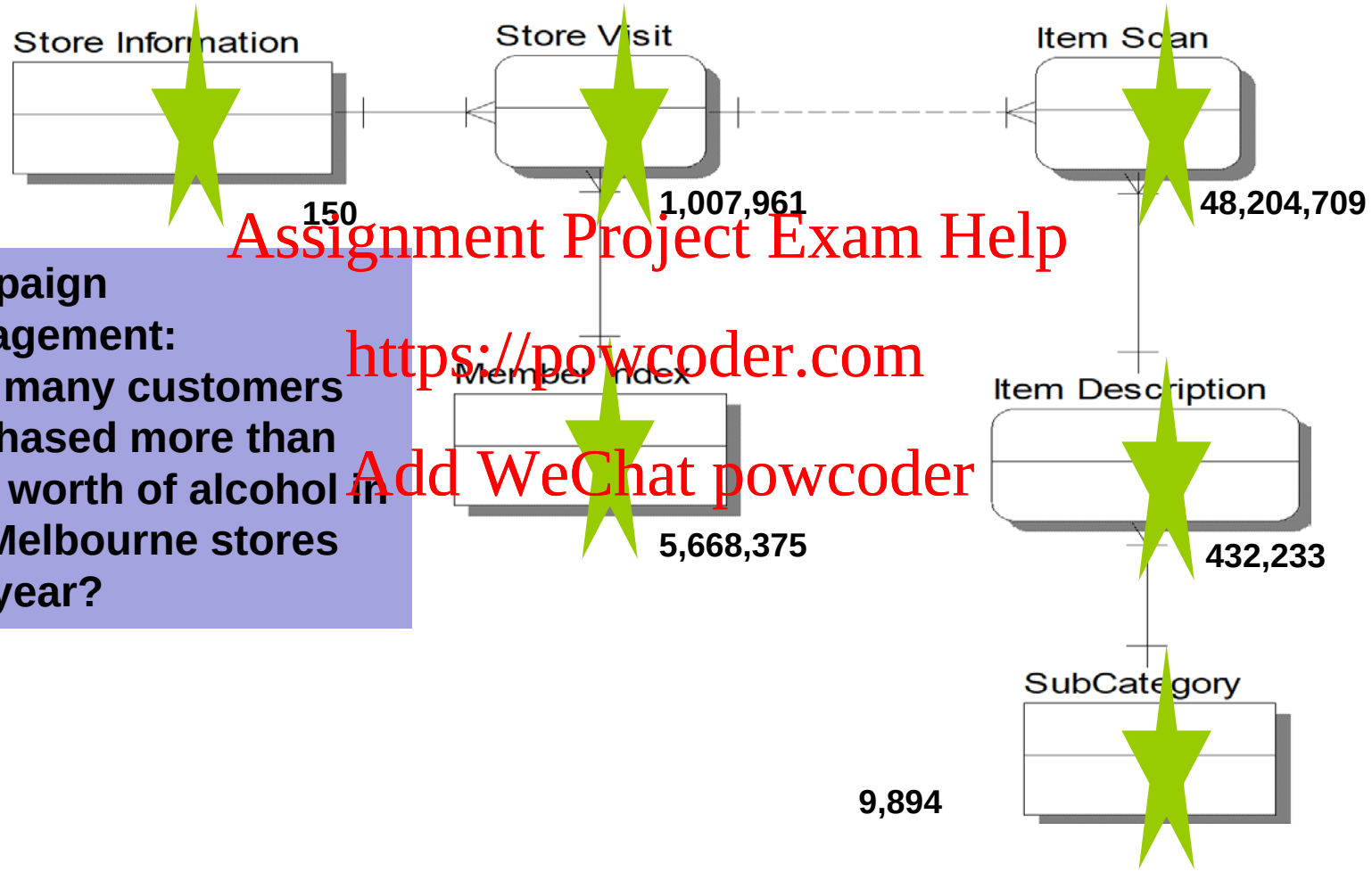
Transactional (Operational) Questions



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

© University of Melbourne

Analytical Questions



Campaign Management:
How many customers purchased more than $500 worth of alcohol in our Melbourne stores this year?

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

© University of Melbourne

DW Supports Analytical Queries

- One is interested in numerical *aggregations*
  - How **many**?
  - What is the **average**?
  - What is the **total cost**?
- One is interested in understanding *dimensions*
  - Sales **by state by customer type**
  - Sales **by product by store by quarter**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

DW will help answer these questions

© University of Melbourne

Characteristics of a DW

- Subject oriented
  - Data warehouses are organised around particular subjects (sales, customers, products)
- Validated, Integrated data
  - Data from different systems converted to a common format: allows comparison and consolidation of data from different sources
  - Data from various sources validated before storing it in a data warehouse

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

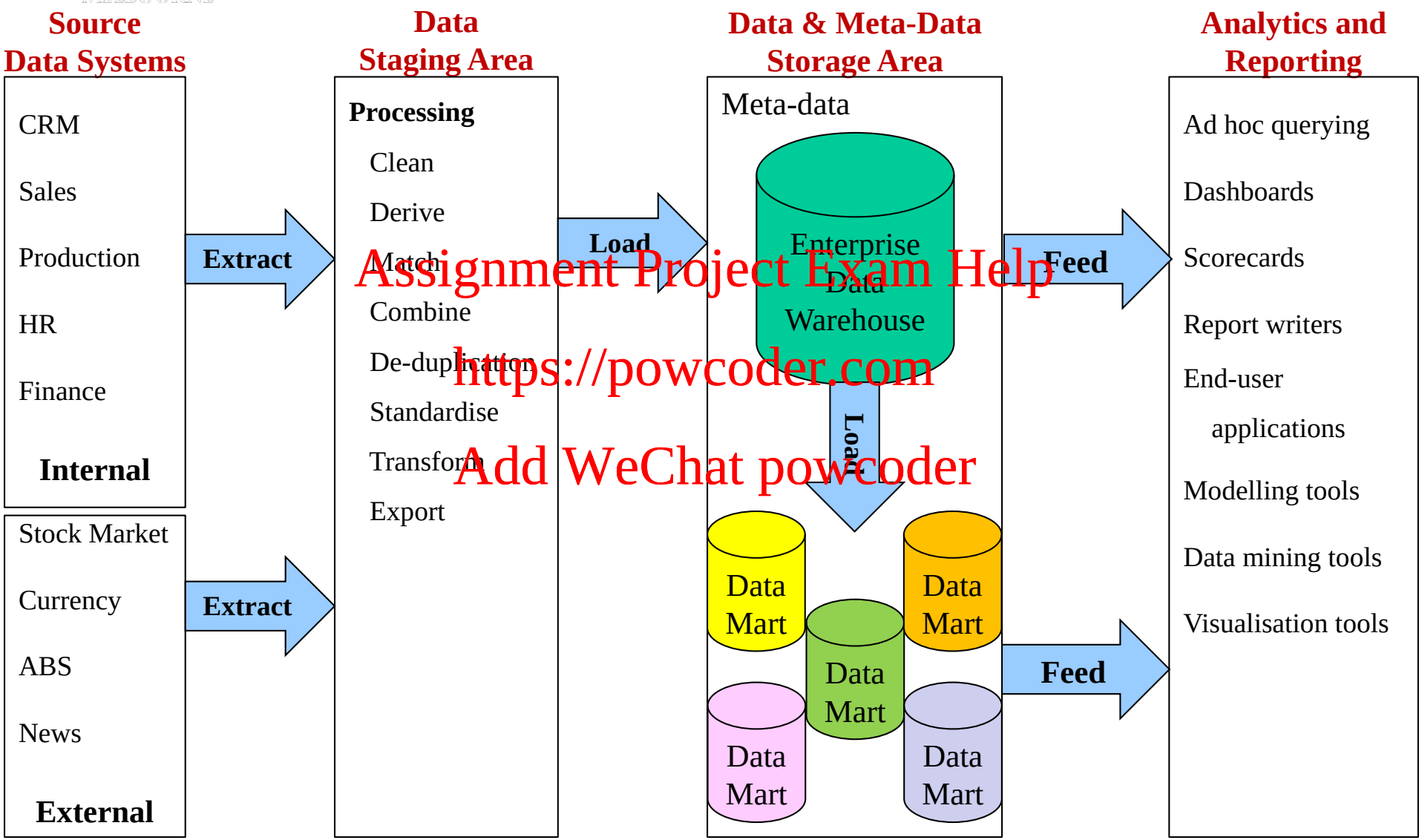
© University of Melbourne

Characteristics of a DW

- Time variant
  - Historical data
  - Trend analysis crucial for decision support: requires historical data
  - Data consists of a series of “snapshots” which are time stamped
- Non-volatile
  - Users have read access only – all updating done automatically by ETL process and periodically by a DBA

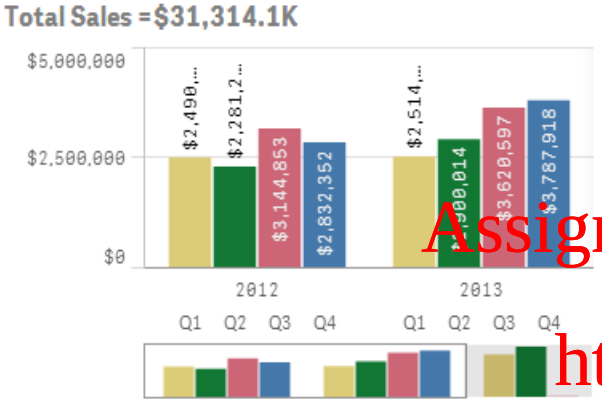
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

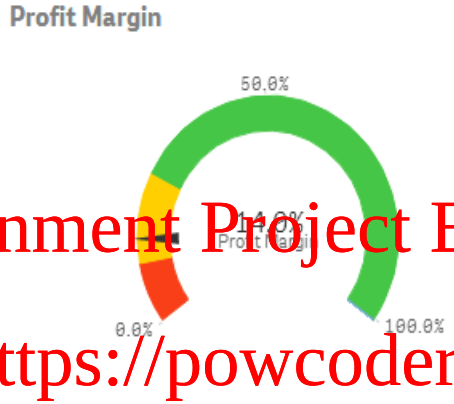
A DW Architecture

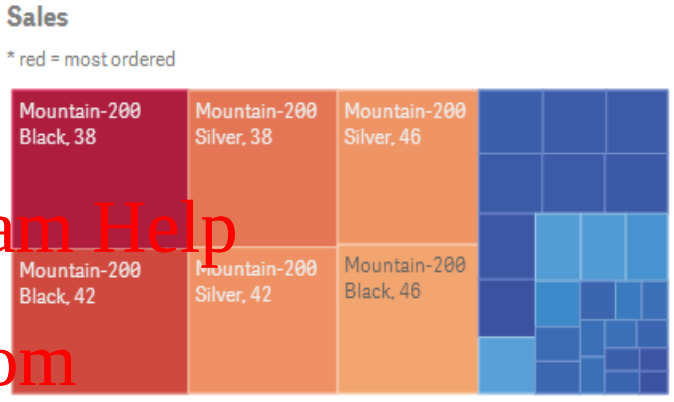


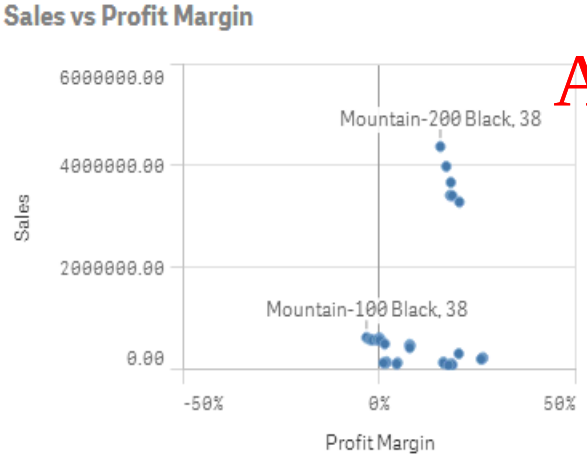
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

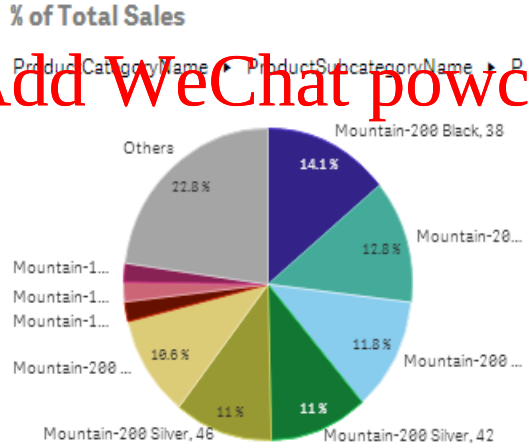
Business Intelligence Dashboard

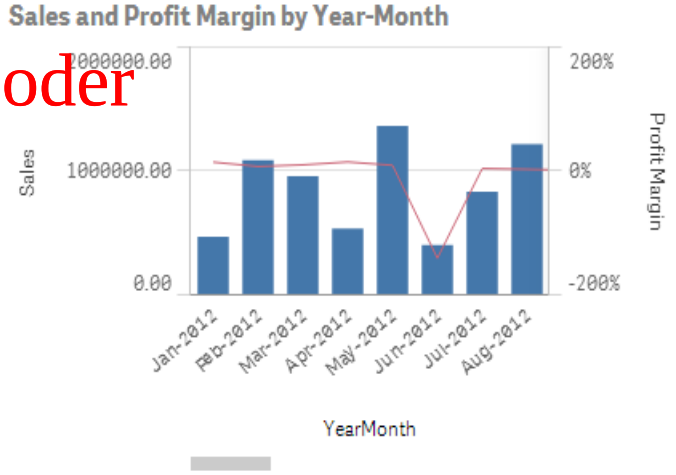








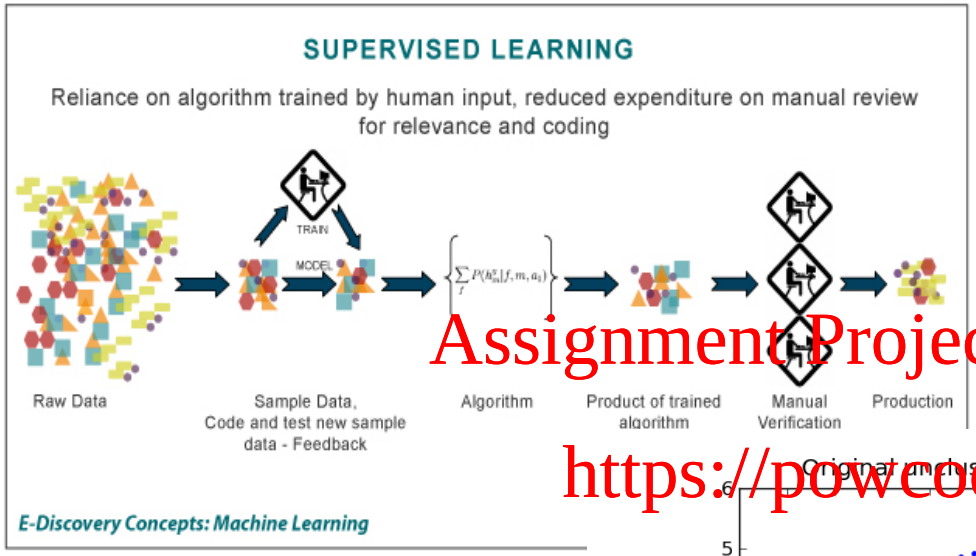




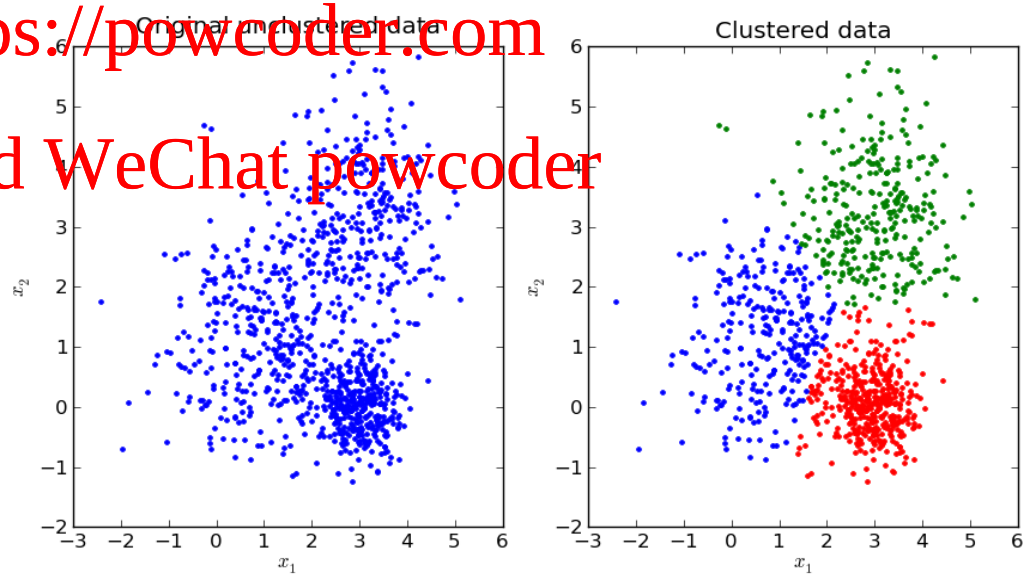
Qlik

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

DW Supports Advanced Analytics



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



http://us.hudson.com/legal/blog/postid/513/predictive-analytics-artificial-intelligence-science-fiction-e-discovery-truth
http://pypr.sourceforge.net/kmeans.html

© University of Melbourne

Part 2- Dimensional Modelling

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Business Analyst World

- How much **revenue** did the **product G** generate in the **last three months**, broken down by month for the south eastern sales **region**, by individual **stores**, broken down by **promotions**, compared to estimates and to the previous version of the product
  - Analysis starts usually with a single indication of something strange, then goes deep into the data, left to a new dimension, right to another, up to the summary, back down and left and right again, until the problem is identified…
  - Dimensional Analysis: To support business analysts view
    - Revenue per product per customer per location?

Fact Dimension Dimension Dimension

Fact Dimension Dimension
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Introduction to Dimensional Modelling

- Popularised by Ralph Kimball in the 1990s
- Based on the *multi-dimensional* model of data and designed for retrieval-only databases
- Very simple, intuitive, and easily understood structure
- Also known as *star schema* design

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Dimensional Modelling

- A dimensional model consists of:
  - Fact table
  - Several dimensional tables
  - (Sometimes) hierarchies in the dimensions
- Essentially a simple and restricted type of ER model

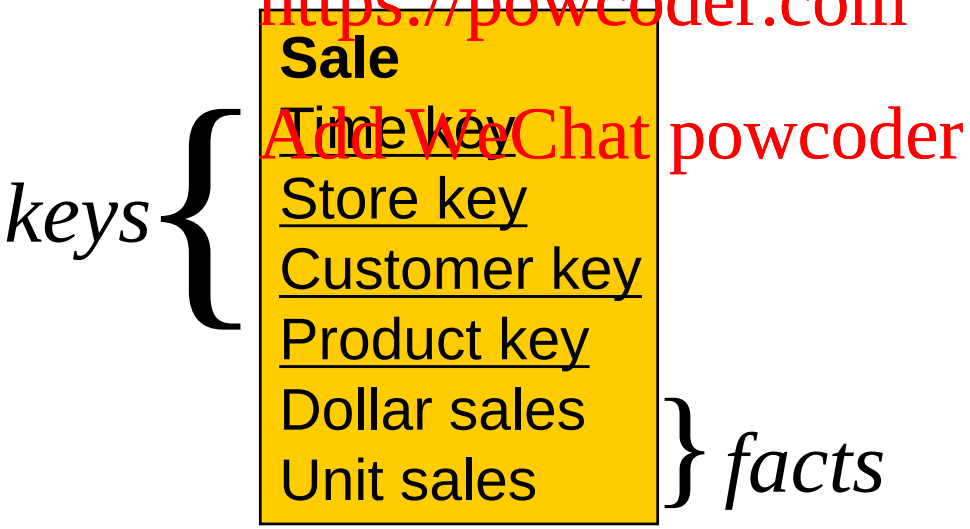
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

© University of Melbourne

Fact Table

- A fact table contains the actual business measures (additive, aggregates), called *facts*
- The fact table also contains *foreign keys* pointing to *dimensions*

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Fact Table - example

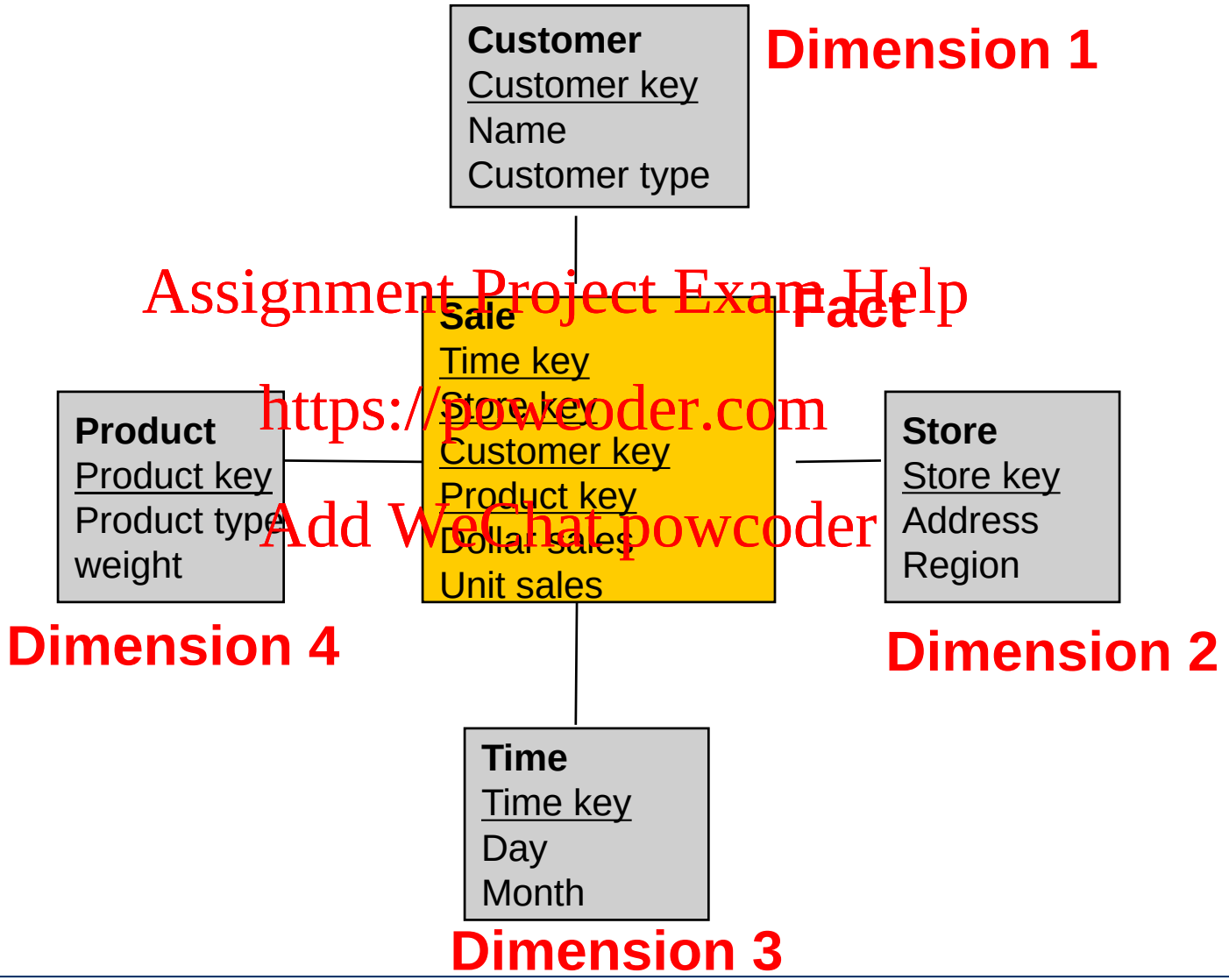
- Actual data might look like this
- Granularity, or level of detail, is a key issue
  - Finest level of detail for a fact table, determined by the finest level of each dimension

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Time-id	*Store-id*	*Cust-id*	*Prod-id*	*Dollar sales*	*Unit Sales*
T100	S303	C101	P98	$120,000	5,000
T101	S303	C256	P98	$240000	10,000
T102	S387	C101	P10	$456,000	27,899
T100	S234	C400	P56	$100,200	5,600

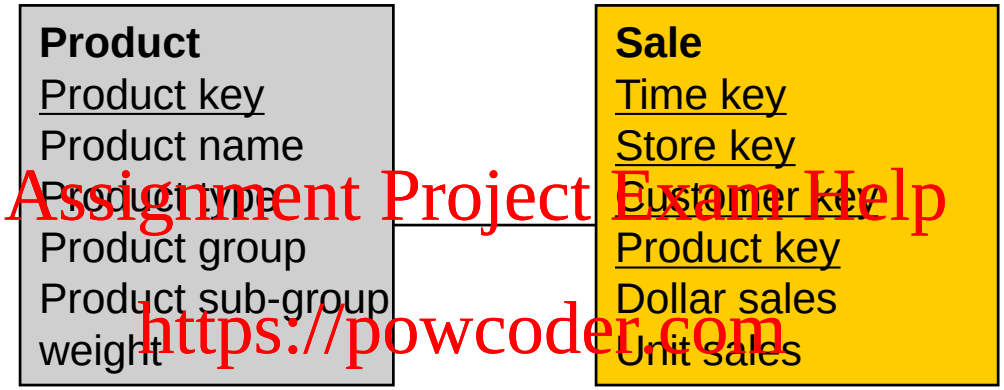
© University of Melbourne

Star schema – dimensional model



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Dimension Hierarchies



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Product name e.g. Hammer
- Product type e.g. Tool
  - Product group e.g. Hardware

Dimension Table - example

- Captures a factor by which a fact can be described or classified
- Actual data might look like this
- Hierarchy evident in data

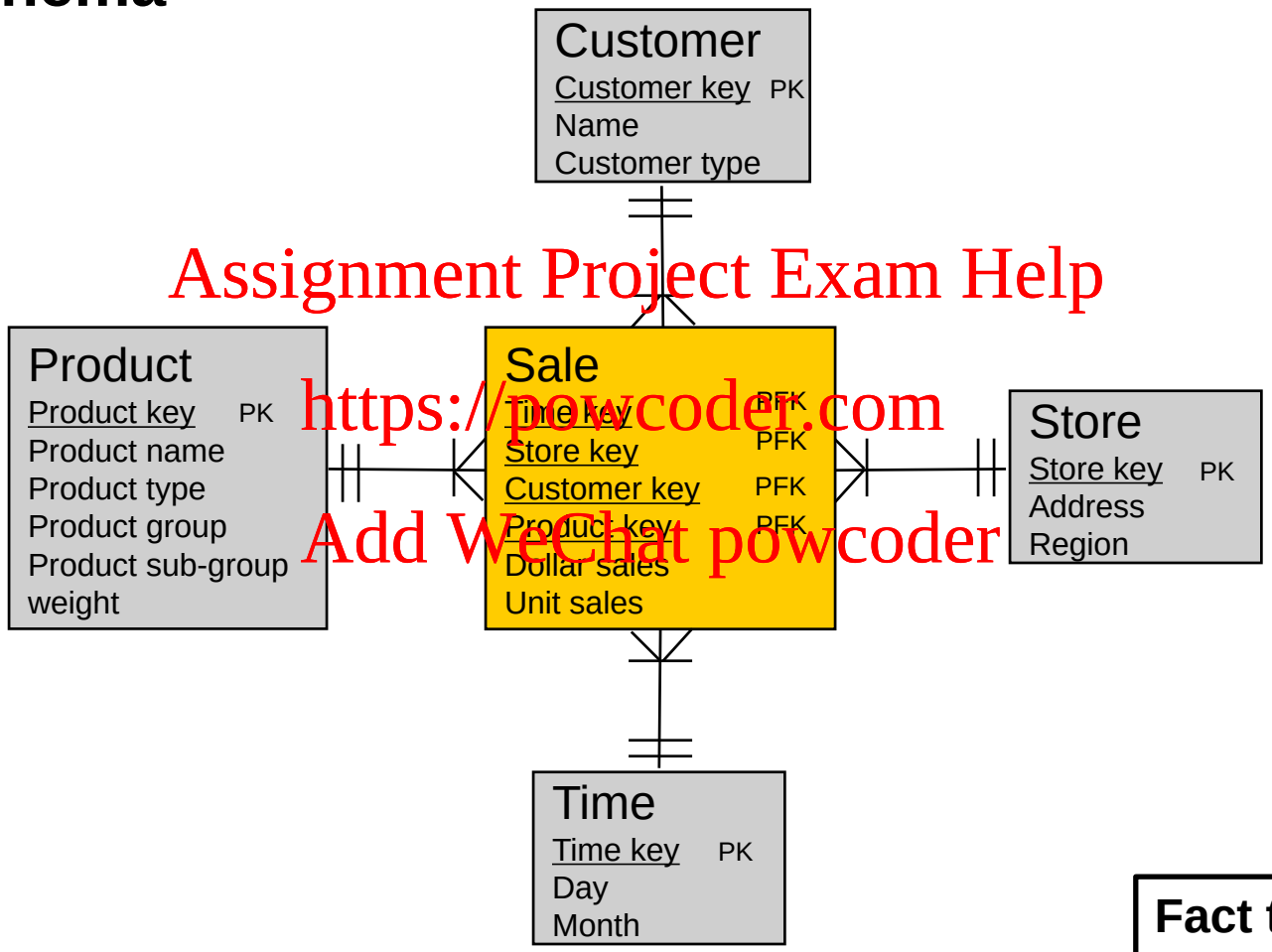
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Prod-id	*Prod-Name*	*Prod-Group*	*Prod-Subgroup*	*Weight*
P10	Hammer	Hardware	Tool	5kg
P56	10cm Nails	Hardware	Nails	1kg
P98	Plastic Pipe	Plumbing	Pipe	1kg

© University of Melbourne

Dimensional model as an ER model

Star schema



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Fact table is an intersection table

© University of Melbourne

Designing a Dimensional Model

Steps:

1. Choose a Business Process
2. Choose the measured facts (usually numeric, additive quantities)
3. Choose the granularity of the fact table
4. Choose the dimensions
5. Complete the dimension tables

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

(Kimball, 1996)

© University of Melbourne

Embedded Hierarchies in Dimensional Tables

Assignment Project Exam Help

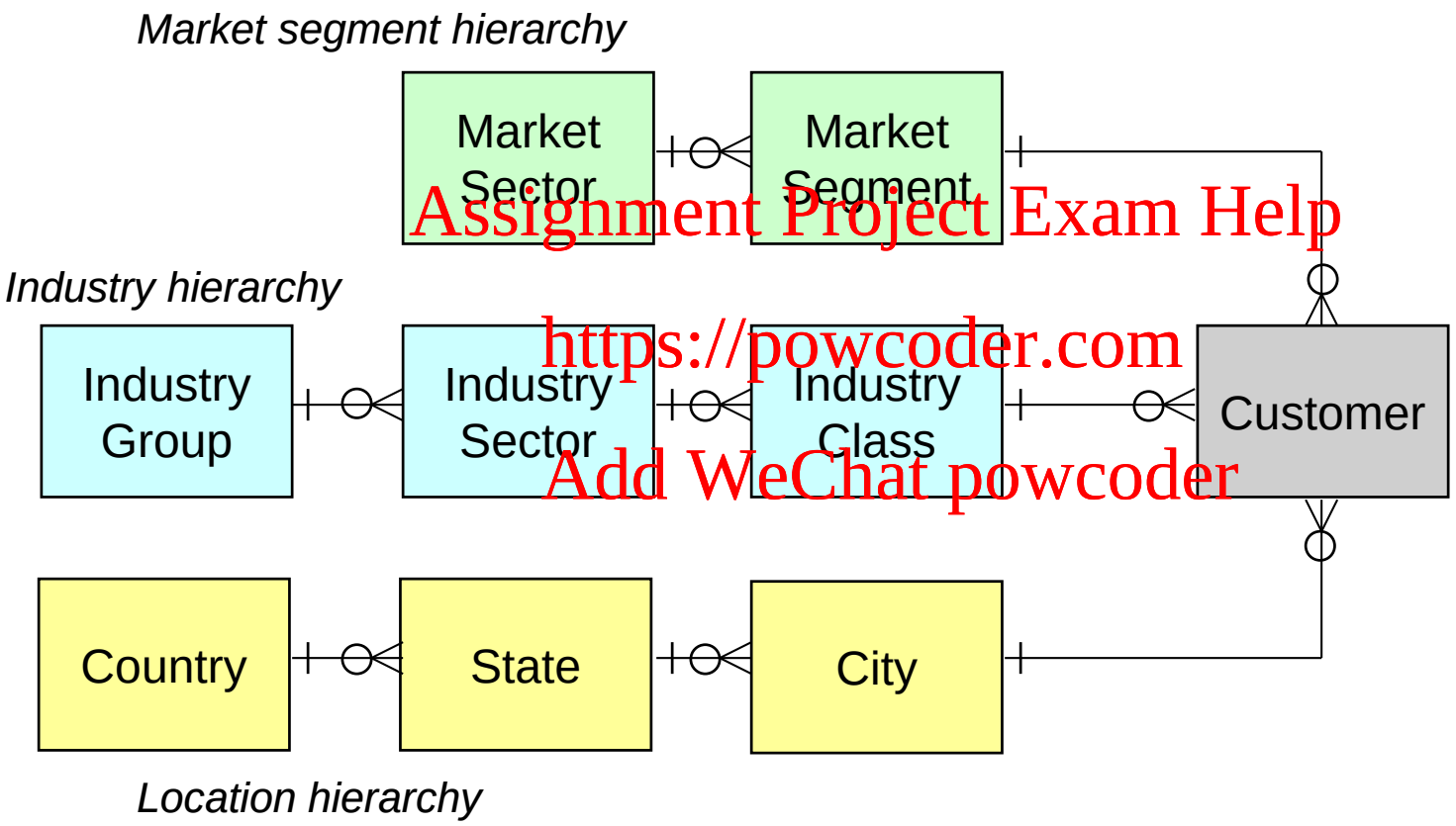
https://powcoder.com

Add WeChat powcoder

Customer-id
Customer name
Market segment
Market sector
Industry class
Industry sector
Industry group
City
State
Country

© University of Melbourne

Embedded Hierarchies in Dimensional Tables

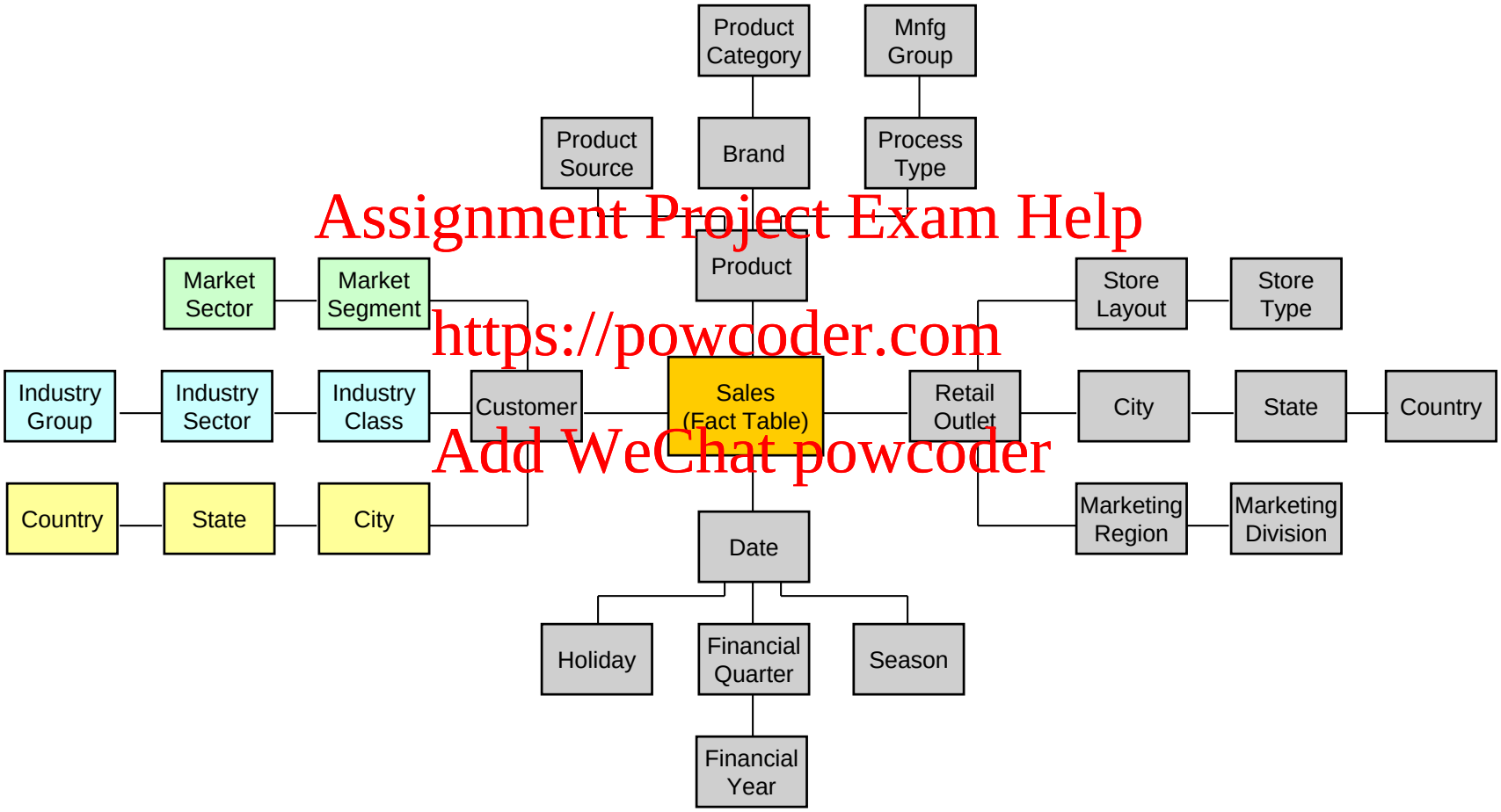


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Customer-id
Customer name
Market segment
Market sector
Industry class
Industry sector
Industry group
City
State
Country

© University of Melbourne

Snowflake Schema: hierarchy in dimensions

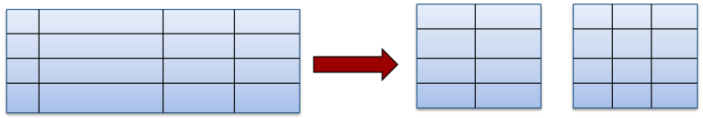


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

© University of Melbourne

Design Outcomes: Normalised or Denormalised?

- Normalisation
  - Eliminates redundancy
  - Storage efficiency
  - Referential Integrity

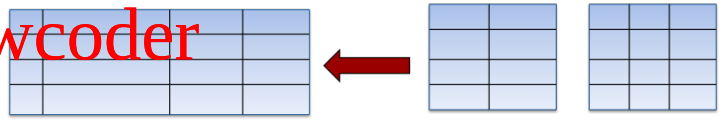


Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- Denormalisation
  - Fewer tables (fewer joins)
  - Fast querying
  - Design is tuned for end-user analysis



© University of Melbourne

Let's try it …

- We are making a data warehouse for a real estate agency. The company wants to track information about the **selling** of their properties. This warehouse keeps information about the **agents** (license#, first name, last name, phone #), **buyers** that come in (buyer id, first name, last name, phone #), and **property** (property#, property address, price). The information managers want to be able to find is **the number of times a property is viewed**, **sales price**. The information needs to be accessible **by rental agent, by buyer, by property** and **for different time** (day, week, month, quarter and year).
- Draw a star schema to support the design of this data warehouse.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

What is Examinable?

- Differences between transactional and informational databases
- Designing a star schema
- Defining facts and dimension tables

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

If you want to know more

More technical details (I won't ask you these things):

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

https://www.youtube.com/watch?v=w-S0fj0fmqg&list=PLdQdqgMB3WzHcEN9PrhADq3CBColhY2hl&index=17

Next Lecture

- Distributed Databases

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

© University of Melbourne