DS-306 Data Warehousing and Business Intelligence

Topic 7: Conceptual Data Warehouse Design

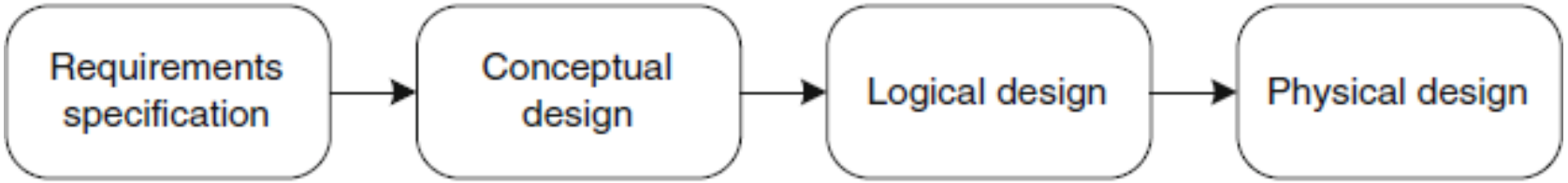
Dr. Khurram Shahzad

Design Specification & Process

- The design process explains a procedure to design the process
- It is a combination of **art** and **science**, but dominated by **art**…

General overview of the method

- Data warehouse is a particular type of database dedicated to analytical purposes

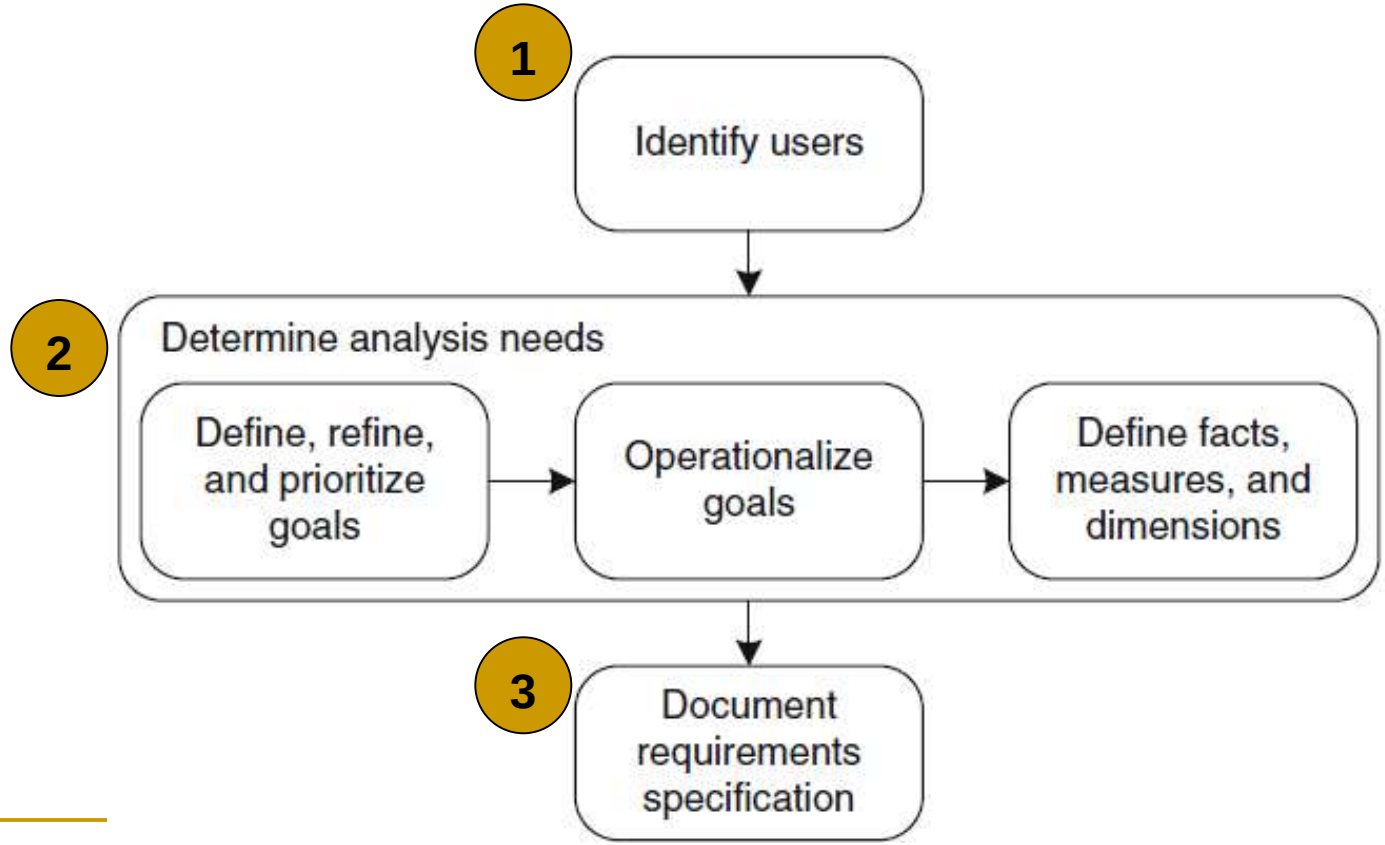


Requirements specifications

- **Requirements specifications** of DW differs significantly from operational system
- In DW, requirement specification determines
  - Which data should be available?
  - How these data should be organized?
  - Also, queries of interest are determined

Analysis driven req. specification

- The Framework



Analysis driven req. specification

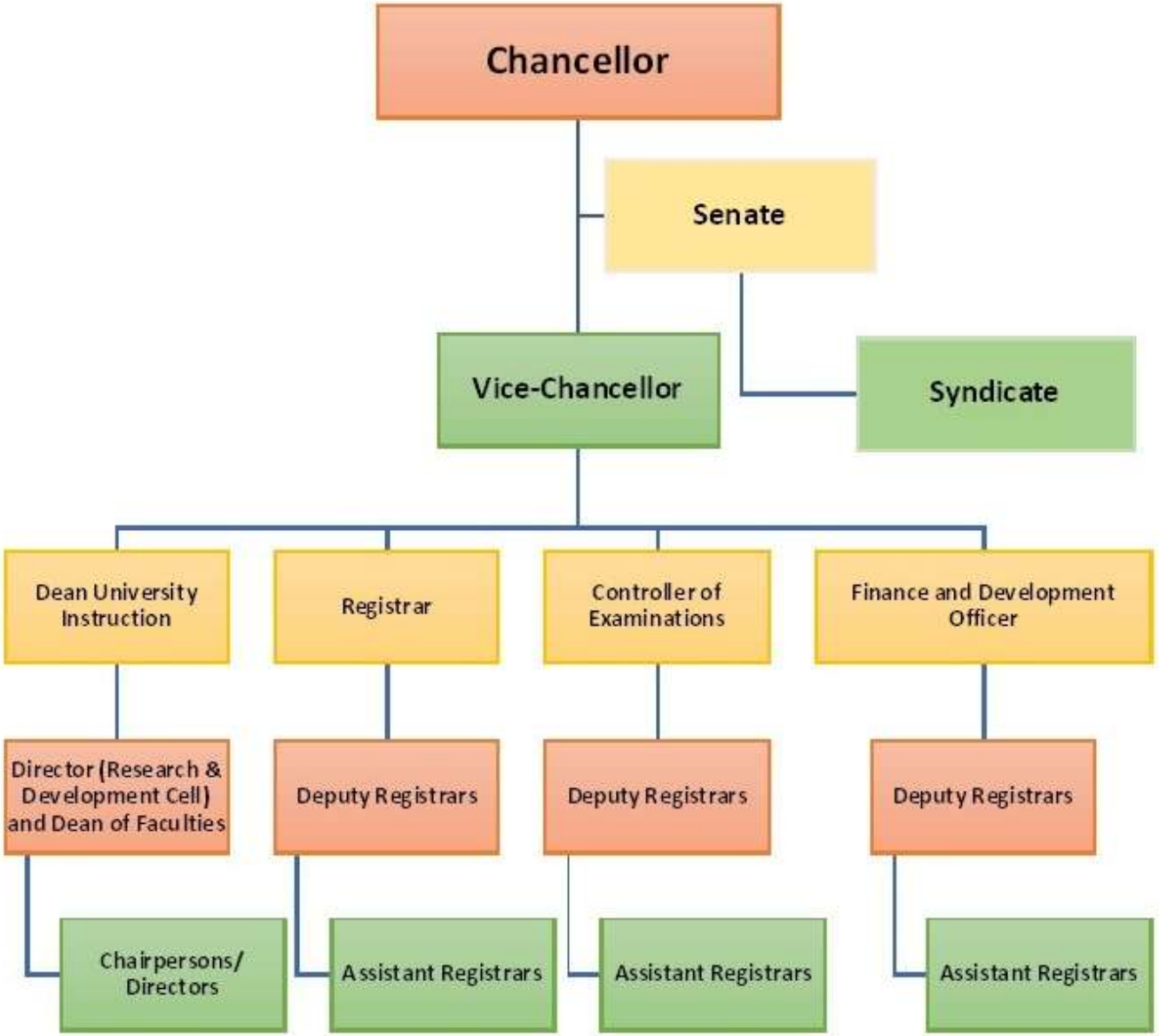
1

- Identify users/ Types of users
  - Executive users
    - Top organizational level
    - Requires global summarized information
    - They help understand goals and overall vision
  - Management users
    - Requires more detailed information pertaining to a specific area of organization
    - Provide more insight into the process or tactics used for achieving business goals
  - Professional users
    - Responsible for specific section or services

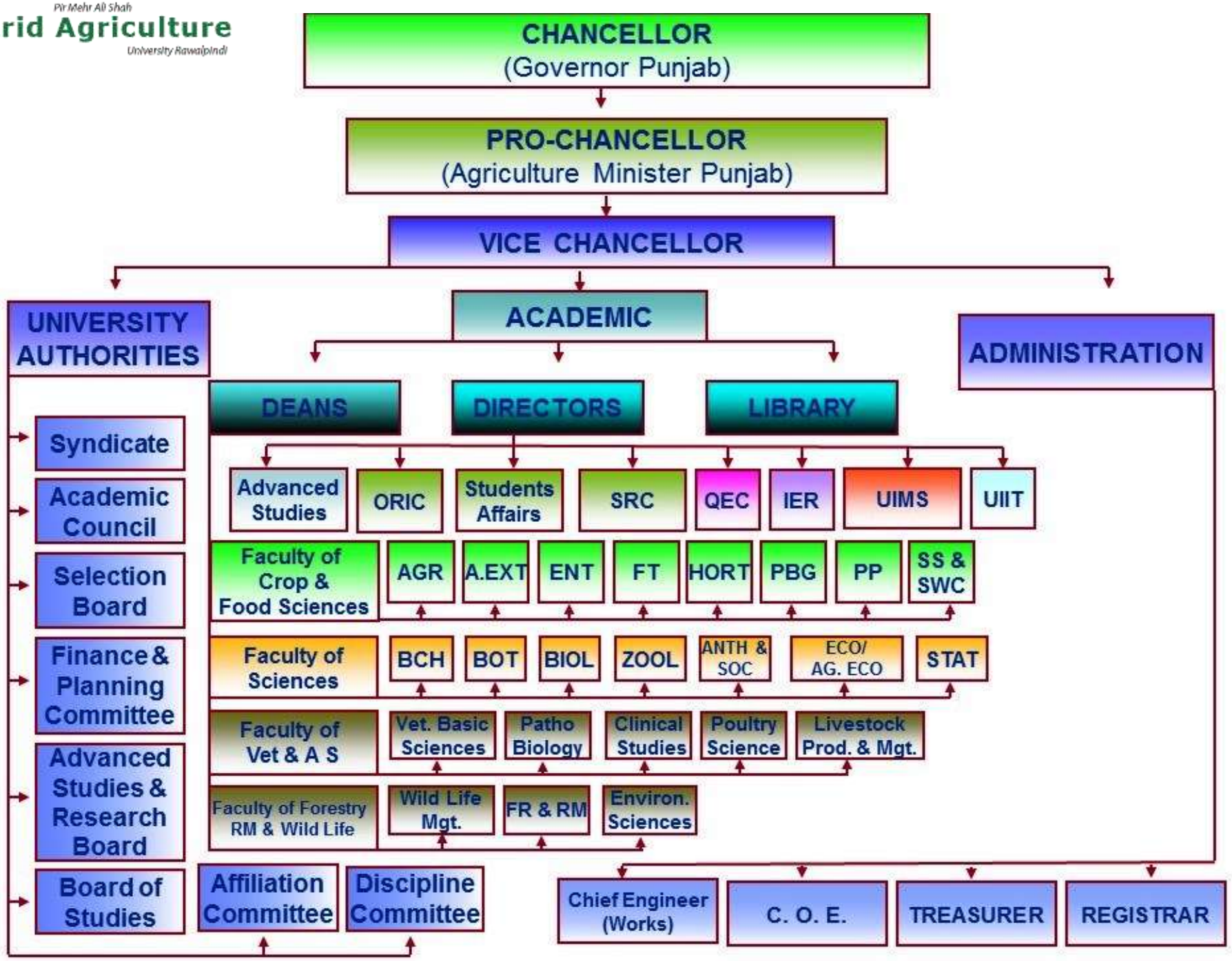
Analysis driven req. specification

- Identification of potential users should also consider different entities in a horizontal division of organization (departments)

Organigram: Panjab University

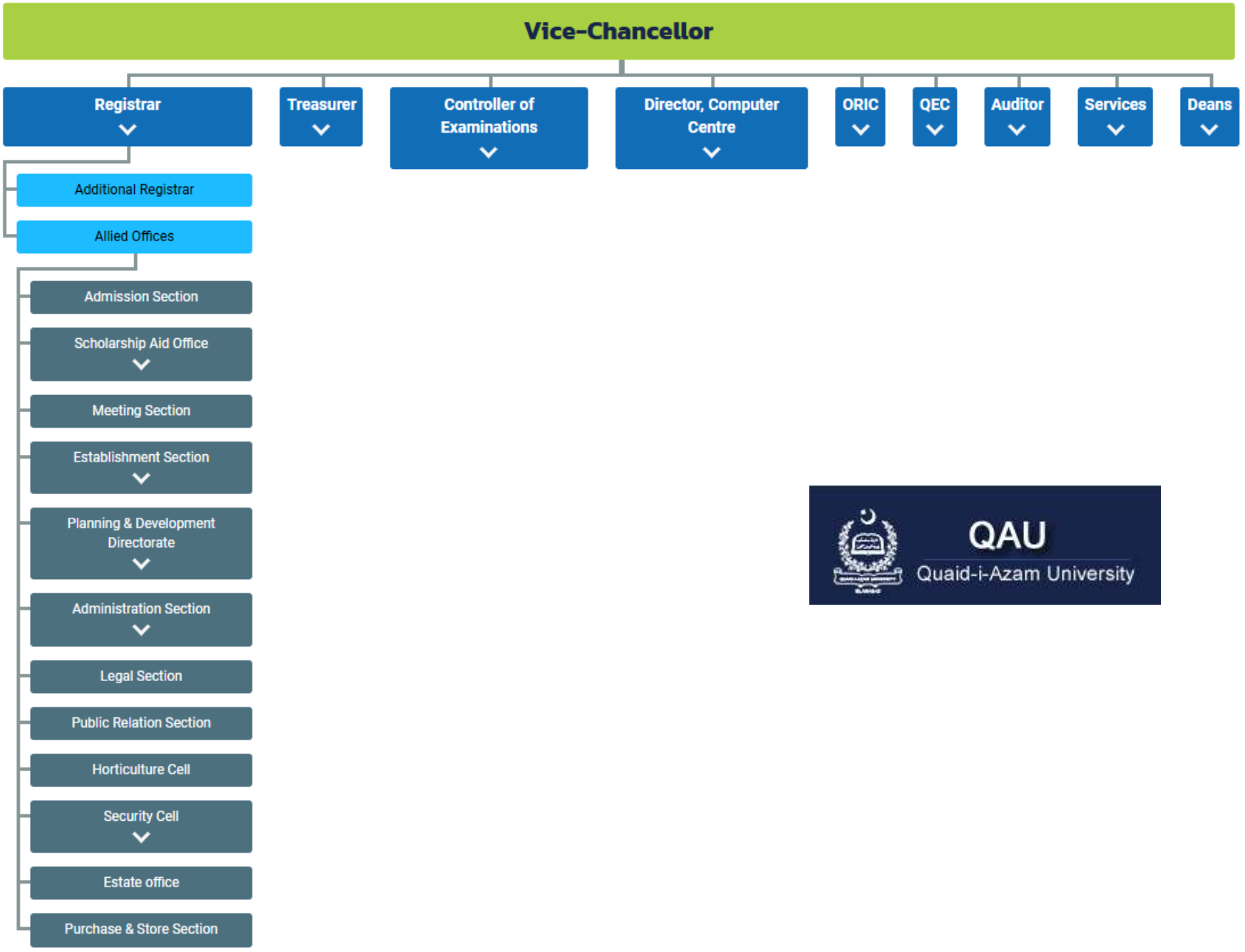


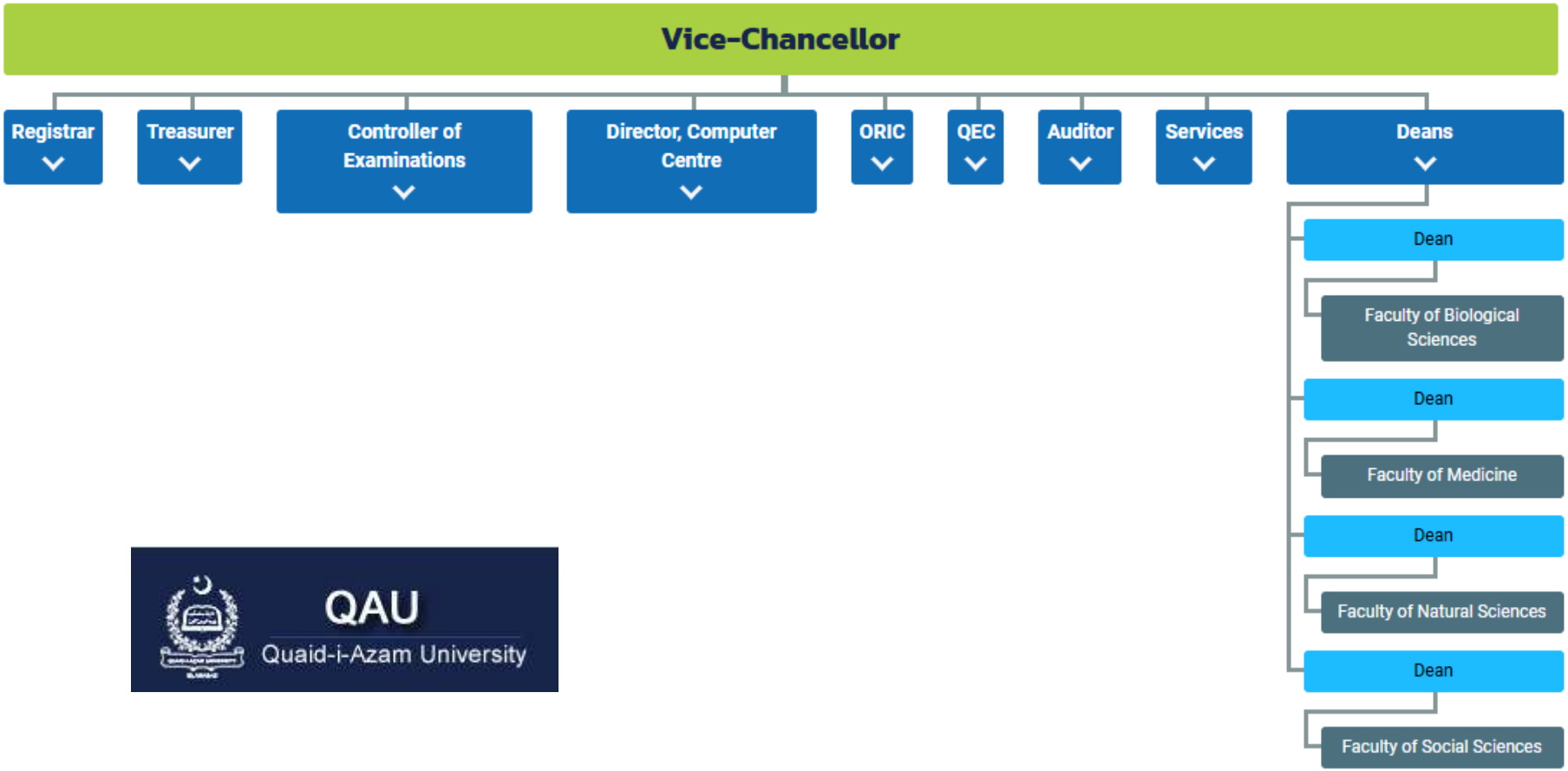
Pir Mehr Ali Shah
Arid Agriculture
University Rawalpindi



QAU
Quaid-i-Azam University

https://qau.edu.pk/organization-chart/


Vice-Chancellor
Registrar
Treasurer
Controller of Examinations
Director, Computer Centre
ORIC
QEC
Auditor
Services
Deans
Additional Registrar
Allied Offices
Admission Section
Scholarship Aid Office
Meeting Section
Establishment Section
Planning & Development Directorate
Administration Section
Legal Section
Public Relation Section
Horticulture Cell
Security Cell
Estate office
Purchase & Store Section
QAU
Quaid-i-Azam University

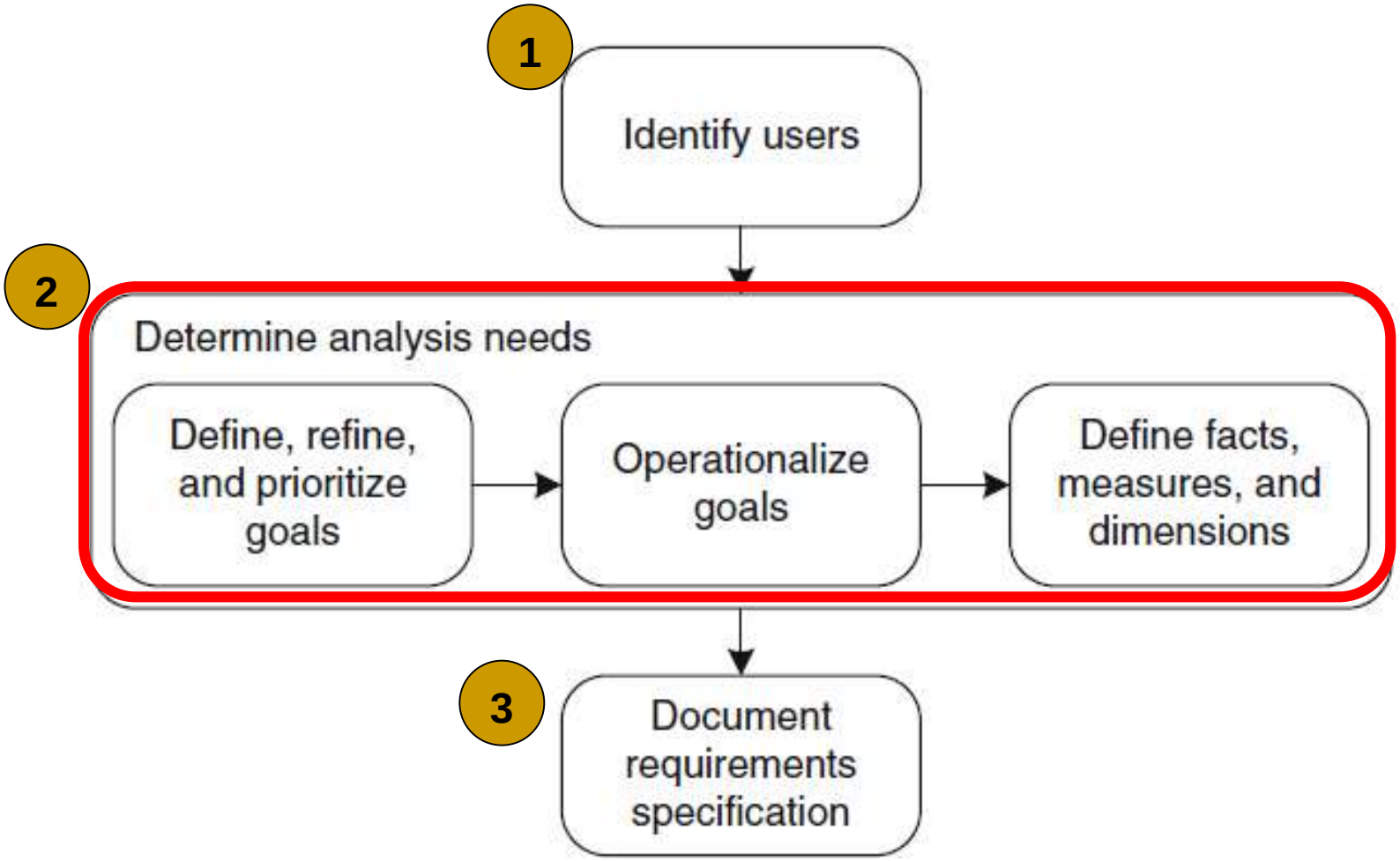


QAU
Quaid-i-Azam University

Powers and Duties of The Syndicate

1. The Syndicate shall be the executive body of the University and shall, subject to the provisions of this Act and the Statutes, exercise general supervision over the affairs and management of the property of the University.
2. In particular and without prejudice to the generality of the foregoing provision, the Syndicate shall have the power:
    a. to hold, control and administer the property and funds of the University;
    b. to govern and regulate, with due regard to the advice of the Finance and Planning Committee in this behalf, the finance, accounts and investment of the University and for that purpose, to appoint such agents as it may think fit:
    c. to consider the annual report and the annual and revised budget estimates and to approve the same6 and to re-appropriate funds from one major head of expenditure to another;
    d. to transfer and accept transfer of moveable or immovable property on behalf of the University;
    e. to enter into, vary, carry out and cancel contracts on behalf of the University;
    f. to cause proper books of accounts to be kept for all sums of money received and expended by the University and for the assets and liabilities of the University;
    g. to invest any money belonging to the University, including any unapplied income, in any of the securities described in Section 20 of the Trust Act 1882 (II of 1882), or in the purchase of immovable property or in such other manner as it may determine, with the like power of varying such investment;
    h. to receive and manage any property transferred and grants, bequests, trusts, gifts, donations, endowments and other contributions, made to the University, and to administer any funds placed at the disposal of the University for specified purposes;
    i. to determine the form, provide for the custody and regulate the use of the Common Seal of the University;
    j. to provide the buildings, libraries, premises, apparatus, equipment and other means required for the purpose of the University, and to establish and maintain halls of residence and hostels or approve or licence hostels or lodgings for the residence of students;
    k. to arrange for the inspection of the colleges and the teaching departments and Institutes;1
    l. to institute Professorships, Associate Professorships, Assistant Professorships, Lectureships and other teaching posts, or to suspend or abolish such posts;
    m. to create, suspend or abolish such administrative, research, extension or other posts as may be necessary;
    n. to appoint University Teachers and other Officers on the recommendations of the Selection Board for teaching and other posts;
    o. to appoint Professors Emeritus on such terms and conditions as may be prescribed;
    p. to confer with the prior approval of the Chancellor, Honorary degrees in accordance with the conditions prescribed;

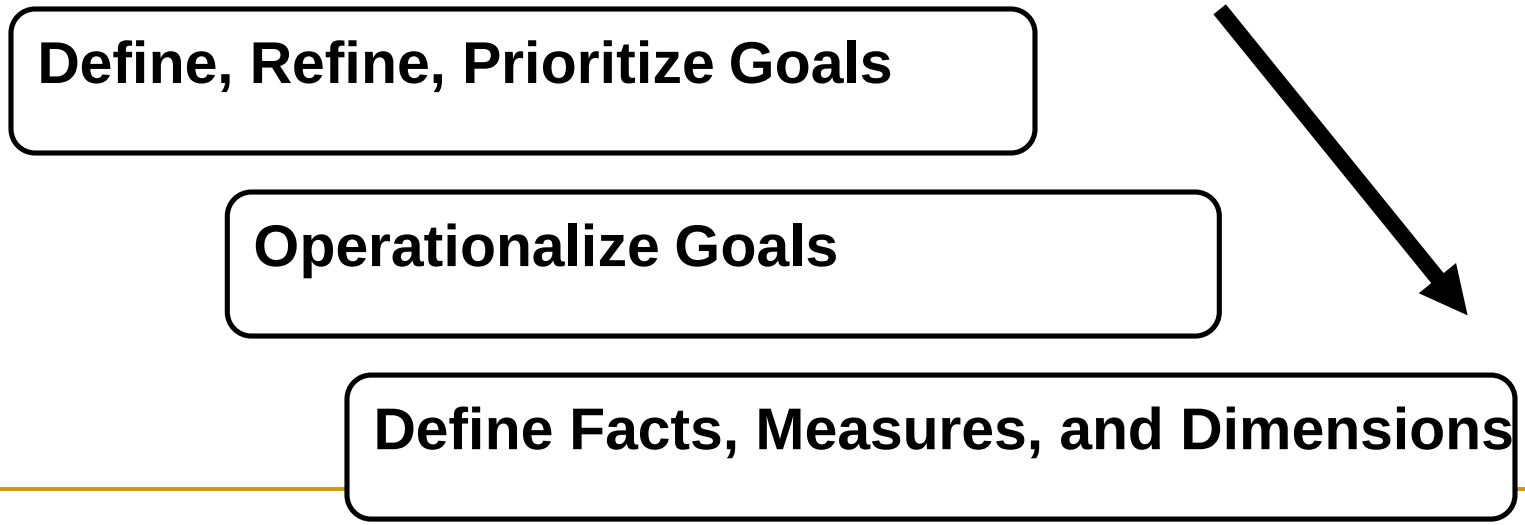
Analysis driven req. specification



Determine Analysis Needs

2

- This phase helps users understand what data should be available to respond to users' expectations
- Steps for Determining Analysis Needs



Determine Analysis Needs

2

- Define, Refine and Prioritize Goals
  - Develop a clear specification of the goals of a company
- Two types of goals
  - General goals
  - Specific goals
- Methods for gathering goals
  - Conduct interviews and brainstorming sessions

Determine Analysis Needs: Define goals

2

- Identify dependencies between goals
- Prioritize goals
- Identify supporting and conflicting goals
- Analyze goals to detect redundancies
  - Some goals could be combined due to similarity
  - Some goals should be discarded because of inconsistency

Determine Needs: Operationalize goals

2

- Make goals concrete
  - By defining subgoals
- For each goal, define measurement criteria (KPIs)
  - Use interviews with users to identify it
- Then, define representative queries
  - Users define these queries
- Each user provides list of queries needed for daily tasks

Determine Needs: Operationalize goals

2

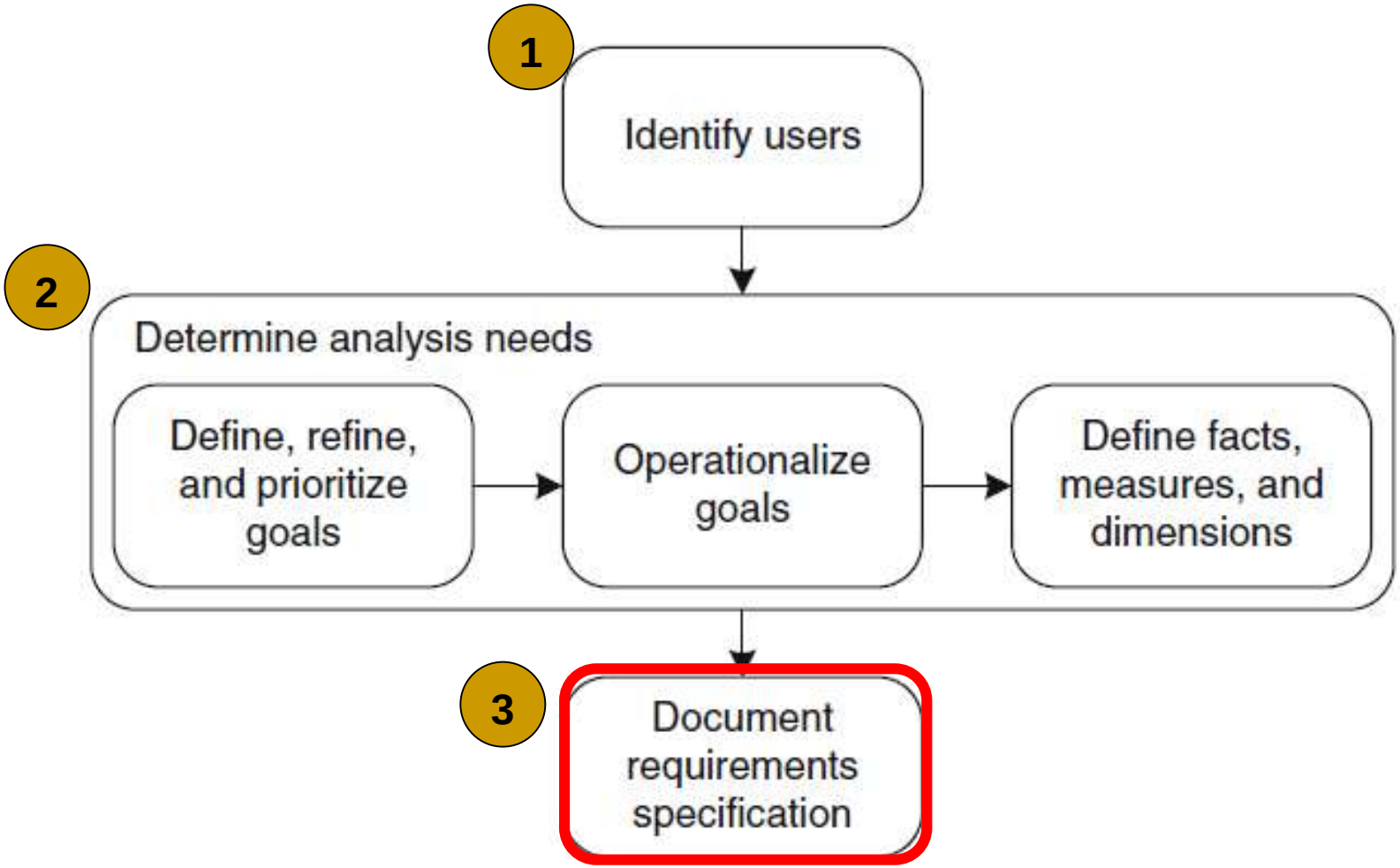
- Each user provides list of queries/reports needed for daily tasks
- Disambiguate query terms
  - For example, best customer
    - Customer with highest total sales amount
    - Customer with high margin, or ROI
    - Customer with highest preference
- Prioritize queries
- Non-functional requirements can be included
  - Response time, accuracy

Determine Needs: Define facts, dim

2

- Manual process to define measures and dimensions
- Analyst tries to identify the underlying facts and dimensions
- Example: ‘top five customer names with monthly average sales higher than 1500$’
- Data elements
  - Customer name
  - Month
  - Sales

Analysis driven req. specification



Document requirement specification

3

- Develop documentation
  - Include all elements required by designers
  - Dictionary of the terminology
  - Organizational structure
  - Policies
  - And, constraints of the business

Example

3

- Sample Data Warehouse Requirements Document

Northwind Case Study

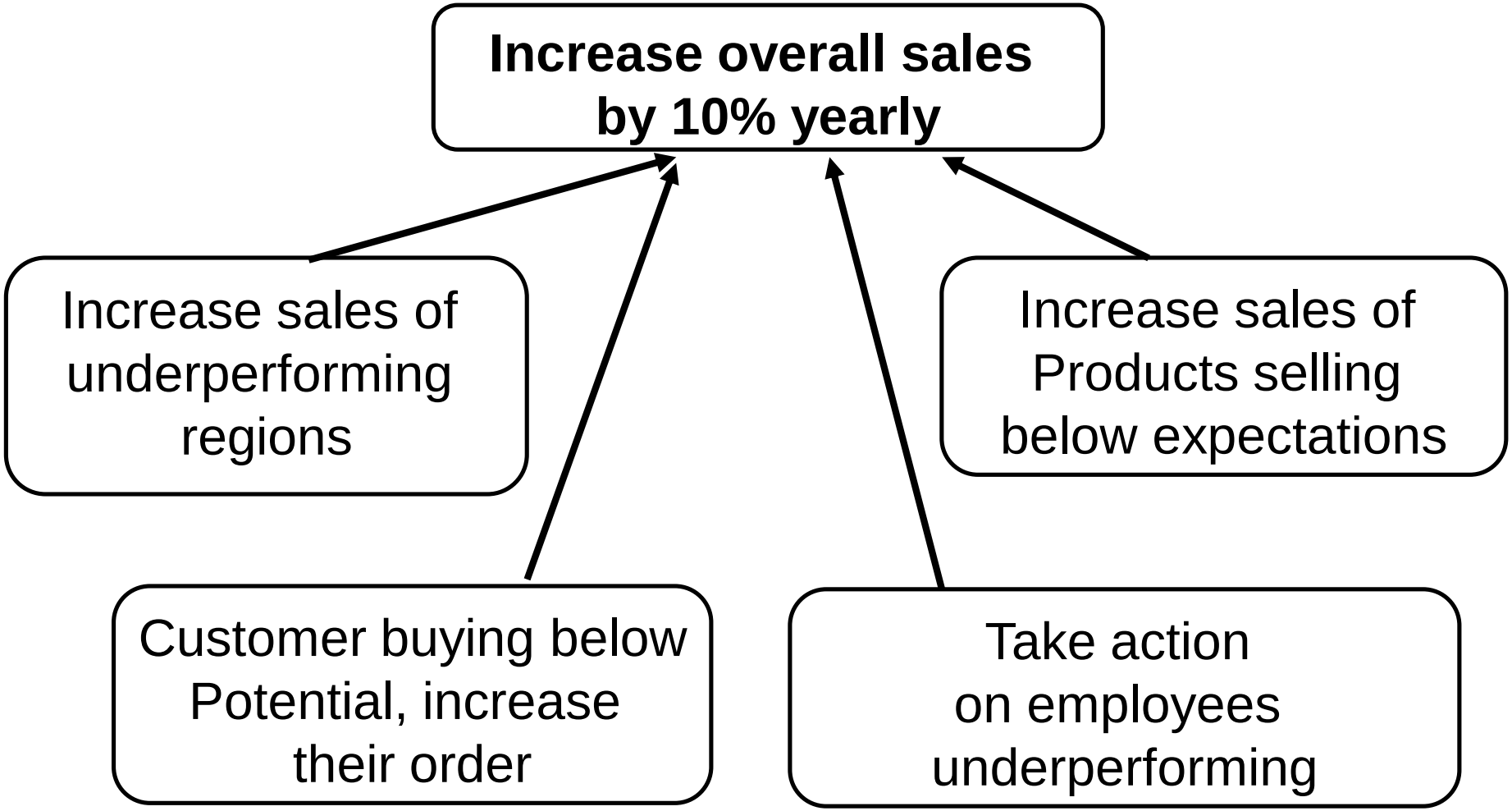
1. Identify users

- Executive:
  - Members of the board of directors of NW who define overall company goals
- Management
  - Managers at department level, e.g. Marketing, regional sales, and human resources
- Professional
  - Personnel who implement indications of the management e.g. Marketing executive officers

2. Determine Analysis Needs

- Increase the overall company sales by 10% yearly
- Subgoals
  - Increase sales in underperforming regions
  - For customers buying below their potential, increase their orders
  - Increase sales of products selling below the company expectations
  - Take action on employees performing below their expectation

2. Determine Analysis Needs



2. Determine Analysis Needs

1. Increase sales in underperforming regions

1a. Five best- and worst-selling (measured as total sales amount) pairs of customer and supplier countries.

1b. Countries, states, and cities whose customers have the highest total sales amount.

1c. Five best- and worst-selling (measured as total sales amount) products by customer country, state, and city.

2. Determine Analysis Needs

2. For customers buying below their potential, increase their orders

2a. Monthly **sales** by **customer** compared to the corresponding sales (for the same customer) of the previous year.

2b. Total number of orders by customer, time period (e.g., year), and product.

2c. Average unit price per customer.

2. Determine Analysis Needs

3. Increase sales of products selling below the company expectations

3a. Monthly sales for each product category for the current year.

3b. Average discount percentage per product and month.

3c. Average quantity ordered per product.

2. Determine Analysis Needs

4. Take action on employees performing below their expectation

4a. Best-selling employee per product per year with respect to sales amount.

4b. Average monthly sales by employee and year.

4c. Total sales by an employee and her subordinates during a certain time period.

2. Determine Analysis Needs

Dimensions /measures	Hierarchies and levels	Analysis scenarios											
		1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	4c
Employee	**Supervision** Subordinate → Supervisor **Territories** Employee ⇆ City → State → Country → Continent	–	–	–	–	–	–	–	–	–	✓	✓	✓
Time	**Calendar** Day → Month → Quarter → Semester → Year	–	–	–	✓	✓	✓	✓	✓	–	✓	✓	✓
Product	**Categories** Product → Category	–	–	✓	–	✓	–	✓	✓	✓	✓	–	–
Customer	**Geography** Customer → City → State → Country → Continent	✓	✓	✓	✓	✓	✓	–	–	–	–	–	–
Supplier	**Geography** Supplier → City → State → Country → Continent	✓	–	–	–	–	–	–	–	–	–	–	–
Quantity	–	–	–	–	–	–	–	–	–	✓	–	–	–
Discount	–	–	–	–	–	–	–	–	✓	–	–	–	–
SalesAmount	–	✓	✓	✓	✓	–	–	✓	–	–	✓	✓	✓
UnitPrice	–	–	–	–	–	–	✓	–	–	–	–	–	–

3. Document requirements specification

- The information includes the specifications of the users' requirements
  - Summarized information
  - Explanation of each element
    - Obtain it by interviewing users or administrative staff

Data Mart Design

The Inmon Approach

Design Specification & Process

- The design process explains a step-by-step procedure to design the artifact
- It is a combination of art and science, but dominated by art…

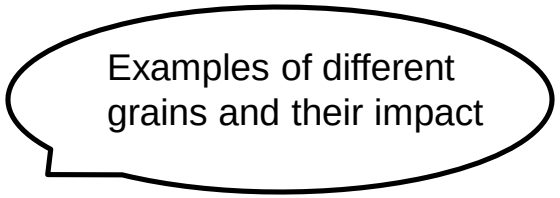
Designing Process of Star

- Steps in Designing dim. modeling
  - **Step 1:** Choose Business Process
  - **Step 2:** Choose the Grain
  - **Step 3:** Choose the Facts
  - **Step 4:** Choose Dimensions
  - Relationship between fact and dimensions

Step 1: Business Process

- A major operational process in organization
- Usually, is represented by a use-case diagram in operational source
- Supported by legacy system
- They are critical areas for the company
- Reporting system plays an important role in identifying process
- Are also called ‘subjects’
- **Difficult** to decide its scope

Step 2: Granularity

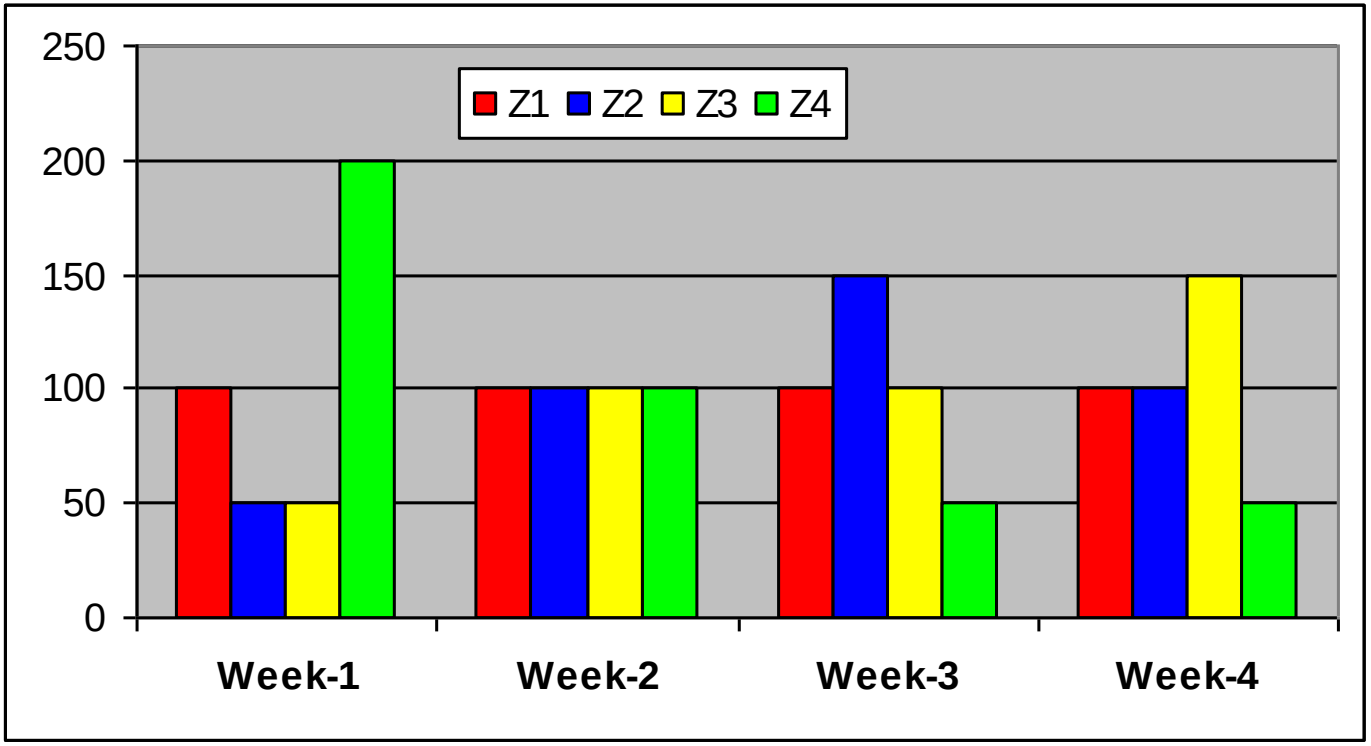


- Determine level of detail
- Grain is, unit of analysis
- Grain varies with facts, it represents what individual fact table record as data
- Typical grains
  - Individual transaction
  - Specific duration
- Examples of grain
  - **Time:** Year, Quarter, Month, Date
  - **Location:** Continent, Country, Division, City
- Supports roll-up and drill-down.

Aggregation hides crucial facts: Example

	Week-1	Week-2	Week-3	Week-4	Average
Zone-1	**Just looking at the averages i.e. aggregate**				100
Zone-2					100
Zone-3					100
Zone-4					100
Average	100	100	100	100	

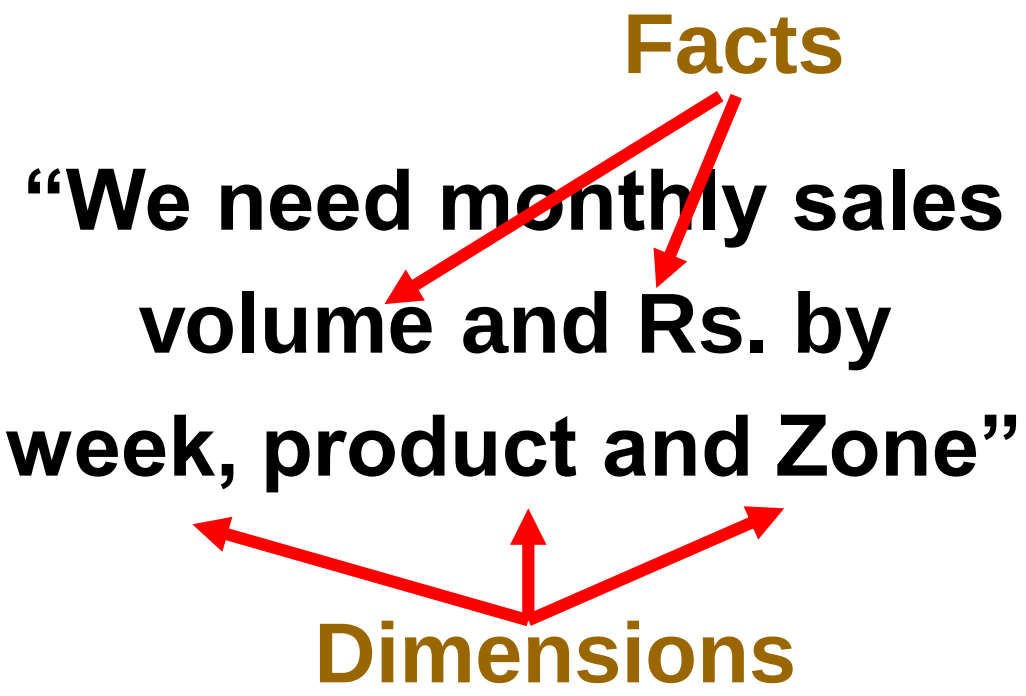
Aggregation hides crucial facts chart



Z1: Sale is constant (need to work on it)
Z2: Sale went up, then fell (need of concern)
Z3: Sale is on the rise, why?
Z4: Sale dropped sharply, need to look deeply.

W2: Static sale

Step 3: Choose Fact Statement



Step 3: Identify Facts

- Aggregates that are critical for business
- Remember that best facts are numeric, summarized data, aggregates
- Example: Profit, Qty sold, Amunt earned
- Must match with the grain
- Fact can be repeated in multiple tables
- Re-Group the facts
  - Facts with same grains and critics should be placed together

Step 4: Identify Dimensions

- Are the aspects, which can be used for analyzing facts
- Time dimension is compulsory
- Dimensions have hierarchies
- A dimension can share across schemas
- A Tip:
  - An attribute can live in only one dimension

Step 5: Identify Relationship

- Link dimensions with fact
- Resolve many-to-many relationship by adding surrogate key
- Relate fact with all possible dimension

Example Case*

GB Video is a members-only video rental organization.

Members join GB Video by filling out an application and giving a credit card as a reference for membership.

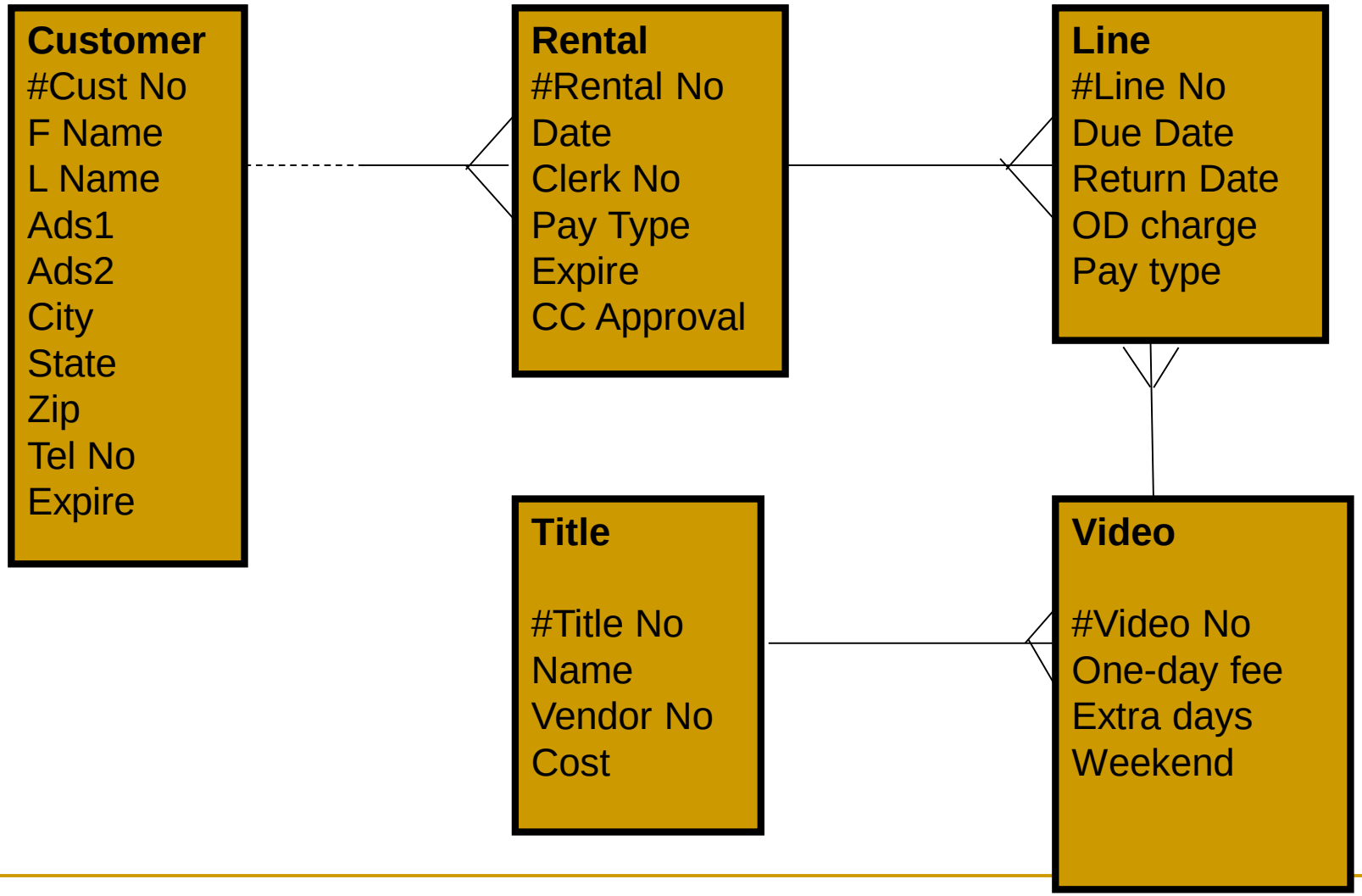
When members come into the store they select the particular videos that they want to take and bring them to a clerk to check out.

The clerk completes an electronic rental form for them and calculates the amount of the bill.

Each video has a one day fee, an additional lower rate for extra days, a premium charge for each weekend day, and a charge for late (overdue) returns.

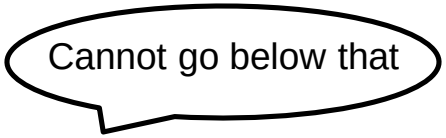
When the customer returns the video, the clerk checks to see if overdue charges apply and returns the video to inventory.

Example Case ERD



Building a DM for GB Video

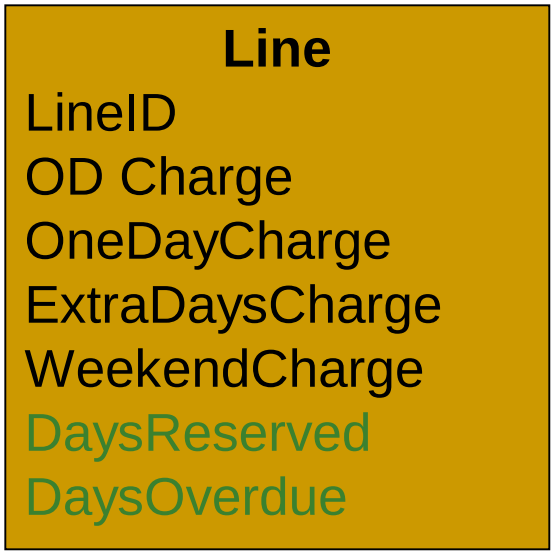
- Choose the Business Process
  - **Select a many-to-many association table as the central fact table for the mart.**
  - Look for the tables which are capturing transactions.
  - If there are more than one association tables in the E-R diagram then the system will probably generate more than one potential Stars
- Choose the Grain
  - The records in the association/transaction table define the *grain* of the fact table
  - The designer should explicitly describe what that grain is in operational terms
  - In this case the grain of the mart is *"individual video rental line describing a video rented to a specific customer on a specific rental form"*



Building DM (Contd.)

- Choose the Facts
  - The most useful facts are numeric and additive
  - Additive facts are those that can be added up across records
  - In sales marts like the GB Video mart the most common facts are ‘cost’ and ‘quantity’.
  - Most order systems can record price, quantity and cost. Price, for example, is not additive, but quantity and cost are.
  - Overdue (OD) charge is an additive fact.
  - One-day fee, extra days, and weekend charge are not additive but because change so often that they are better facts than dimensional attributes.

Partial Fact Table

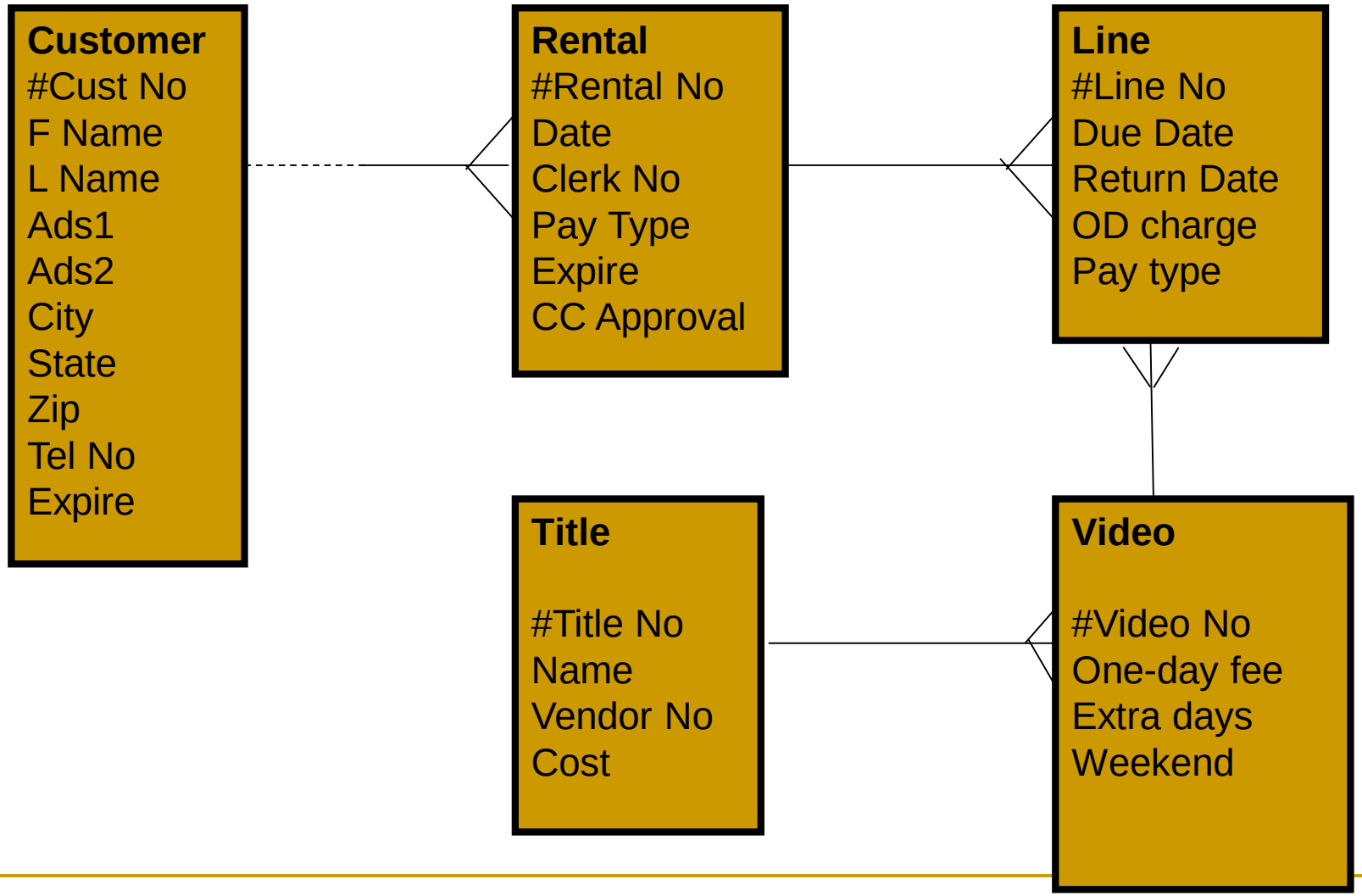


- Now we need to connect this fact table to dimensions.
- First it requires finding dimensions and their descriptive attributes.

Building DM (Contd.)

- **Dimension Tables**
  - **Create dimension tables from entities by promoting foreign keys into the fact table**
  - For example, the relationship “Requestor of” connecting records in the Customer table to individual Rental occurrences would be implemented by a foreign key in the Rental table.
  - Moving that foreign key from the Rental table into the Line table maintains the information content and makes for a simpler query structure.
  - Promote dates to the fact table and create a date dimension to replace the actual date value by an ID.

Example Case ERD



Dim. Model for GB Video