FUNDAMENTALS OF DATABASES

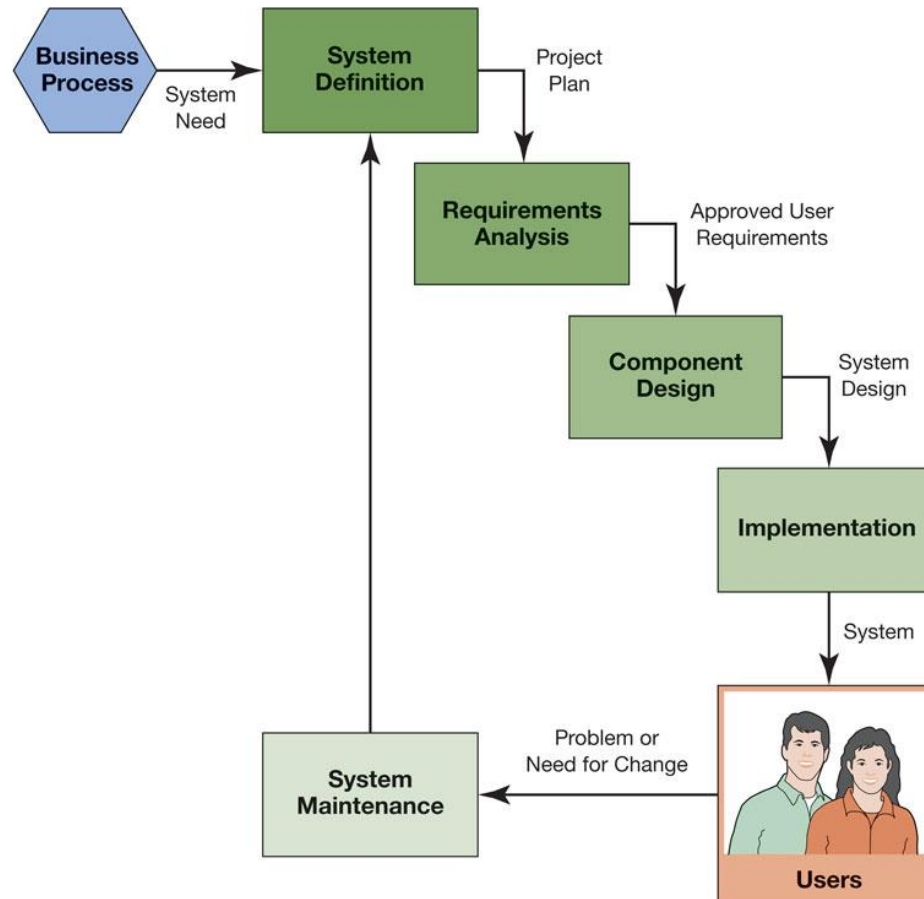## Database development process

NGUYEN Hoang Ha

Email: nguyen-hoang.ha@usth.edu.vn

SDLC

Systems Analysis and Design

- Systems analysis and design is the process of creating and maintaining information systems.

- The classic methodology used in systems analysis and design to developed information systems is called the systems development life cycle (SDLC) and is composed of the following steps:

    - Systems definition

    - Requirements analysis

    - Component design

    - Implementation

    - System maintenance

The SDLC in Use

The System Definition Step

- Input: the need for an information system to support a business process

- Output: project plan

- This step includes:

  - Define the information system project goals and scope

  - Assess the feasibility of the project (cost, schedule, technical, organizational)

  - Form the project team

  - Plan the project (specify tasks, assign personnel, determine task dependencies, set schedules)

Goals of AMS

- • Efficient Information Management: Centralize student, faculty, and course information for easy access and management.

- • Enhanced Communication: Facilitate effective communication between students, faculty, and administration.

- • Student Progress Tracking: Monitor and track student performance, grades, attendance, and overall progress.

- • Resource Allocation: Optimize resource allocation, including classrooms, and faculty.

- • Administrative Efficiency: Simplify administrative tasks such as admissions and record-keeping.

- • Data Security and Privacy: Ensure data security measures to protect sensitive information.

Scope of AMS

- • User Access and Security: Implement user roles, access controls, and data encryption for security

- • Student Information Management: Capture and maintain student profiles, enrollment details, academic records, etc.

- • Faculty Information Management: Store faculty profiles, qualifications, schedules, and performance evaluations.

- • Course Management: Facilitate course creation, scheduling, allocation of resources, and curriculum updates.

- • Attendance and Grading: Automate attendance tracking and grading systems, providing real-time updates.

- • Reporting and Analytics: Generate reports on student performance, faculty workload, and other key metrics.

The Requirements Analysis Step

- Input: project plan

- Output: a set of approved user requirements

- This step includes:

  - Conduct user interviews

  - Evaluate existing systems

  - Determine needed new forms/reports/queries

  - Identify needed new applications features and functions

  - Consider security

  - Consider the 05 components of an information system: hardware, software, data, procedures, people

The Component Design Step

- Input: user requirements

- Output: final system design.

- This step includes:

  - Determine hardware specifications

  - Determine program (software) specifications

  - Create the database design

  - Design business procedures

  - Create job descriptions for business personnel

The Implementation Step

- Input: final system design

- Output: a final installed and functioning information system

- This step includes:

  - Build system components

  - Conduct component tests

  - Integrate the components

  - Conduct integrated component tests

  - Convert to the new system

	Hardware	Software	Data	Procedures	People
Component Design Step	Determine hardware specificiations	Select off-the-shelf software if available. Design custom programs if necessary	Design database and related application components	Design user and operational procedures	Develop job descriptions
Implementation Step	Obtain, install, and test hardware	License and install off-the-shelf software. Create custom programs if necessary. Test programs	Create database. Populate with data. Test database and data.	Document procedures. Create training programs. Review and test procedures	Train personnel. Hire new personnel if necessary.
	Integrated Testing and Startup				

The System Maintenance Step

- Input: the implemented system

- Output: an updated system/ OR a request of a new SDLC cycle to enhance the information system

- This step includes:

  - Update the system with patches, service packs, and new software releases

  - Record and prioritize requests for system changes of enhancements

DATABASE DEVELOPMENT PROCESS

Steps in the Database Development Process

1. Requirements analysis

2. Data analysis using ERD

3. Database design
   1. Logical
   2. Physical

Sources of Requirements for a Database Applications

- User interviews

- Forms

- Reports

- Queries

- Application programs

- Websites

- Use cases

- Business rules

Interview to get Ideas

- What kind of objects objects/entities do you want to manage?

  - List of Academic years

  - Students

  - Class rooms, each locate in a building

  - Subjects which are implemented into classes in specific Academic years. Each class is organized into sessions. The attendance status of each student for each session can be recorded. The assessment results of each student are recoreded

  - Teachers: internal and external (Adjunct)Suppliers who provide products for Northwinds

  - Departments. Each student and internal teacher belong to a department.

Example of Notes from an Interviews

- Date: November 10, 2024

- Interviewee: Dr. Sarah Lee, USTH
  Role: Registrar, oversees enrollment, course scheduling, student records, and academic progress tracking

- Desired Features for the New AMS

  - Integrated Course Scheduling Module:

  - Centralized Student Records Database:

  - Enhanced Reporting and Analytics:

Example of Notes from an Interviews (cont')

- Security and Compliance Requirements
    - Data Security: Ensuring all student data is protected and encrypted, with role-based access to maintain data privacy.
    - Compliance: System must adhere to university policies and relevant government regulations (e.g., FERPA, GDPR for student data protection).
    - Audit Trail: System should maintain logs of changes in student records, enrollment, and grade adjustments for accountability.
- Other Considerations
    - User Training: Staff and students need clear training and support resources for the transition to the new system.
    - Mobile Access: The system should be mobile-friendly to allow students and staff to access essential functions on various devices.
    - Scalability: The AMS should be flexible and scalable to handle increased enrollment and evolving academic requirements over time.

User requirements of AMS

- Authentication
- Manage user
- Manage academic year
- Manage department
- Manage classroom
- Manage subject
- Manage class
- View timetable
- View student list of the class
- Check Attendance
- Enter assessment results
- View academic result
- Generate reports

The use case diagram of AMS

Validating the Data Model

- After the data model (e.g ERD) has been completed, it needs to be validated:

  - The most common way is to show it to the users and obtain their feedback

  - prototyping is commonly use to validate forms and reports

  - A data model needs to be evaluated against all use cases

NORTHWIND SAMPLE PROJECT

Context

- Northwind a fictious company that imports and exports specialty foods from around the world.

- Northwind's office is in the Seattle, Washington, USA and London, British Isles.

- Mission: develop a RDB for business of Northwind

GETTING IDEAS

Interview to get Ideas

- What kind of objects /entities do you want to manage?
  - List of Suppliers who provide products for Northwinds
  - List of Customers who buy products
  - Products: list of product information
  - Purchase orders: the order that Northwind place with their suppliers.
  - Orders: the selling orders
  - Shippers: the company helps Northwind to ship products to customer according to the order
  - Invoices: the bill issued once the order completed
  - Employees: the employees' information. As the sale persons, the employees support client make orders, to follow up the order, to select shipper for each order.
  - order to by products from Suppliers.

Interview to get Ideas

- **What is the detailed information of each entities/objects**

  - Products: are the products that Northwind trades in. It has a ID, Name, List Price, and Category name.

  - Customers: who buy products from Northwind with following properties: Name, Email (unique), Phone, Address.

  - Suppliers– who supply to the company with following information: Name, Email (unique), Phone, Address.

  - Orders: the order that customers place with Northwind. Each Order consists OrderCode which is unique, of Order Date, and some lines which are often called Order Details with following information: which Product, Quantity, and Unit Price.

  - Purchase Orders (a.k.a. PO): refer to purchase order transaction that Northwind makes with a Supplier. Each PO has Creation Date, an ordinal number which is unique with each Supplier, Payment Method (cash, check, or transfer), and a list of Purchase Order Details describing which Product, Quantity, and Unit Cost.
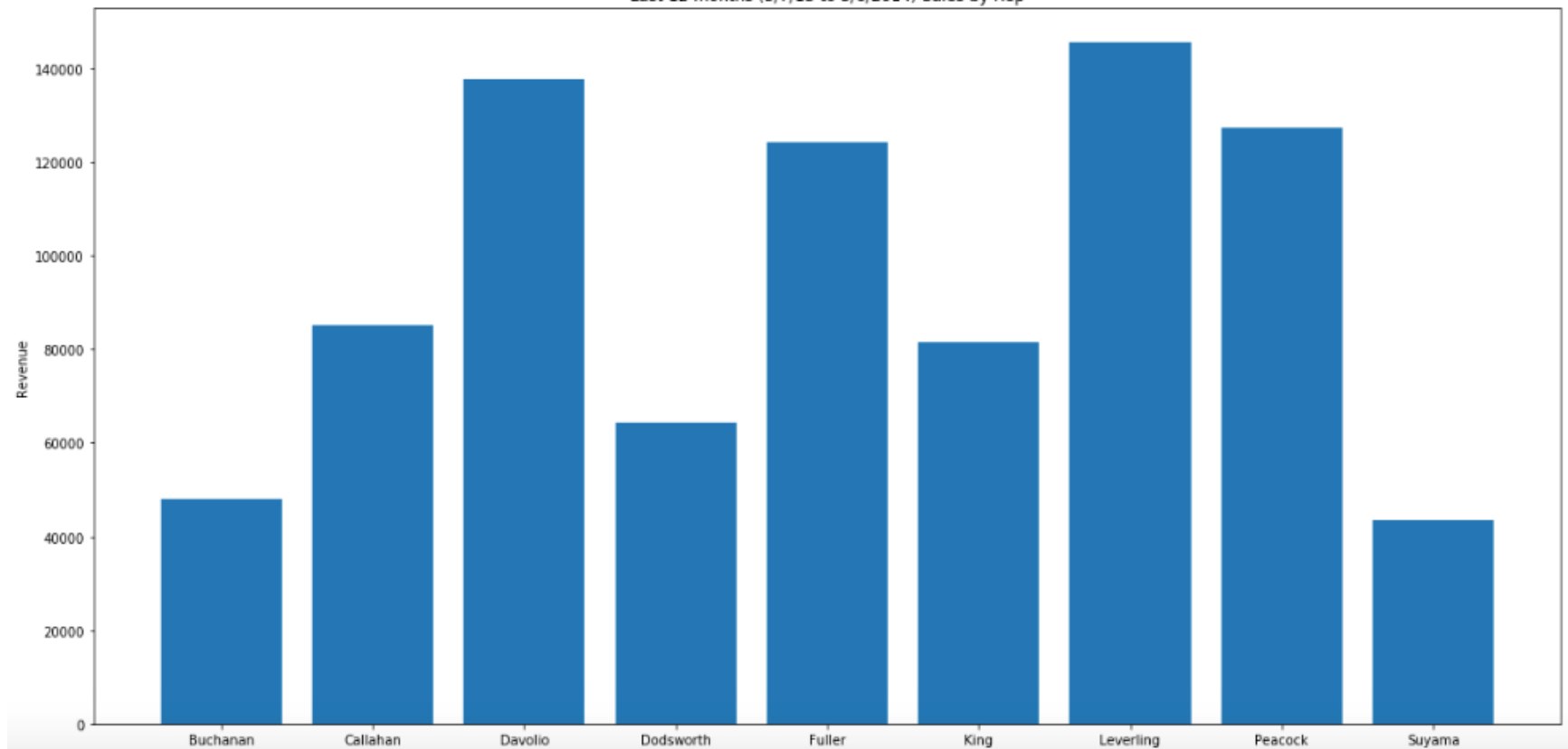
Bussiness processes

- Northwind wants to Insert, update, delete, make a report on the list of: suppliers, customers, products, POs, orders, shippers, invoices, employees.

  - But do not allow to delete products, employees, customers, suppliers, shipper who/which already invovled in an order/PO/invoice.

- When staff of Northwind record an order/PO/invoice sucessfully, he should be able to print out it.

- The manager of Northwind wants to see some reports on the busuniess to be able to make better decissions.
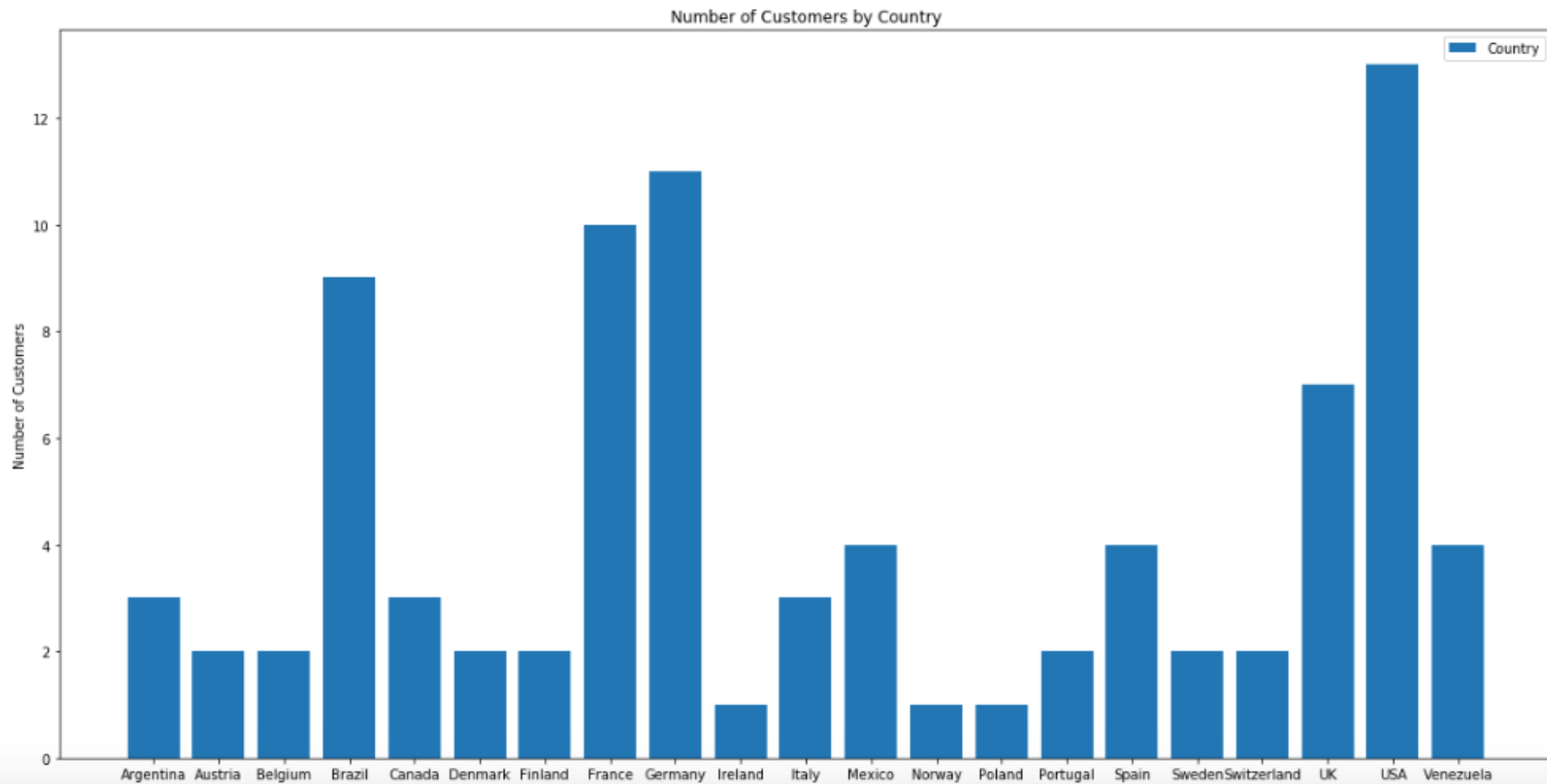
Some reports for the manager

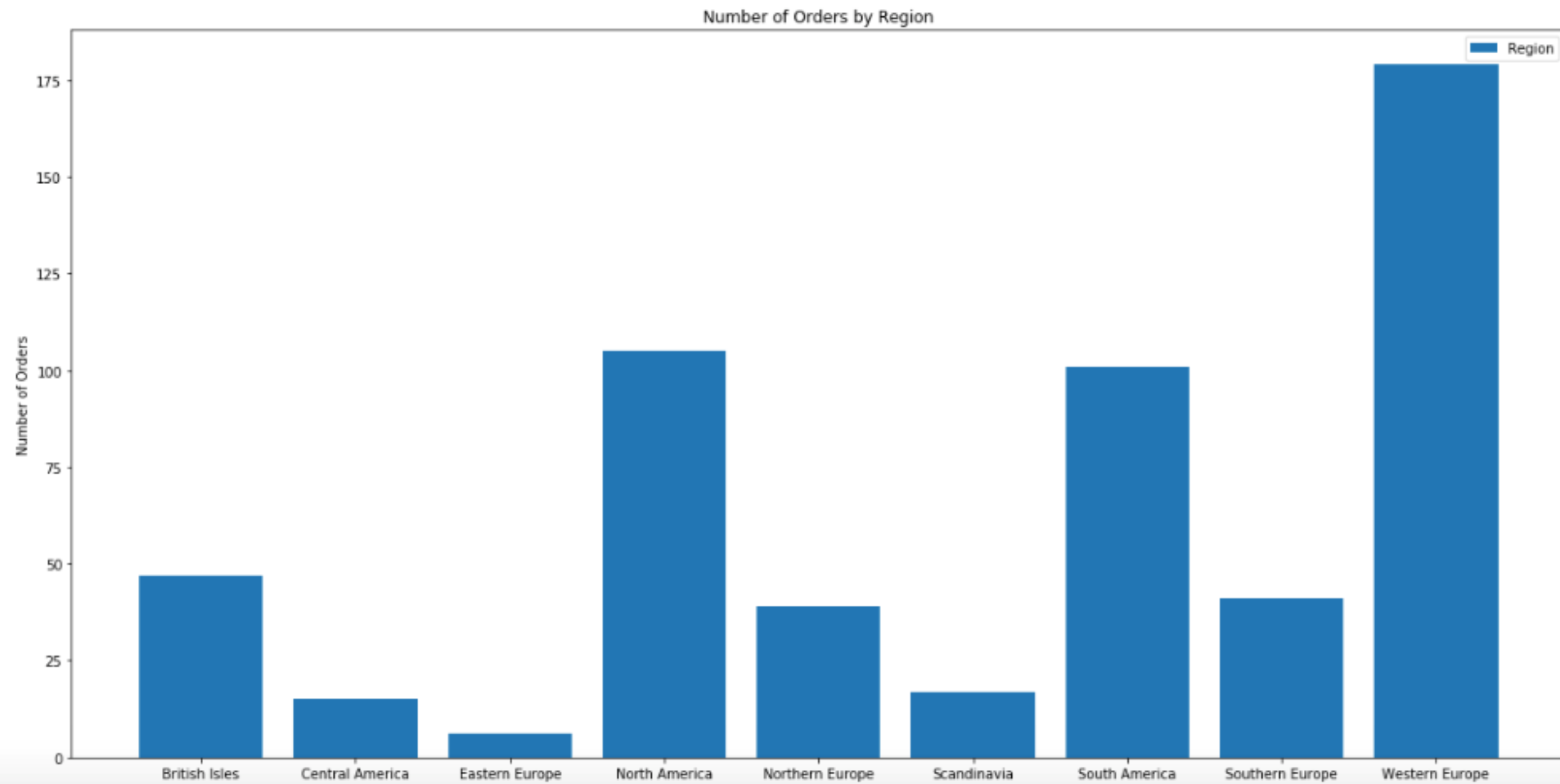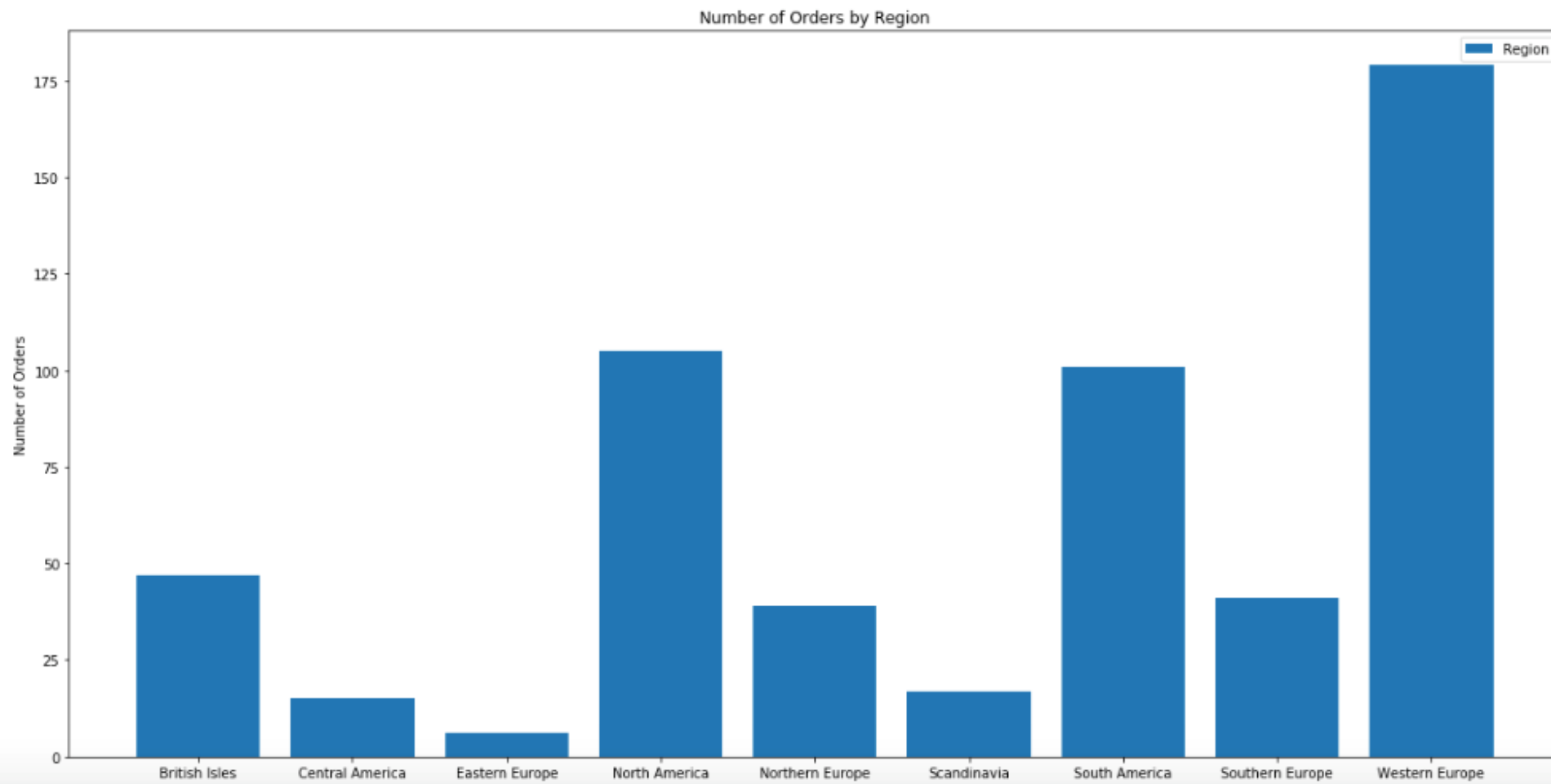Last 12 months (5/7/13 to 5/6/2014) Sales by Rep
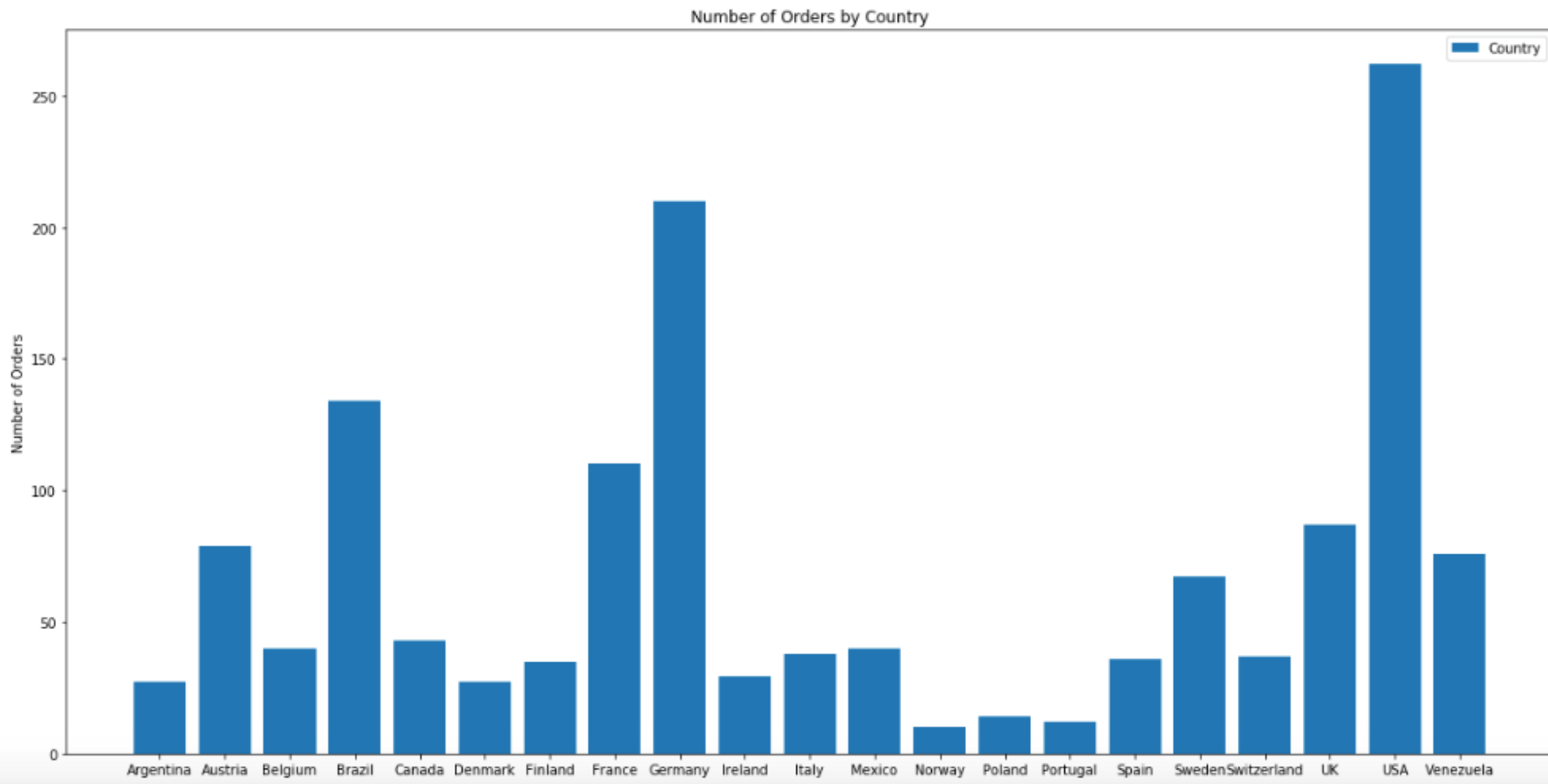
What are the reports do you want

- What are the reports needed?

Number of Customers by Region

Number of Customers by Country

Number of Orders by Region

Number of Orders by Region

Number of Orders by Country

Other Reports

- Show the product (all columns) whose deviation between List Price and Standard Cost is the largest.

- Show the best seller product (which is one being ordered the most so far)

- Compute the average List Price of all products

- Compute the average List Price and the standard deviation of each category

- Count how many employees in each city

- Count how many items in each category

- Show list of product categories with their average prices from small to big

- Show the list of product names that have not been ordered ever.

- Show how many orders each Customer ID has placed.

- Identify the employee ID corresponding to the employee having the highest number of orders

Other reports

- Show the last name of the employee having the highest number of orders

- Compute the subtotal of each Order ID where the subtotal is the sum of values of all the corresponding order lines. Note: the value of each line was mentioned in question 11 of Assignment 3, which is (Quantity * [Unit Price])*(1-Discount)

- Show the Order ID(s) with a subtotal greater than 1500 in the descending order.

- Count how many orders each Employee ID have processed. If an Employee ID has no corresponding order, his results should be 0. A part of the results will look like the below figure.

- Compute the sale of each product. The results should consist of 3 columns: Product ID, Product Name, and Sale.

- Show products who List Price(s) are greater than the average List Price of all products in the table Products.

- Compute the average shipping fee to each city of orders with status "Closed". The resulting rows consist of only average values greater than or equal to 100, and is sorted in ascending sequence.

- Show products who List Price(s) are greater than the average List Price of all products in the same category. The results includes 4 columns: Product Name, List Price, Category, and AVG Price of Category

- Compute how many days in average each Employer ID processes orders (from the Order Date to the Shipped Date). Note: only orders with Shipped Date filled are considered in this query.

- Retrieve information of orders with the following columns: Customer's Company, Employee's Last Name, Order Date, and Order Status Name.

DERIVING THE ERD FOR A HIGHT LEVEL DESIGN

Suggested Tools: draw.io/Visio/Word

1st step: select Entity sets/Properties

- Underlines all the nouns from the business descriptions

- Entity Set: select the nouns you need to manage as a lists, things contains the detail information
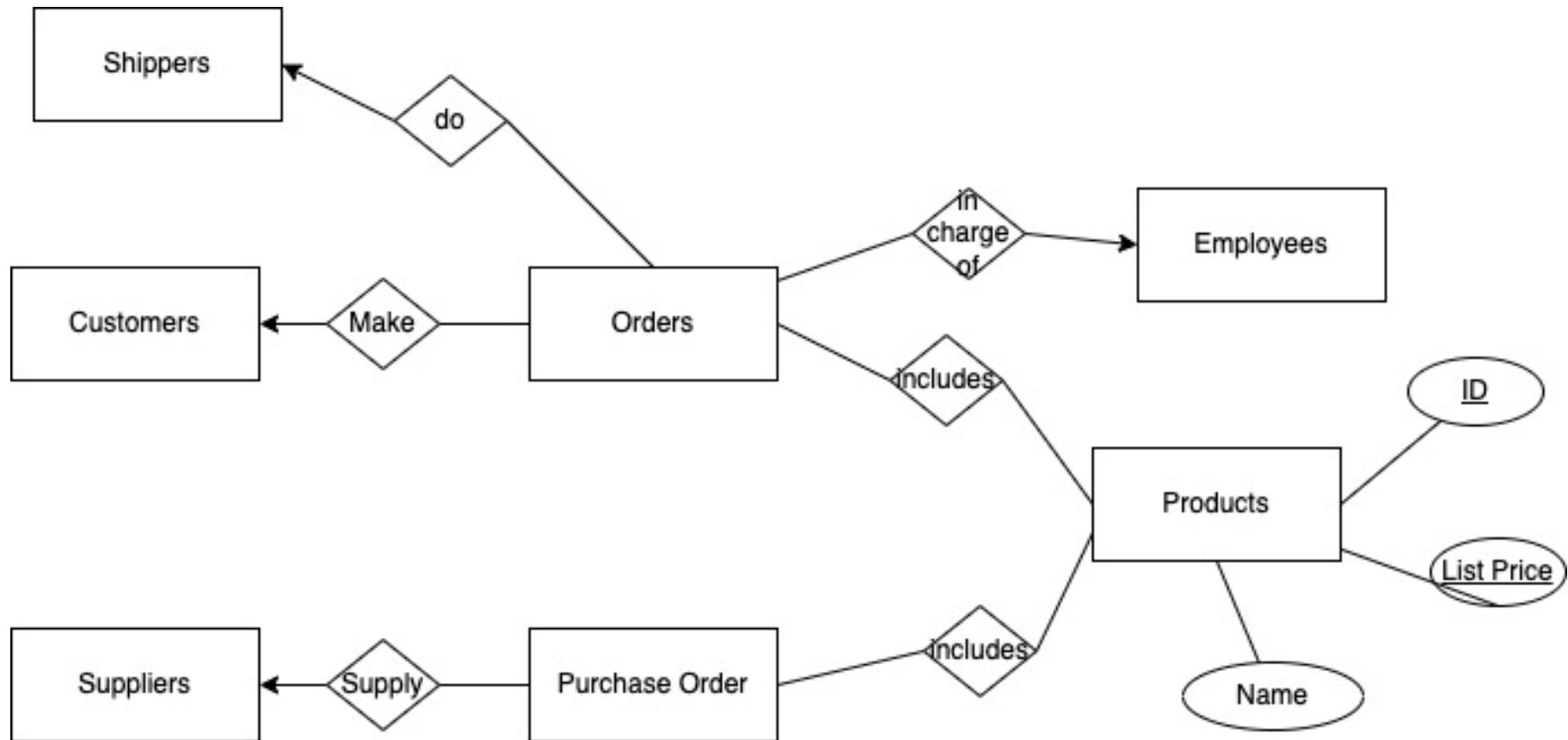
- Properties: simple data belonging to an Entity set

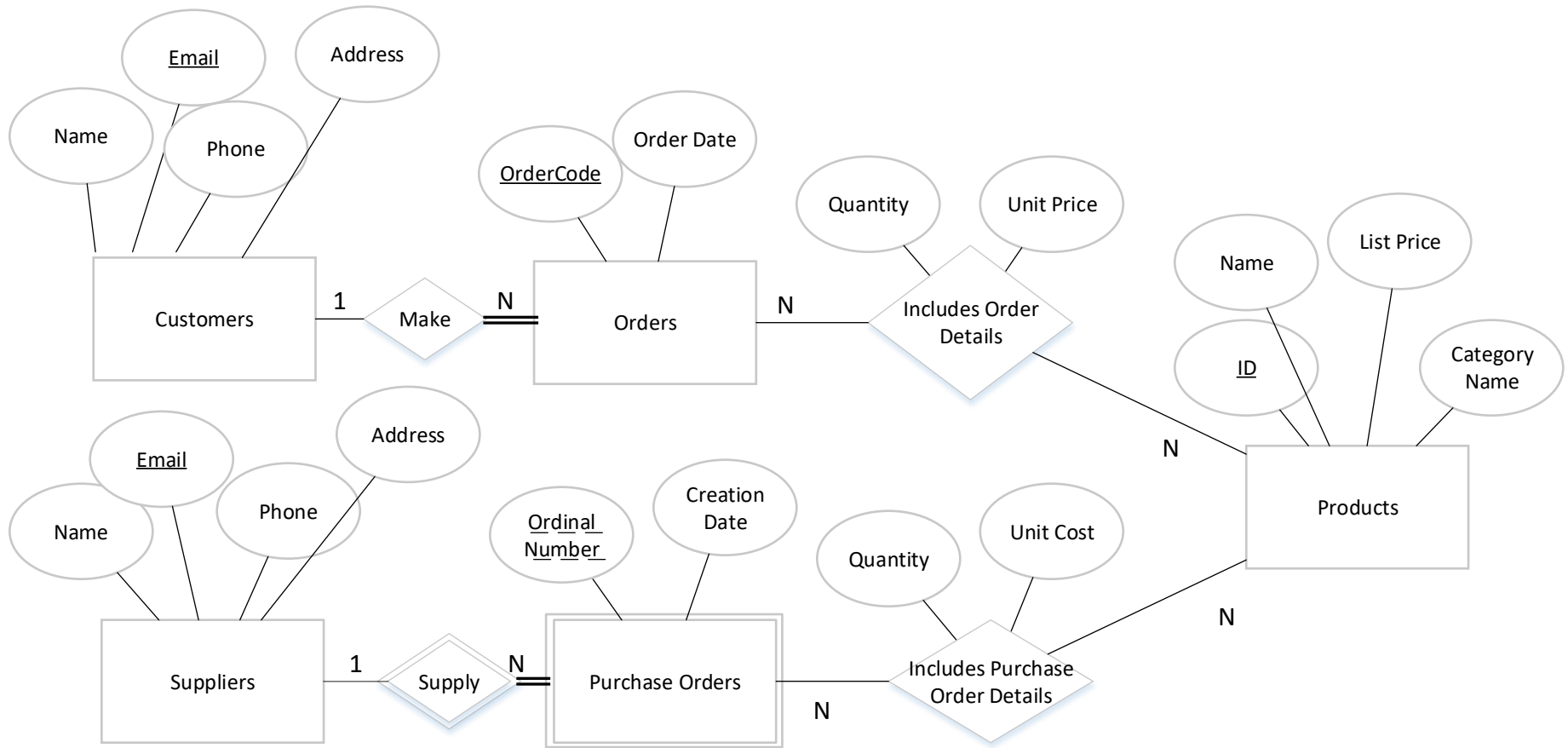# 2nd step: Propose the relationships

- Focus on the verbs in the business decriptions

- Chosing cardinality (1-1, 1-n, n-n) for each relationship

3nd step: refine/revise the ERD

- Decide each Entityset is weak or not

- If weak, should we add an artificial ID to make it strong?

- Test the reports: is this design convenient to build these reports?

41

CONVERT ERD TO DB SCHEMA

Tables

- All entity sets to become tables:

  - Products

  - Employees

  - Suppliers

  - Customers

  - Orders

  - Shippers

- N-n relationships to becomes tables, add attributes of theses relationship to colums of tables

  - Order details (OrderID, ProductID, Quantity, Price, Discount)

  - Purchase order details

What are columns of tables?

- Attributes of entity sets → columns of table

- For 1-n relashipship, use the Primary Key of the 1-side entity set to become Foreign Key of the n-side entity set

  - Eg: Orders (OrderID, CustomerID, …) where CustomerID is the Foreign key

Final database Schema

- List all tables with their attributes and contraints (PR, FK, …)

SQL FOR REAL DBMS

SQL statements

- DDL: already ready for Northwind
  - Define db
  - Define tables with contraints, index if any
  - Other objects: views, procedures, triggner…
- DML
  - How to CRUD: insert, update, delete
  - How to perform queries for require questions
- Write all SQL statement in *.sql files