SIT103/SIT772 Data and Information Management

Week 1

Introduction to Database

Dr Iynkaran Natgunanathan,

email: iynkaran.natgunanathan@deakin.edu.au,
Phone: +61 3 924 68825.

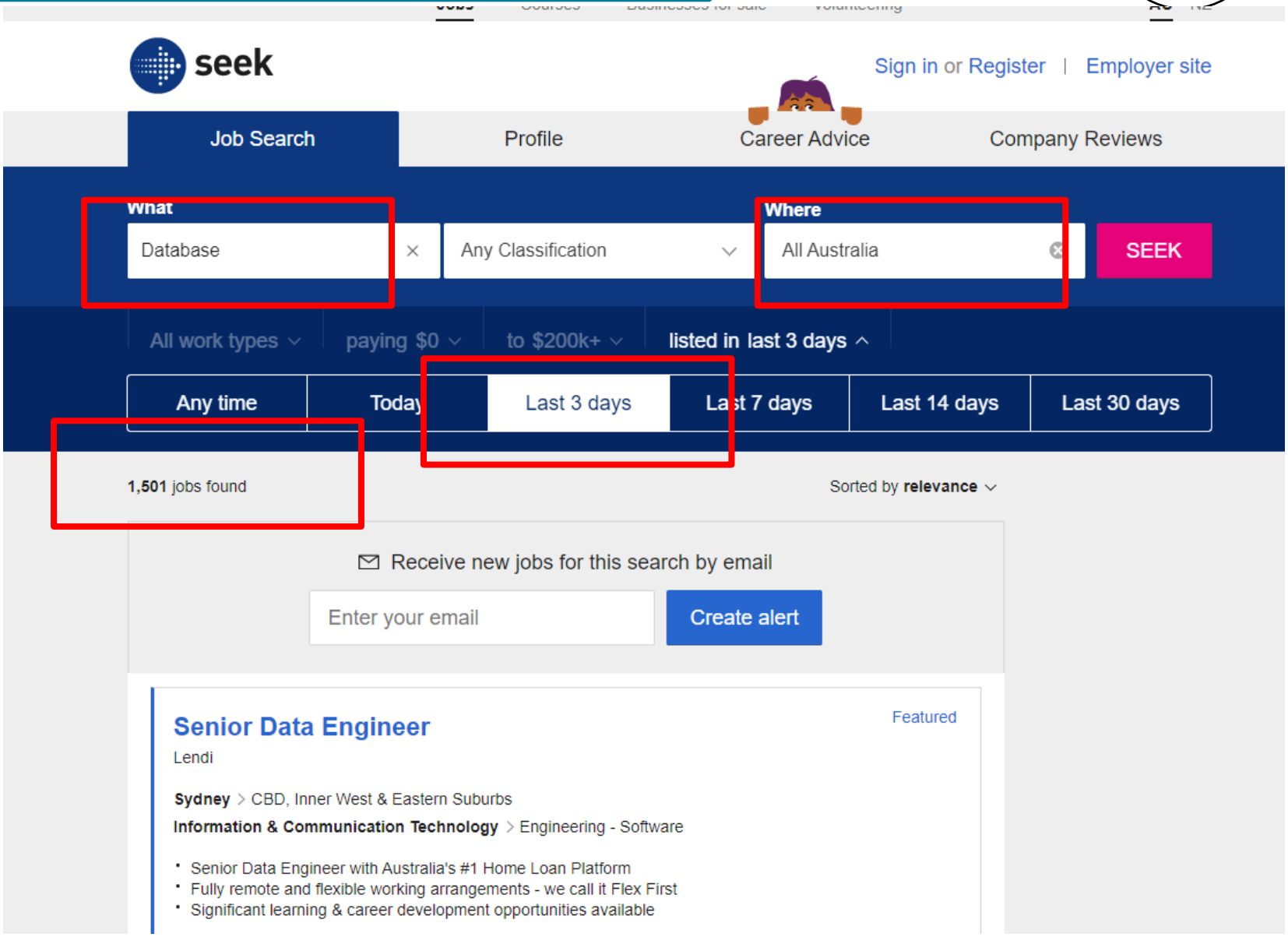
Why are you doing this unit?

- We live in the world of data, data is everywhere around us
  - we use and generate a lot of data every day
- Data are stored in a repository – **Database** or **DB** in short
- Almost all IT systems we use have a database at the backend
  - e.g., facebook, internet banking, cloudDeakin, etc.
- Most systems use **Relational DB** – data are stored in tables
- This unit discusses how to **design**, **implement**, and **manage** a **relational DB**

Database Careers

TABLE 1.3	DATABASE CAREER OPPORTUNITIES	
JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	QL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization

Job opportunities



Unit Learning Outcomes

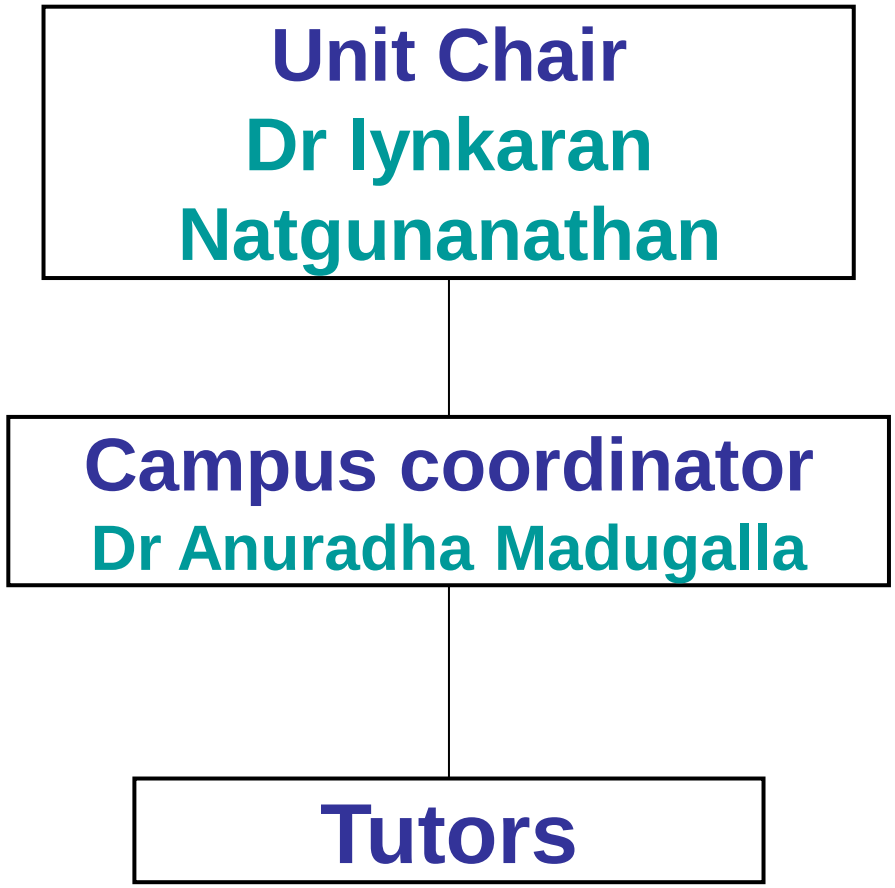
- ULOs are what this unit is about.
- At the end of this unit, students will be able to:

ULO1. Describe the techniques used in storing and retrieving data.

ULO2. Evaluate data models and apply data modelling techniques to capture the data aspects of real-world situations.

ULO3. Design and develop relational databases by using SQL and a DBMS.

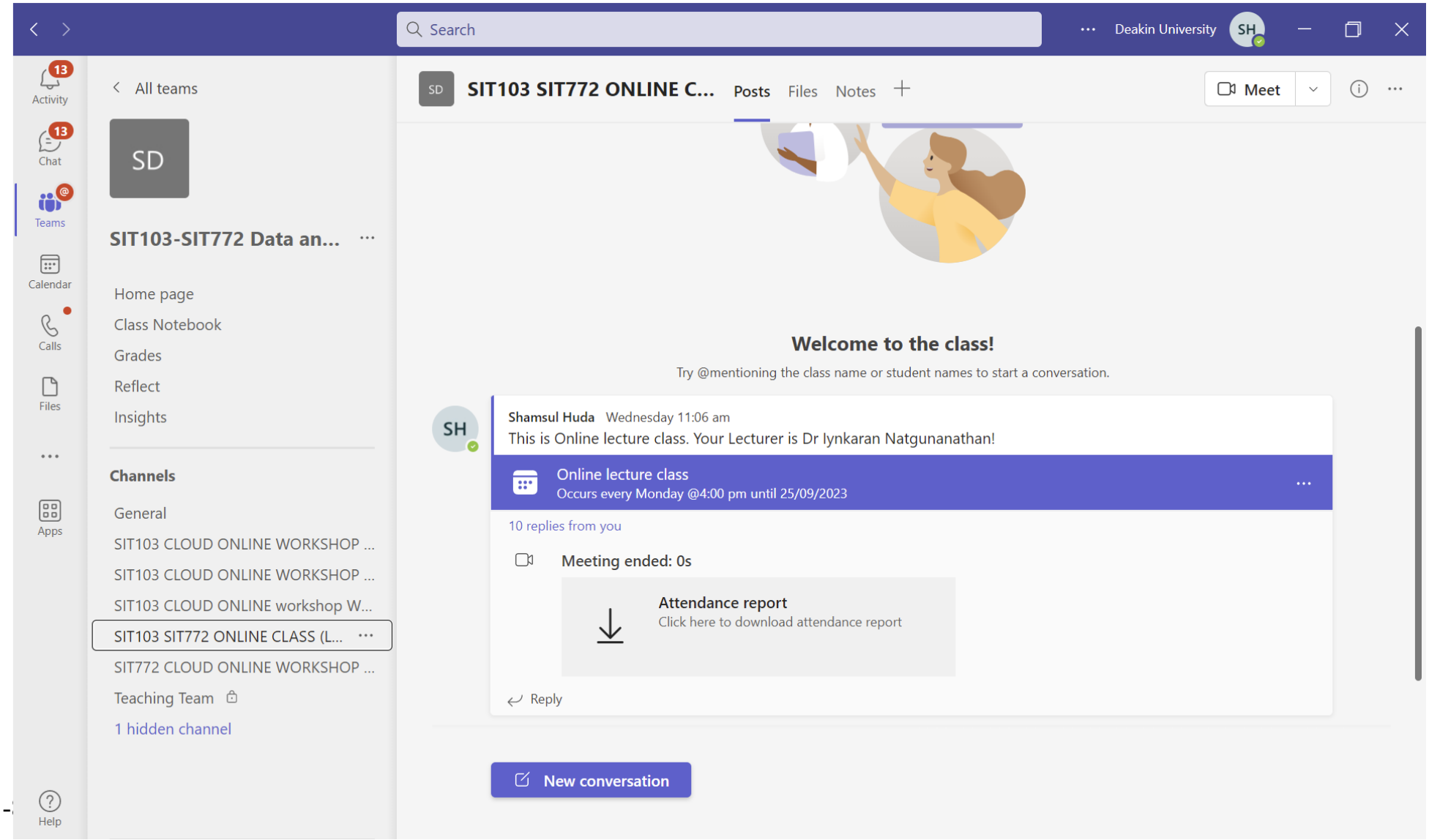
Who you learn with?



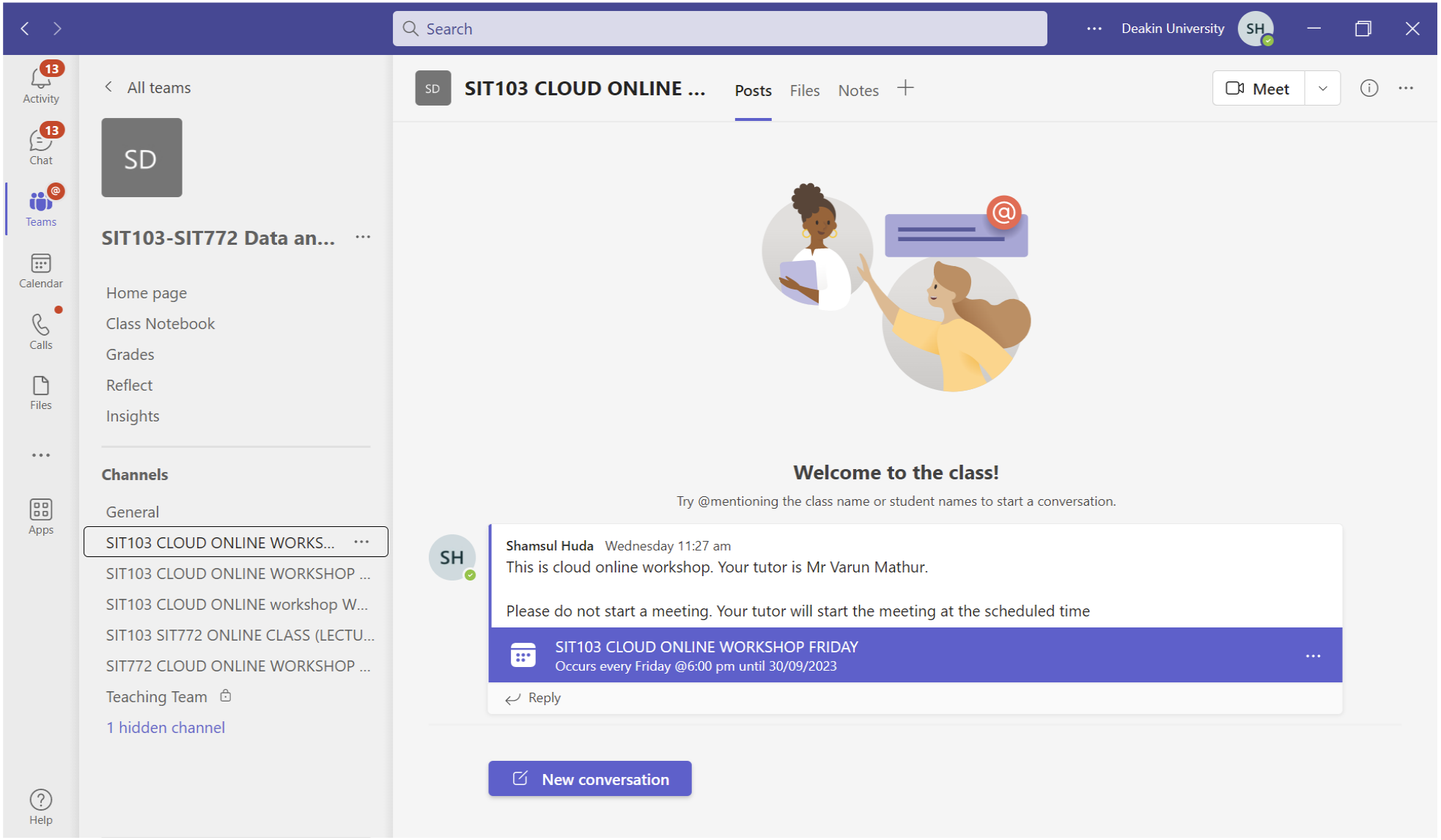
Unit Delivery

- 1 x 2 hrs class per week – online via MS Team
- 1 x 2 hrs workshop per week – on-campus/online via MS Team
  - **On-campus students – on-campus workshops**
    - Please enrol accordingly via STAR
  - **Cloud students – online workshops in MS Teams**
    - Feel free to join a session that suits you

Unit Delivery-ONLINE MS TEAM CLASSES



Unit Delivery-ONLINE WORKSHOPS



Classes and Workshops

- Active discussion rather than just content delivery

  - expect you to go through the content before class/workshops

  - contribute to the discussion

  - ask questions to clarify your doubts

- Helps in achieving ULOs and completing assessments

- Class – more conceptual/theoretical discussions

- Workshop – hands-on activities and implementation

Communication

- MS TEAM UNIT CHANNEL
  - for all teaching, content, assessment related queries
  - your question and our reply will be useful to others too

--POST YOUR MESSAGES IN MS TEAM CHANNEL ONLY

- we monitor MS Teams only during online class/workshops

Emails/MS Teams direct messages

- private/confidential discussions (pls have **SIT103/SIT772** in the subject line)

- We endeavour to reply you within 3 business days.
- **Please keep all communications respectful!**

Unit's Content Overview

DEAKIN UNIVERSITY

- Database overview
- Database design
  - Understanding business requirements
  - Conceptual and logical models
  - Data models (Relational models, Entity relationships models)
  - Relational algebra
  - Normalization
- Implementation & management of database using MySQL
  - Structured Query Language (SQL) using MYSQL
- Case studies based on database design and development
- Business Intelligence and Data Security

Assessment

- No Final Examination or Assignments
- Portfolio using OnTrack
- **You target a grade and complete weekly tasks accordingly**

Pass – Scaffold the concepts learned (discussed in workshops)

Credit – Interpret/Explain the concepts learned (some support)

Distinction – Translate concepts to any problems (self-directional, minimum support)

High Distinction – Go/Extend beyond the unit scope (Aspirational, sky is the limit, we only guide you)

- Mark and Grade based on your **Final learning portfolio –** all tasks you complete during the trimester + your learning summary report at the end

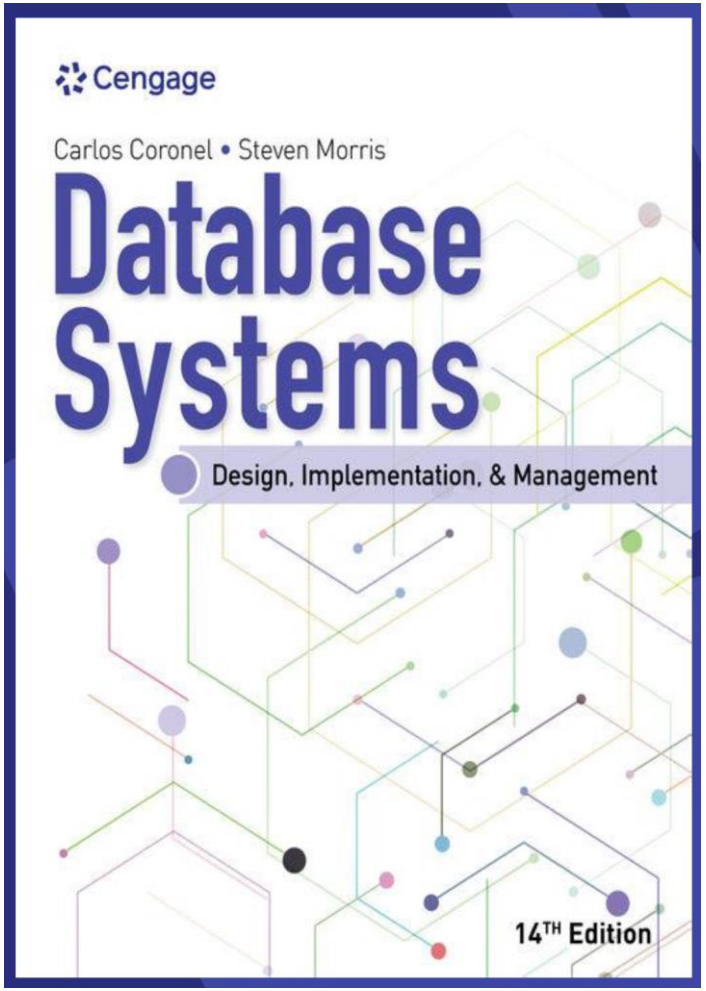
Academic Integrity

- See Academic Integrity slides available on the unit site under week 0 resources
- Anyone using cut-and-paste or copying of other people's work will be easily identified by Turnitin and the markers.
- The outcome of such actions will be a disciplinary committee hearing which can have very serious outcomes.
- Contract cheating? DON'T DO IT, see here https://blogs.deakin.edu.au/deakinlife/2018/09/18/dont-ruin-your-career-dont-contract-cheat/

Unit CloudDeakin Site

- Homepage - SIT103_SIT772 - Data and Information Management (deakin.edu.au)
- Lecture materials will progressively be updated and released as we go through the trimester
- All materials will be placed in the CloudDeakin unit site.
  - Unit Guide (already there)
  - Resources
  - Discussions
- OnTrack tasks will be released progressively as we go.

Prescribed Text



Database Systems: Design, Implementation, & Management

14th Edition

Coronel and Morris

Cengage publisher

E-book available through the library

10% discount for students

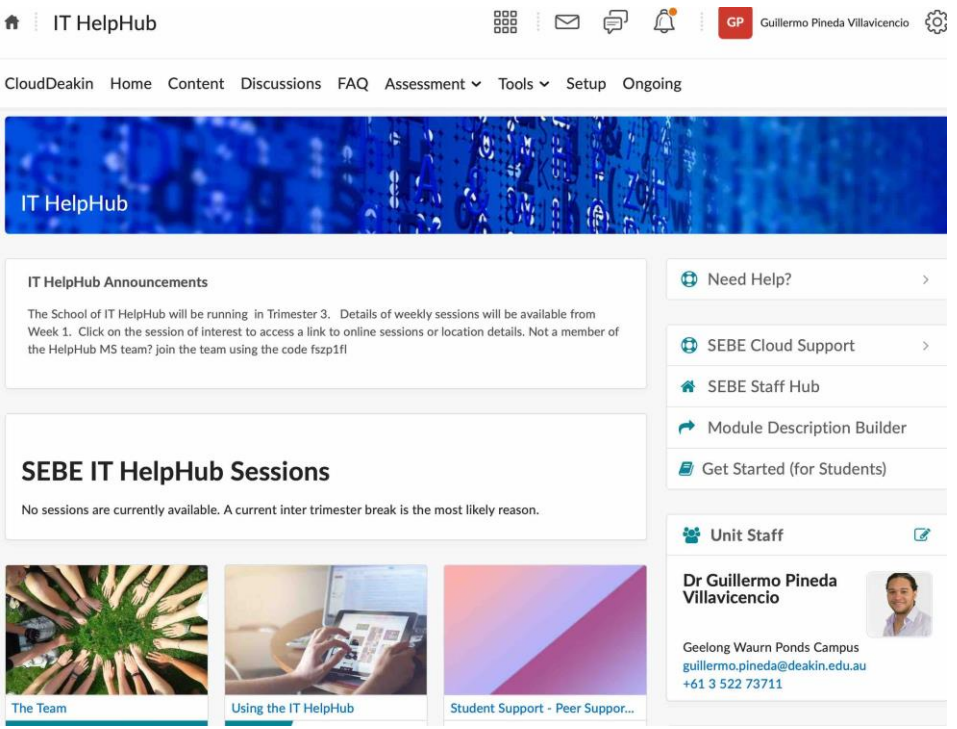
Discount code: WOW10

Help?

- For many of you, this is your first time at uni
  - Uni study is different from high school study
  - You are in-charge of your learning
  - You need to go beyond what is discussed in class/practicals
  - The transition can be difficult/challenging
- There are helps available, just seek for those
  - Unit chair and me
  - Other members of the teaching team (your tutor, campus coordinator)
  - Other services (S**IT Helphub**, Library, Peer Support, DUSA, student central)

SIT HelpHub

- The SIT HelpHub suppor students enrolled in SIT units by:
  - ✓ Answering your questions
  - ✓ Showing you where to find information
  - ✓ Demonstrating how to solve problems and understand concepts that will help you with your assessments and the unit in general
- The HelpHub is supported by tutors and volunteers, who are selected by your unit chairs and are experts in your unit topics.
- Regardless of your enrolment, you can join any helphub session.
- Details of Help Sessions are available on our [helphub site]



Any questions so far

About unit administration and management?

Why Databases?

- We use them in our everyday life
  - Google to search for information on the Web
  - Supermarkets or online stores to buy something
  - ATM machines to withdraw money from bank
  - Library catalogues to look for books
  - Etc.
- Almost all modern business systems **rely on databases**.

Why Databases?

A Day In Susan's Life

See how many databases she interacts with each day

Before leaving for work, Susan checks her Facebook and Twitter accounts

Where is the data about the friends and groups stored?

Where are the "likes" stored and what would they be used for?

On her lunch break, she picks up her prescription at the pharmacy

Where is the pharmacy inventory data stored?

What data about each product will be in the inventory data?

What data is kept about each customer and where is it stored?

After work, Susan goes to the grocery store

Where is the product data stored?

Is the product quantity in stock updated at checkout?

Does she pay with a credit card?

At night, she plans for a trip and buys airline tickets and hotel reservations online

Where does the online travel website get the airline and hotel data from?

What customer data would be kept by the website?

Where would the customer data be stored?

Then she makes a few online purchases

Where are the product and stock data stored?

Where does the system get the data to generate product "recommendations" to the customer?

Where would credit card information be stored?

Figure/table is from Coronel & Morris (2018)

Data versus Information

- Data consists of raw facts
  - Not yet processed to reveal meaning to the end user
  - Building blocks of information
  - Usually stored in databases
- Information
  - Produced by processing raw data to reveal meaning
  - Requires context
  - Bedrock of knowledge
  - Reveals the **meaning** of data
  - Enables **knowledge creation**
  - Should be accurate, relevant, and timely to enable decision making

Data vs. Information (2)

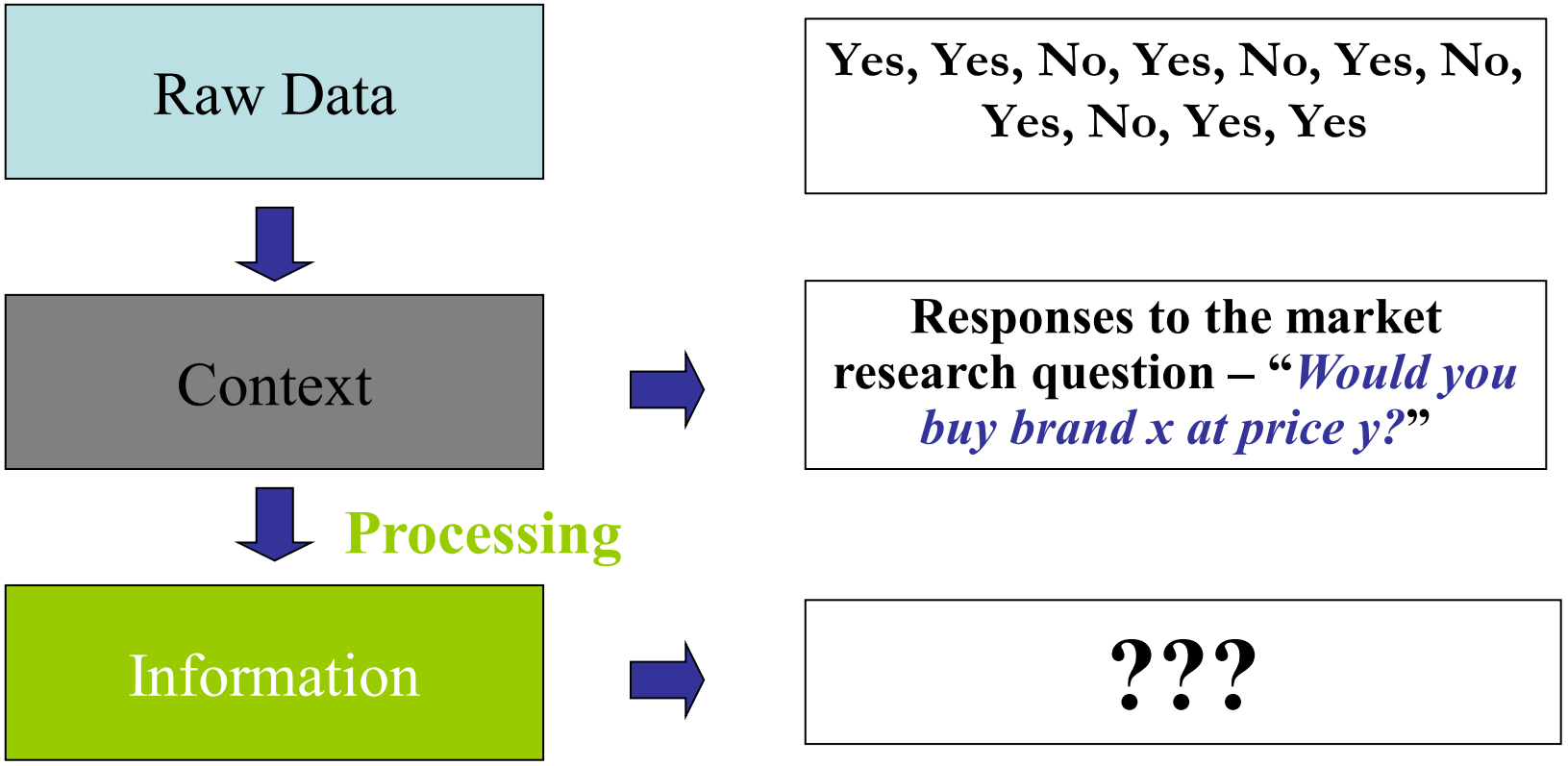
- Student Marks (data or information?)

Student_ID	Name	Major	Marks Assignment-1, Q1	Marks Assignment-1, Q2
8912345	Lewis, A.D.	MG	10	6
9023456	Baker, G. P.	CS	9	9
9134567	Hunter, S. L.	IS	7	2
9145678	Grant, G. D	CS	9	10
…	…	…	…	

- Unit profile, a summary report (data or information?)

Grade	%	No. of Students
HD	12	15
D	17.6	22
C	28.8	36
P	31.2	39
N	10.4	13

Data vs Information (3)



Why Databases? (2)

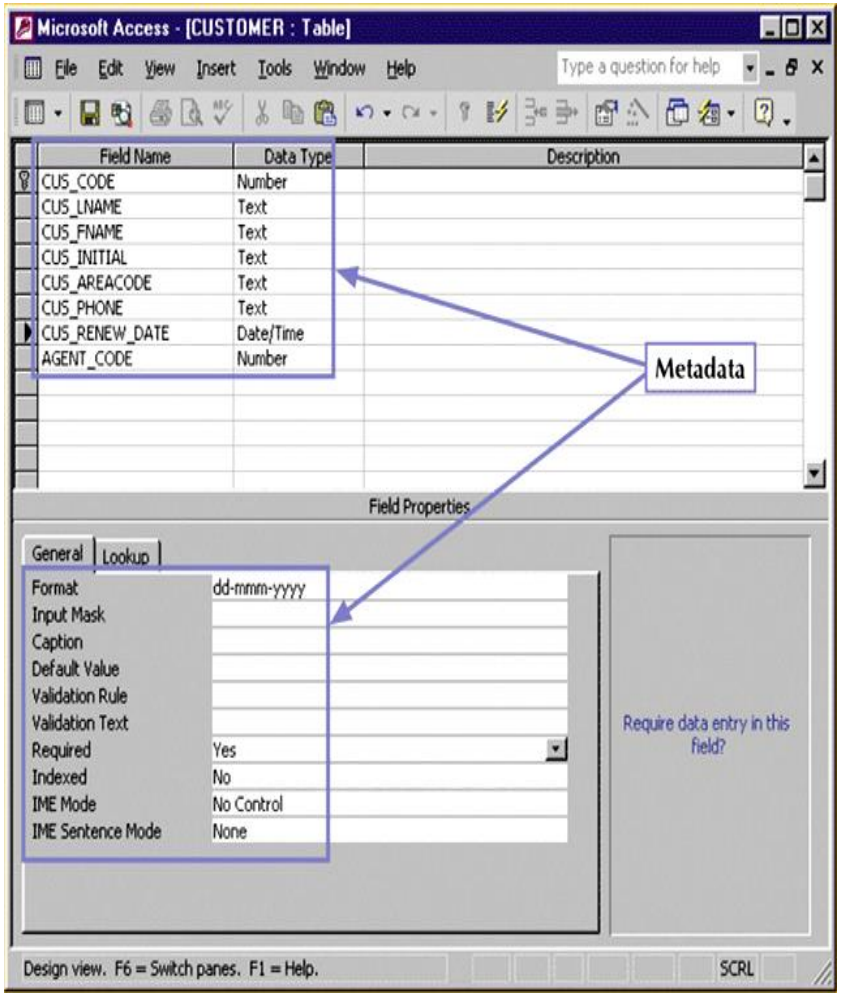
DEAKIN UNIVERSITY

- Characteristics of data in today's world
  - ubiquitous (i.e., abundant, global, and everywhere)
  - pervasive (i.e., unescapable, prevalent, and persistent)
- We generate and consume a lot of data every day
  - started from our birth – birth certificate
- Essential for businesses to survive and prosper
  - Collection, storage, aggregation, manipulation, dissemination, and management of data (e.g., products, transactions, customers, etc.)
- Databases make data persistent and shareable in a secure way
  - Specialized structures that allow computer-based systems to store, manage, and retrieve data very quickly

What is a Database?

- A shared, integrated structure that stores data.
- Two types of data:
  - **End user data** (raw facts of interest)

    1010, Larson, John, J.L., 02, 42514987, 02-09-2021, 5
  - **Meta-data** (data about data)
    - through which the end-user data is integrated and managed
    - Describes data characteristics and relationships



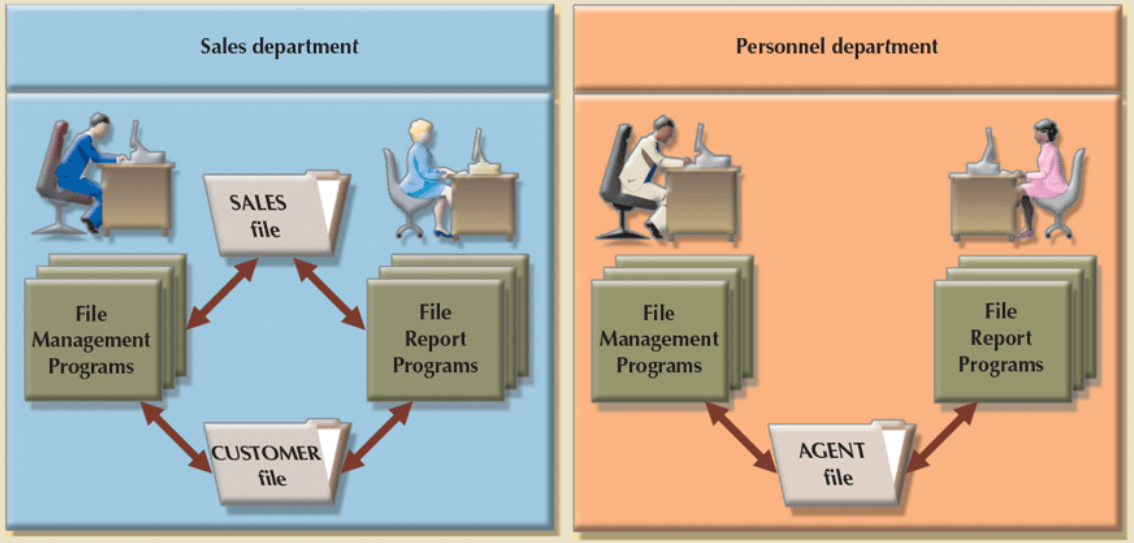
Types of Databases

User Type	Single-user, Multi-user (Workgroup, Enterprise)
Location	Centralised, Distributed, Cloud
Data Usage	Operational (a.k.a. transactional or production), Analytical (Data Warehouse)
Data Type	General-purpose, Discipline-specific
Data Structure	Structured, Semi-structured, Unstructured
New Type	NoSQL (Non SQL), not the traditional database, NoSQL is the name given to a broad array of non-relational database to handle (e.g. social media on the Internet) - Unprecedented volume of data - Variety of data types and structures - Velocity of data operations

Evolution of data storage

- Manual file systems
  - Accomplished through a system of file folders and filing cabinets
- Computerized file systems
  - Data processing (DP) specialist created a computer-based system to track data and produce required reports
- Database and DBMS

FIGURE 1.9 A SIMPLE FILE SYSTEM

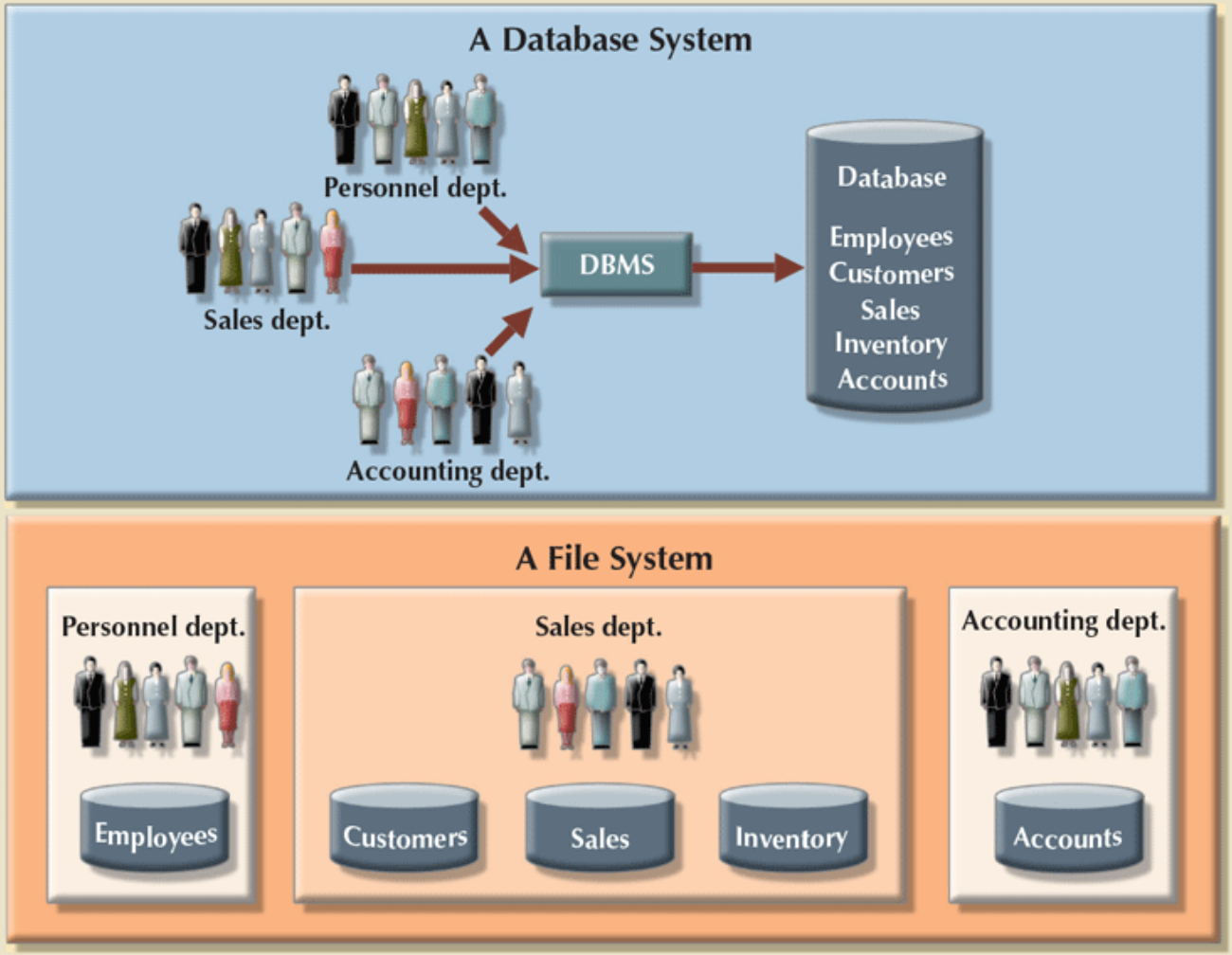


Issues with file systems

- Difficulty of getting quick answers
- Complex system administration
- Lack of security and limited data sharing
- Data redundancy - unnecessarily storing the same data at different places
  - Islands of information (i.e., scattered data locations)
  - Increases the probability of having different versions of the same data

Database vs File Systems

FIGURE 1.10 CONTRASTING DATABASE AND FILE SYSTEMS



Database Management System

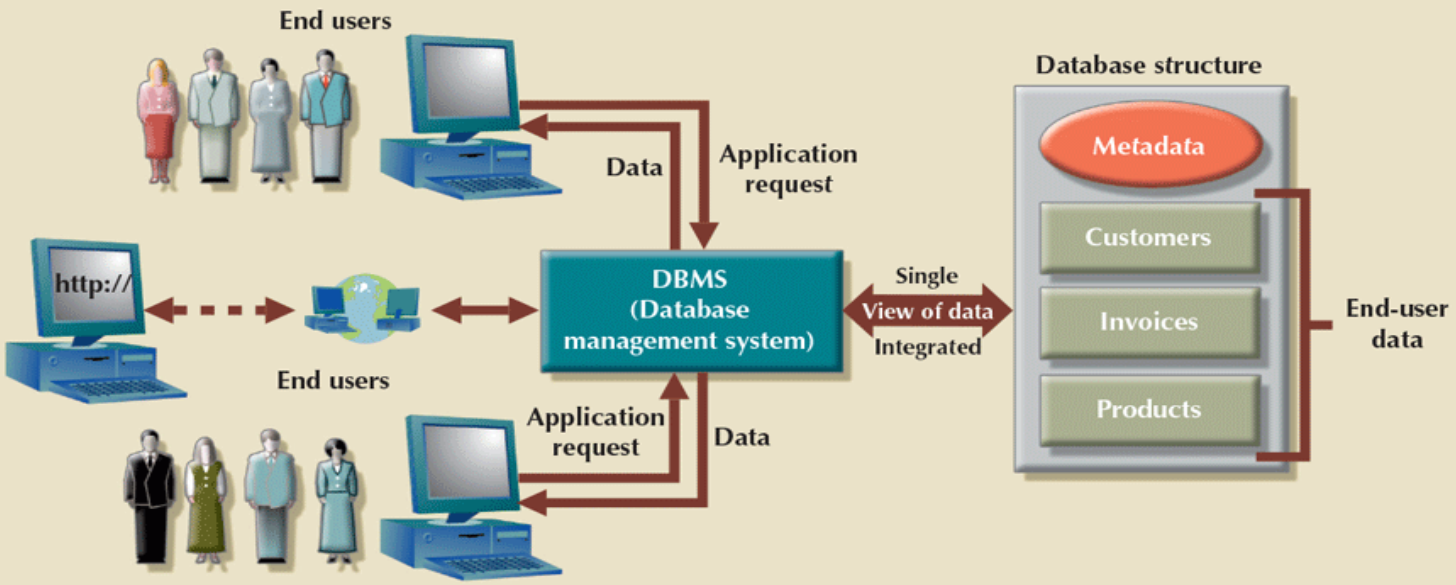
- DBMS = Database Management System
- A collection of programs that
  - **manage** database structures
  - **control** access to data stored in the database
  - **facilitate the sharing** of data among multiple users and applications
  - intermediary between the user and the database
  - Presents the end user with an integrated view of data
  - Provides more efficient and effective data management
  - Improves sharing, security, integration, access, decision-making, productivity, etc.

DBMS

The DBMS manages the interaction **between the end user and the database**

- The DBMS receives all application requests and
- Translates them into the complex operations required to fulfill those requests.
- The DBMS sends back an answer (result set) to the application.

FIGURE 1.4 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE



Figure/table is from Coronel & Morris (2018)

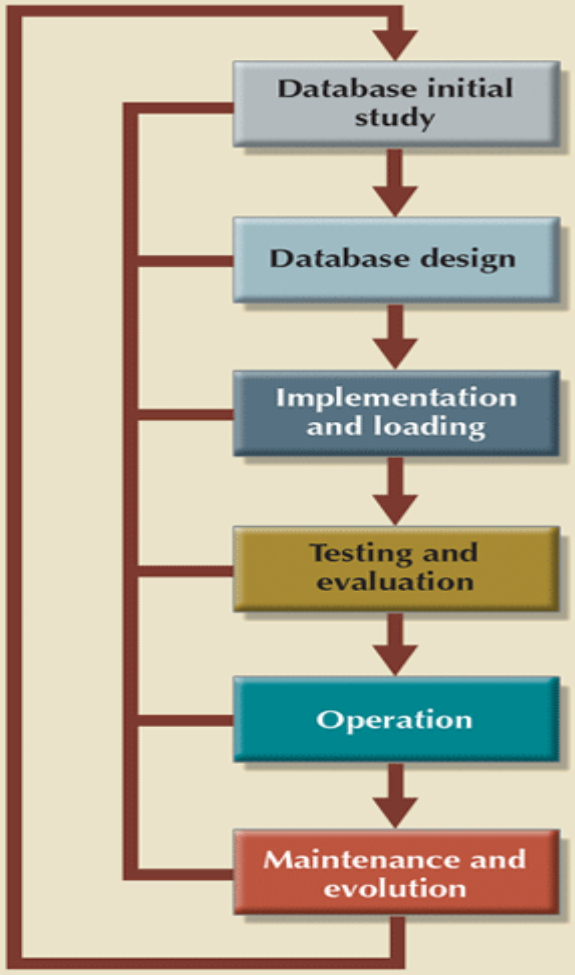
The Database Life Cycle

- The Database Life Cycle (DBLC): A cycle that traces the history of a database within an organization's information system.
- DBLC contains six phases
  - Database initial study (understanding data requirements)
  - Database design
  - Implementation and loading
  - Testing and evaluation
  - Operation
  - Maintenance and evolution

The Database Life Cycle

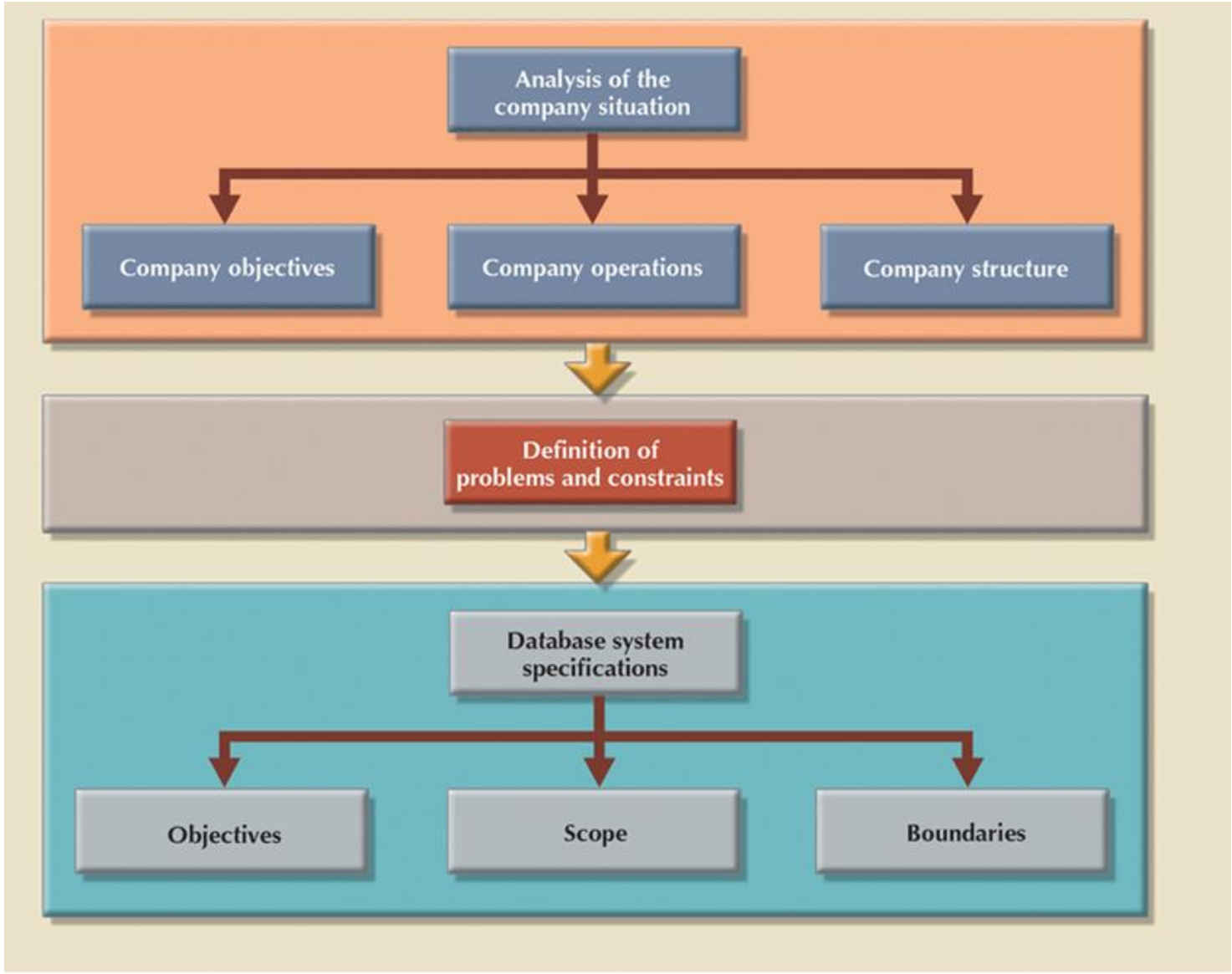
Figure/table is from Coronel & Morris (2018)

FIGURE 9.3 THE DATABASE LIFE CYCLE (DBLC)



Action(s)	Section
Analyze the company situation Define problems and constraints Define objectives Define scope and boundaries	9-3a
Create the conceptual design DBMS software selection Create the logical design Create the physical design	9-3b
Install the DBMS Create the database(s) Load or convert the data	9-3c
Test the database Fine-tune the database Evaluate the database and its application programs	9-3d
Produce the required information flow	9-3e
Introduce changes Make enhancements	9-3f

The Database Initial Study

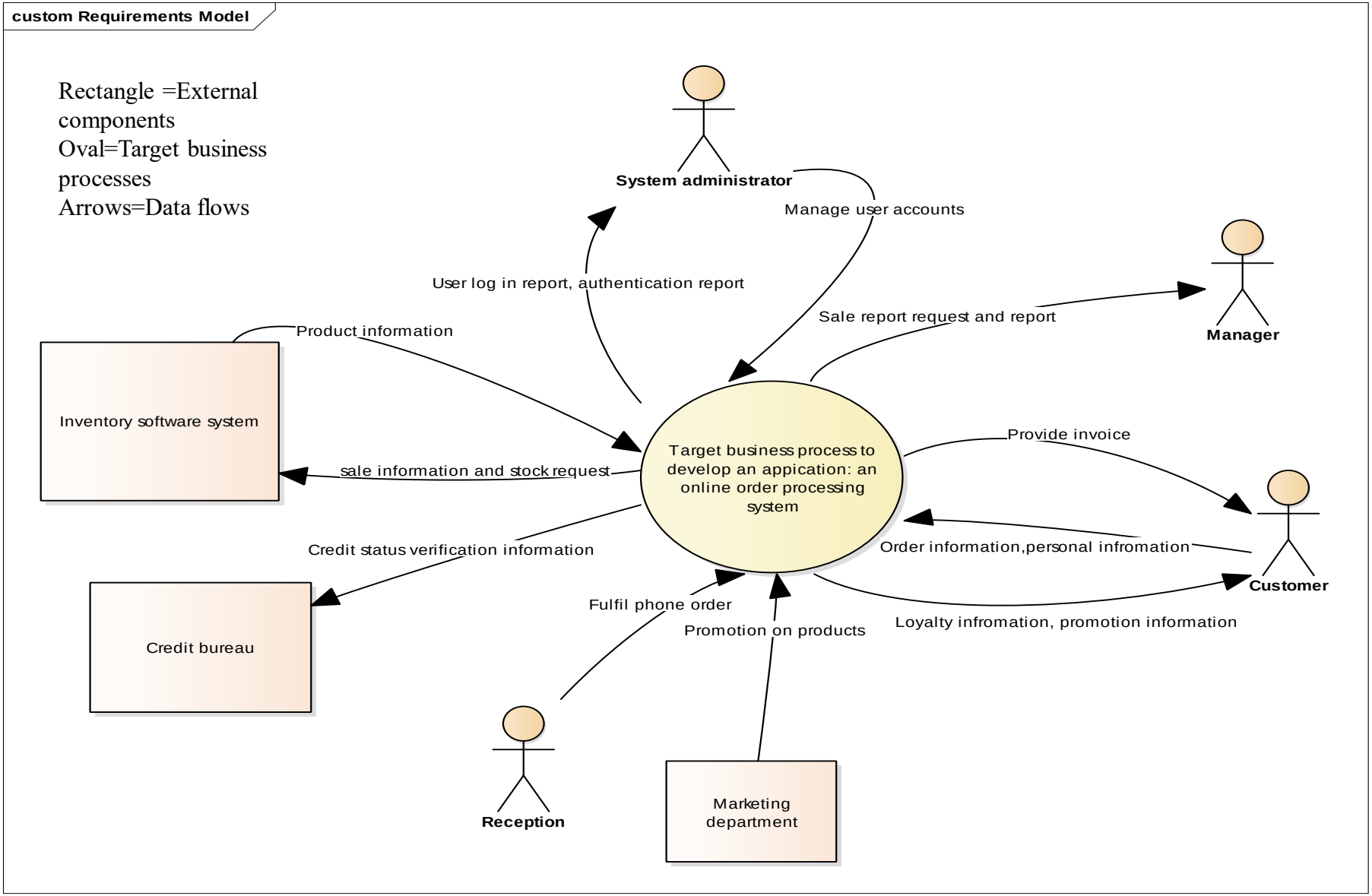


Initial study

- Understanding data requirements of a system/business
  - Examines the current system's operation within the company,
  - how and why the current system underperforms or fails.
  - Interview the company people
  - Read company documents
  - Read reports
  - Physically observe its steps of business process
- define problems (what are the problems in current systems),
- Constraints (hardware, software, Budget)
- Objectives (What the new system should do, purposes),

Understanding system's data requirements

Context Level Data Flow Diagram (DFD) of an online order system



Context level DFD

- Shows flow of information (input/output) to/from the system
- helps to identify data that system has to store/maintain to meet the needs of external systems/users
- In this example of order processing system:
  - Customer information
  - Product information
  - Loyalty information
  - Promotion information
  - Order information
  - Sales information
  - Financial information
  - Stock/inventory information

Database Design

- Start thinking about how to store and manage the data that system has to maintain
- What exactly is required to be stored for each data component/entity
  - **Customer**: Name, Address, Username, password, Last login, Credit card
  - **Product**: Name, Category, Price, Stock level, Supplier, Supplier Address

  Etc.
- Develop a structure that links different data components (e.g., Customer, Product, Order, etc.) together
  - Customers place Orders and Orders include Products

Database Design (2)

- Focuses on the **database structure** that will be used to store and manage data
- A database that meets all user requirements does not just happen;
  its structure must be designed carefully
- An easy-to-use DBMS **does not mean** a good database design
- Even a good DBMS will perform **poorly with a badly designed DB**

Goals of DB design

- General goals of DB design
  - Avoid **redundancy**
  - Provide efficient but controlled **access** to data
  - Enable a **fast response** to a query
- **Well-designed database**: facilitates data management and generates accurate and valuable information
- **Poorly designed database:** causes difficult-to-trace errors that may lead to poor decision making

Week 1 Summary

- Data versus Information
- What is a database and why it is important?
- User data and meta data
- File Systems vs Database
- Database life cycle
- Understanding system's data requirement

  - Context level DFD

- Database design and its importance

This Week's OnTrack Tasks

- **1.1P** Reflection on three data-driven information systems you use in your daily life

  - What are the systems, where the data comes from, what would have happened if that system/data was not available to you?

- **1.2P** Installing and setting up MySQL Environment

  - MySQL community Server

  - MySQL Workbench

- Please check the task sheets and start working on them.

Thank you

- **Any questions/comments?**

- Workshop starts from this Week!
- Please make sure you are allocated to at least one workshop session
- Online workshops are available in MS Teams
  - Strictly for cloud students only
- On-campus students, please enrol to on-campus workshop and come to uni
  - we want to see life back at uni

See you next week!

More on Database Design

DEAKIN UNIVERSITY

Next Week

Making sure that the final product meets user/business requirements

- Conceptual design
- Logical design
- Physical design
- Relational Model

Readings and References:

- Textbook Chapter 1

Database Systems : Design, Implementation, & Management 14TH EDITION, by Carlos Coronel (Author), Steven Morris (Author)