



INFO90002

Database Systems & Information Modelling

Week 12

Pre-exam wrap-up

“The fundamental move from the cinematic form to the database form is a move from narrative and linearity to relationality. The database has no inherent narrative structure; there is no natural sequentiality to objects in a database. What structure the database provides is one of relations – between objects and their properties, from object to object, and so on.”

- Paul Dourish (2017) *The Stuff of Bits*. p106

Learning about databases is essential because databases are:

- ubiquitous
 - there is a database behind most of today's exciting technologies
- standard
 - 30+ years of standardization mean that the relational database skills you learned will be applicable in any IT career
- starting to change
 - but recent challenges to the relational standard are creating significant opportunities for new players



- Conditions
 - see <https://students.unimelb.edu.au/admin/exams>
 - closed book = don't bring anything except pens and student card
 - 15 minutes of reading time + 3 hours writing
 - covers a range of material from week 1 to week 12
 - about half is on Data Modelling and SQL (weeks 1-6)
 - the rest is on other topics (weeks 7 to 12)
- Hurdle requirement:
 - to pass the subject students must obtain at least:
 - 50% of the marks from non-exam based assessment
 - 50% of the marks from the exam
 - <https://handbook.unimelb.edu.au/2019/subjects/info90002/assessment>

Exam page 1

School of Computing and Information Systems
End of Semester 2 2019 Examination

INFO90002 DATABASE SYSTEMS & INFORMATION MODELLING

Reading time: 15 minutes

Writing time: 180 minutes

This paper has 5 pages, including this page.

Authorised Materials:	No materials are authorised.
------------------------------	------------------------------

Instructions to Invigilators:

The examination paper IS TO REMAIN in the examination room.
Students are to be provided with standard answer books.

Instructions to Students:

Ensure your student number is written on all script books.

The total mark for this paper is 70 marks, representing 70% of your final assessment.

This exam paper has 9 questions, some with multiple parts. Attempt all questions.

Answer all questions on the *lined* pages of your answer book. The *unlined* pages of answer books are for draft working and notes and will not be marked.

If you fill your first answer book, you may ask for an extra one.

Start the answer to each numbered question on a new page in the answer book.

Write legibly in blue or black pen.

Textual answers can be in point form.

All electronic devices (including mobile phones, watches and calculators) must be switched off and remain under your desk throughout the examination.

No items, including electronic devices, may be taken to the bathroom.



Reading time

(lecturer is present to
answer questions)

Writing time (lecturer not present)



Material examined

Data Modelling
and SQL

further topics

Week	Lecture Hour 1	Lecture Hour 2	Lecture / Tutorial	Hoffer Chapter	Extra reading	Assessment
1	Intro to Subject	Intro to Databases	Designing a Database	1. Database Environment	Hoffer video History of Database	
2	Data Modeling 1	SQL 1	Tutorial MySQL Workbench Setup SQL Set up Scripts Tutorial Solution	2. Modelling data 6. Introduction to SQL	SE Radio 'Relational Databases'	Assignment 1 Released
3	Data Modelling 2	SQL 2	Tutorial (DM/SQL 1) Tutorial Solution	4. Logical Design & Relational Modeling	Simsion ch 1 , Hoffer video	
4	Data Modelling 3	SQL 3	Tutorial (DM/SQL 2) Tutorial Solution	7. Advanced SQL	Simsion chapter 3 , Hoffer video	
5	Normalisation	Physical Design	Tutorial (SQL 3) Tutorial Solution	5. Physical DB Design	Kent (1983) Normalization Hoffer video MySQL data types	Assignment 2 Released
6	Data Dictionaries	wrapup and Q&A on part 1	Tutorial (normalization) Tutorial Solution	1. section on Metadata	example data dictionary	Assignment 1 Due midnight Friday
7	Databases in Applications	Web Applications	Tutorial (SQL 4) Tutorial Solution	9. Application Environment	Multi-tier architecture O'Reilly video: Intro to Web	
8	Transactions & Concurrency demo script	Distributed Databases	Tutorial (SQL 5) scott.sql Tutorial Solution	13. Distributed Databases	WP on transactions Discussion on distributed databases	Assignment 2 Due midnight Friday
9	Database Administration	continued	discuss Asst 1	12. Data and Database Administration	MySQL database administration Oracle database administration	
10	guest: Martin Tomko on GEOM90018 Spatial Databases	guest: Rao Kotagiri on COMP90050 Advanced Database Systems	discuss Asst 2			
11	NoSQL Databases	NoSQL continued	Revision 1: you choose the topics		How Facebook stores data Martin Fowler on NoSQL	
12	Industry Trends	Wrapup and exam prep	Revision 2: you choose the topics			



- Half (47%) = Data modelling and SQL
 - draw an ER diagram
 - write some SQL code
- Half (53%) = other material
 - short answer questions
 - must write in black or blue pen
 - DON'T use a red pen
 - pencil is only for drafting, not the answer to be marked
 - examiners must be able to read your answers clearly
 - Since there are total 70 marks,
and you have 3 hours writing time,
you should budget about $180/70 \approx 2.5$ minutes per mark.

Short-answer questions: How much should I write?

- The exam does not prescribe a set quantity of text to write per question.
- We suggest that you budget your time according to the marks that each question is worth.
- Question words like “*list*”, “*describe*” and “*explain*”, in combination with the marks allotted to a question, give an indication of the amount of detail you are expected to provide.
- This is illustrated in the sample on the right.

Q. List three relational database vendors

(1 mark)

- Oracle
- Microsoft
- IBM

Q. Describe three relational database vendors

(3 marks)

- Oracle Corporation, based in California, sell a range of software including their well-known database management system.
- Microsoft, based in Washington state USA, is the largest software manufacturer in the world, and offer a RDBMS called 'Sql Server'.
- IBM or "International Business Machines" began in 1911 and sell the relational DBMS "DB2".

Q. Describe the relational database vendor "Oracle", explaining its success.

(6 marks)

Oracle is the second-largest software company in the world, second only to Microsoft. The company was founded in 1978 by Larry Ellison, who is still Oracle's CTO. "Oracle Database" has the largest market share of any relational DBMS.

A number of factors contribute to Oracle's success, including:

- It offered the first commercial RDBMS
- It has been a frequent innovator of new RDBMS features
- It offered RDBMS that ran on emerging Unix servers in the 1980s
- It offers a range of products + services including application software and training courses
- It has adapted to changes in the I.T. industry such as cloud + NoSQL.

In 2010, Oracle acquired Sun Microsystems and thereby became the owner of Java, MySQL, and a Unix server OS.

Oracle's corporate campus at Redwood Shores features buildings designed to look like hard disk drives.

Drawing ER diagrams by hand: what style should I use?

- [from the exam ...]
- “Use crows-foot notation for relationships, and join the lines to the related columns. Show the cardinalities of relationships and whether they are optional or mandatory. Don’t add surrogate PKs when other attributes exist that can form a suitable PK. You do not need to draw “dotted” lines. You do not need to add names to relationships. You do not need to write a data dictionary: only your ER diagram will be marked.
- Note that marks are largely based on a workable model that enforces all the constraints stated in the case. Marks may be lost for incorrect entities, attributes, relationships, cardinalities, data types or notation, lack of detail, or internal contradictions. “

Drawing ER diagrams by hand: what style should I use?

- You can use this style to draw an ER diagram on paper. This contains most of the information you'd find in a Visio or Workbench diagram, except:
 - No need to worry about nullable/mandatory columns
 - No need to worry about dotted lines for non-identifying relationships

Summary of the course

- the following slides present a summary of the course
- they include *example* exam questions, to help you study
- these are probably NOT the actual exam questions!



- Weeks 1-2
 - Brief history of data storage and processing
 - paper, punch cards, computer files, relational databases
 - Overview of relational databases
 - tables, SQL language
 - Advantages of relational db over file processing
 - Data independence, Minimise redundancy, Improved consistency, Improved sharing, Reduced program maintenance, Better application development productivity, Improved data quality, Standards, Ad-hoc data queries ... and some disadvantages
 - Database development lifecycle
 - Planning, Definition, Requirements Analysis, Design, Implementation, Data loading, Testing, Maintenance
 - Design = conceptual, logical, physical stages

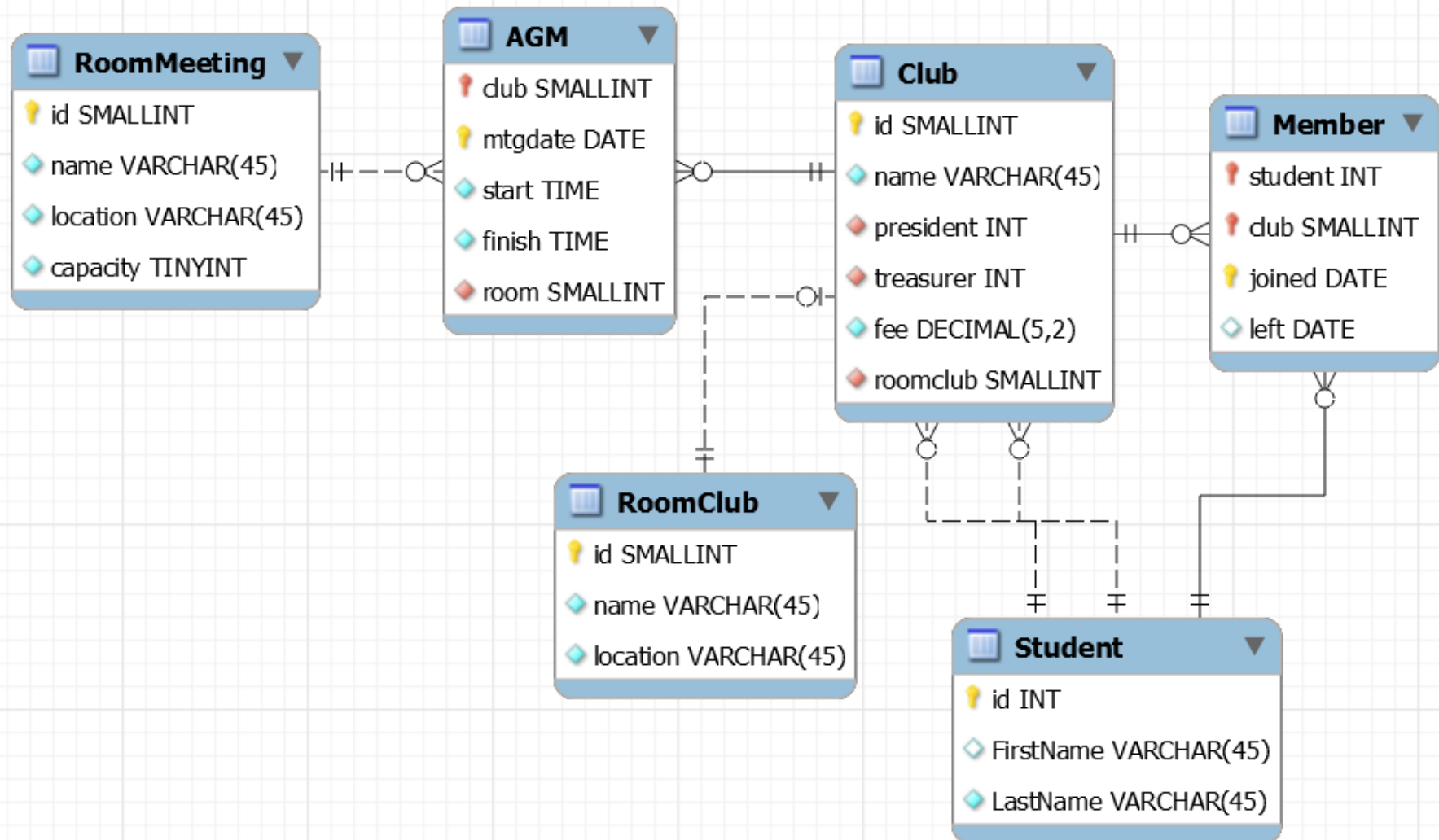
- Describe the phases of the database development lifecycle
- Differentiate the three steps in the design phase
- What are the advantages of relational databases?

- Weeks 3 – 5
 - Database Design
 - Conceptual, logical and physical
 - ER diagrams
 - How to move from conceptual to logical to physical design
 - SQL programming
 - DML statements
 - DDL statements
 - DCL statements

- Given “this description of a business case” produce a Conceptual / Logical / Physical data model
 - use homework examples for study: answers provided
 - and examples on next few slides: answers not provided
 - textbooks are full of examples
- Given “this data model”, write SQL queries to:
 - insert
 - update
 - delete
 - select data

- The Student Union wants a database to help in the management of student clubs. Each club has many members and any particular student can be a member of many clubs. Each club has a president and a treasurer. A student can be the president or treasurer in more than one club. Each club has a different membership fee.
- Each club has a clubroom where it can hold small meetings. Each clubroom is assigned to one club at most. The student union wants to keep track of which clubroom belongs to which club, but is not interested in keeping any information about the meetings held in the clubrooms.
- The clubrooms are too small for clubs to use for their annual general meetings (AGM), and so the Student Union also provides a small number of shared meeting rooms specifically for this purpose. Each club informs the Student Union of the date and time it wishes to hold its AGM, and the Union then chooses a meeting room that is free at that date and time that is large enough to hold that club's AGM.
- Initially, information about club's finances, assets, etc. will not be maintained in this database.

Suggested model for student clubs





- Draw an ER model of the following situation (state any assumptions you believe you have to make in order to develop a complete diagram). The Department of Computing and Information Systems is offering degrees in 10 locations within Australia. Each year about 10,000 students graduate. The Department keeps track of each graduates' student number, birth name, country of birth, current country of citizenship, current name, current address, and the name of each subject the student successfully completed.
- In order to maintain strong ties to its graduates the Department holds many functions (a dinner, reception, conference, or seminar) around Australia. Events have a title, date, location and type. The Department keeps track of which students have attended each event. For each attendance of a graduate at an event a comment is recorded if new information about the graduate is discovered.
- The Department also keeps track of students via phone, mail and email. In addition a member of staff in the department is assigned as a mentor for each graduand. Additional information stored includes the location where the graduand attended the IS course. Because of the nature of the course a graduand may have attended the course at several locations.

- A shipping magnate owns many container ships.
- Containers are collected at one port and delivered to another port.
- Customers pay a negotiated fee for the delivery of each container.
- Each ship has a sailing schedule that lists the ports the ship will visit over the next six months.
- The schedule shows expected arrival and departure dates at each port.
- The daily charge for use of each port is also included.

- Sally operates a cinema chain. She has given you the following requirements ...
- I have many customers. Each cinema can have multiple theatres. Movies are shown throughout the day starting at 11AM and finishing at 1AM. Each movie is given a two hour time slot. We never show a movie in more than one theatre at a time, but we do shift movies between theatres as the required number of seats changes.
- I am interested in knowing how many people, classified into adults and children, attend each showing of a movie. I vary ticket prices by movie and time slot. For instance, Mad Max 4 at 10AM costs \$10 for everyone, but at 11PM it is \$15.

- Week 5
 - Normalisation
 - functional dependency, determinants
 - 1st, 2nd, 3rd, BC normal forms
 - what happens if the database isn't normalised
 - an example of normalising a set of relations
 - Physical Implementation
 - inputs to and decisions made in the physical design stage
 - choosing correct data types - why this is important
 - de-normalisation
 - when to create indexes

- What is normalisation and why is it used?
- What problems arise if normalisation is not carried out?
- Describe 1st, 2nd and 3rd normal form
- Given a relation, carry out normalisation to 3rd normal form.
- Are the following relations in 3rd (or 2nd or 1st) normal form?

- What are the inputs to the (or what are the decisions made in the) Physical design process?
- What is de-normalisation, how is it used, give an example?
- How does indexing work, when should we use indexing?
- What data types should we use for the following columns?



- Week 7 – Database + Application Software
 - Reasons we put applications between users and databases
 - Presentation / UI
 - Business logic
 - Embedding SQL in application software
 - 1, 2, 3-tiered architectures
 - Web applications

- Why should the presentation, business logic and data be separated in internet-based database applications?
- What are the advantages of 2/3/4 tier application architectures?
- How has the web impacted database design and administration?



- Week 8 - Transactions and Concurrency
 - Definition of a transaction
 - Transaction properties
 - ACID & Serializability
 - Transaction logging
 - Concurrency
 - problems:
 - lost updates, uncommitted data, inconsistent retrieval
 - how to handle these – The Scheduler
 - locking and its various types and granularity

- What does “atomic” mean?
- Explain the three concurrency problems
- What is the purpose of locking,
and what types of locks can be used?
- Locking can occur at different levels - explain



- Week 8 - Distributed databases
 - Defined the term distributed database
 - Advantages and disadvantages
 - Looks at the objectives and trade-offs in distributed database
 - ease of access / location transparency, local autonomy
 - asynchronous / synchronous updates
 - Data replication
 - types
 - advantages and disadvantages
 - description and comparison of options

- What is a distributed database?
- Discuss the advantages and disadvantages of different distributed database options
- What is partitioning, how does each type work?
- What things does a DDBMS provide in addition to DBMS functionality?

- Week 9 - Database Administration
 - The DBA and Data Administrator roles
 - Architecture of a RDBMS
 - components and their interactions
 - Capacity Planning
 - estimate future capacity requirements of a database
 - Performance
 - what affects database performance
 - common approaches to monitoring and tuning performance
 - Security
 - threats
 - access control
 - web app security, SQL injection
 - Backup and Recovery
 - types of failure, types of backups
 - other methods to reduce likelihood of data loss

- Describe and differentiate the DBA and DA roles
- Calculate the capacity requirements for this database
- Describe common mechanisms used to improve the performance of database servers
- Security measures for databases include access control and encryption – describe and explain these
- What are the different types of backup?

- Guest lectures in weeks 10 are not directly examined



- Week 11 and 12 – NoSQL and Industry Trends
 - The relational status-quo
 - major vendors
 - current offerings
- Challenges to the relational status quo
 - big data
 - cloud storage
 - object-oriented applications
- Responses
 - NoSQL
 - NewSQL
 - Hadoop
 - In-memory databases



- Who are the leading commercial and open-source vendors?
- What are the pros and cons of storing data in the cloud?
- Describe the four main types of NoSQL database
- Explain the CAP theorem
- What is "eventual consistency"?

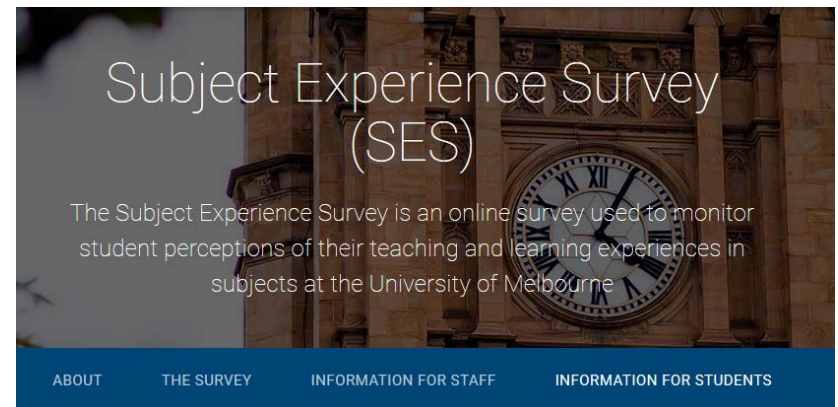


Student Experience Survey

- The SES is available now for you to fill in
- <https://ses.unimelb.edu.au/#students>
- Your feedback helps us to continuously improve our subjects
- You may win a \$100 Coles gift card !!!

We are particularly interested in your views on:

- assignments – more/less?
- guest lectures – more/less?
- modelling – more/less?
- SQL – more/less?
- NoSQL – more/less?
- labs – more or less?
- what would you add or remove from the course?





Thank you!

- Thank you for taking INFO90002
- We hope you find it useful for your career
- Good luck in the exam!