MONASH University

MONASH
INFORMATION
TECHNOLOGY

Week 1 - Introduction
FIT2094 - FIT3171 Databases
Clayton Campus S1 2019.

GROUP OF EIGHT AUSTRALIA

Your Databases Teaching Team - Clayton

Chief Examiner:
Dr Marc Cheong

Lecturers:
Dr Marc Cheong (Thurs) /
Mr Brendon Taylor (Fri)

Head TAs:
Mr Peter Huynh (Admin)
Mr Manoj Kathpalia (Oracle DB/Academic)

(*De facto adviser*: Mr Lindsay Smith)

Your Databases Teaching Team - Clayton

Tutors (alphabetical order)

1. Ahmed Shifaz
2. Andre Pinto
3. Arif Hidayat
4. Harsha Jarugu
5. Harsha Perera
6. Jiang Ou
7. Joe (Zhou) Shao
8. Jonny Low
9. Joseph (Daniel) Jitnah
10. Rodion Sharlov
11. Thejani Dineshika Wedikkara Arachchi
12. Vidya Saikrishna

[CLAYTON] Communication Channels

- As this unit is very large in enrolments…
  - –...plus in terms of staffing = 2 Clayton lecturers + 2 Clayton head tutors…
- We have a unified email role account: **FIT2094-FIT3171.AllCampuses-x@monash.edu**
  - –Monash Gmail will autocomplete it once you type FIT2094…
- **Please don't send individual staff your emails**
  - –...as these may go to the WRONG person, and confusion will arise!
  - –If you insist on doing so :-(... staff reserve the right to reply ONLY the following courtesy message: **"Please redirect your email to the role account"**
  - **–If there are any major issues, contact mgmt via role account - we do not tolerate any abuse of our staff.**

Overview (Hour 1)

- **Communication channels - important!**
- How are the two units run?
  - –incl. weekly activities, overview of assessments
- Unit Guide
- Moodle
- **Penalties for Plagiarism, Collusion - important!**
- Teaching Method (Peer Instruction in Lecture)

… then COFFEE BREAK!

How are the two units run?

- FIT2094 and FIT3171 have the SAME lecture!
  - –Either Thursday on campus - Marc's lecturing it
  - –Or Friday on campus - Brendon's lecturing it
  - –Or participate in either LiveStream (a staff member will be the chat moderator/host).
- FIT2094 and FIT3171 have the SAME tute!
- FIT2094 and FIT3171 have the SAME assignments

- **HOWEVER: the units have DIFFERENT exams, each exam tests different LEARNING OUTCOMES based on difficulty level.**

A typical week...

- BEFORE class:
  - –Pre-reading: textbook - either buy or borrow from library and share in your class/study group (next slide).
  - –Pre-reading: tute sheet and some pre-lecture notes (GDocs, linked via Moodle)
  - –Week 1 onwards: Quiz open from Wednesday 12:00, closes Tuesday 23:59, auto-marked by Moodle. (Australian timezone)
- Tutes
  - –(tutes and labs are the same thing, terms used interchangeably!)
  - –2hrs face-to-face session, guided by expert TAs
  - –complete tasks/discussion/coding, do hands-on work and discussion as per the Tute sheet.
- Lectures (sometimes after tutes in the week)

Textbook

- Options include:
  - –Borrow from library and share in your class/study group.
  - –Students can use older copy (12th Ed) to save $$$.
  - –Discounted textbook promo code →

DATABASE SYSTEMS DESIGN, IMPLEMENTATION & MANAGEMENT, CORONEL/MORRIS, 13th edition, ©2019- ISBN 9781337627900

FANCY A 10% DISCOUNT? Purchase directly from CengageBrain by following the steps below!

Step 1: Go to the Cengage Brain site: http://www.cengagebrain.com.au/shop/index.html new students simply **Create a Cengage Brain Account** or if you have an existing Cengage Brain Account, just enter your Username and Password.

Step 2: Click on this link for the prescribed text for your unit: http://www.cengagebrain.com.au/shop/search/9781337627900

Step 3: Select '**eBook**' and click 'add to cart'.

Step 4: Enter the Coupon code **STUDYSMART** to receive your 10% off!

Now you have access to your Course Material!

HAVE QUESTIONS?

Contact customer support Monday to Friday: 8am – 5pm AEST via PH: 1300 790 853 or Email: anz.techsupport@cengage.com or for FAQ's: https://www.cengagebrain.com.au/shop/faq-page

Overseas students: Students based overseas, can purchase the eBook via link: http://learn.cengage.com/overseas

Best of luck with your studies!

Outline and assessments

Assessment summary

Examination (2 hours): 50%; In-semester assessment: 50%

Assessment task	Value	Due date
Pre-Lecture Quizzes	5%	Weekly prior to the lecture (Weeks 2 to 11)
Assignment 1 Part A - Initial Database Design	5%	Week 6, Monday 9 am
Assignment 1 Part B - Database Design	20%	Week 8, Monday 9 am
Assignment 2 - Database Implementation and Maintenance	20%	Week 12, Friday 11.55 pm
Paper 1	50%	To be advised

Study Program

Week	Activities	Assessment
0		No formal assessment or activities are undertaken in week 0
1	**PART I: The Relational Model** Introduction to Database	
2	Relational Model	Pre-lecture Quiz Questions due weekly prior to the lecture (Weeks 2 to 11)
3	**PART II: Database Design** Conceptual model - E/R Diagram and UML	
4	Logical model - E/R Transformation	
5	Normalisation	
6	Database Implementation (DDL)	Assignment 1 Assignment 1 Part A - Initial Database Design
7	**PART III: The SQL Database Language** SQL I	
8	Update, Delete and Transaction Management	Assignment 1 Part B - Database Design
9	SQL II & Triggers	
10	SQL III	
11	**PART IV: Web Database Implementation** Database Connectivity and Web Technologies - Querying Data	
12	Web Technologies - Manipulating data	Assignment 2 - Database Implementation and Maintenance
	SWOT VAC	No formal assessment is undertaken in SWOT VAC
	Examination period	LINK to Assessment Policy:http://policy.monash.edu.au/policy-bank/academic/education/assessment/assessment-in-coursework-policy.html

Unit Guide

- … is your best friend!
- **UNIT GUIDE CONTENTS DIFFER PER UNIT (FIT2094 vs FIT3171) esp. LEARNING OUTCOMES.**
- Let's check them NOW.

Moodle

- … is another of your best friends!
- FIT2094 and FIT3171 have the **same Moodle throughout** for BOTH units…
  - –(except for any EXAM REVISION parts).
  - –NB: Moodle has just been upgraded, so navigation can be confusing if you used it before.
- Moodle links to weekly ‘eBooks’ or pre-reading content…
  - –Tute Notes and Pre-Lecture (Preview) Slides
  - –On Google Drive as a Doc file (read-only)
  - –...or a low-bandwidth (‘Lite’) published version.

Penalties for Plagiarism and Collusion

- DO NOT copy from other students, the Internet, textbooks, sample code...
- DO NOT complete assignments/quizzes in a group (unauthorised), or have someone else do your work...
- **SEVERE PENALTIES - INCL. LOSS OF MARKS, SUSPENSION, AND EXCLUSION FROM UNI...**

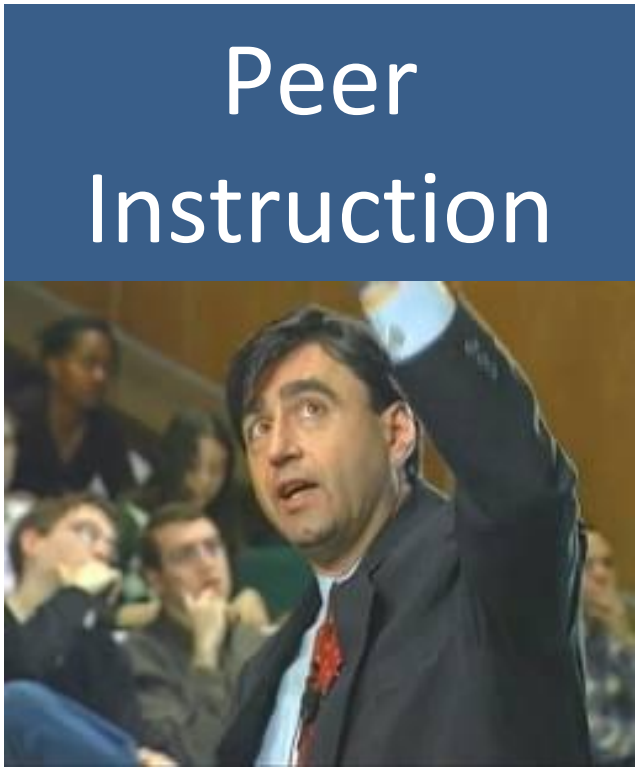
  –Always follow the policy: https://www.monash.edu/students/admin/policies/academic-integrity

  –The teaching team knows how to find out if your work is not your own… and the ASSOCIATE DEAN is aware of the methods used to cheat in OUR SPECIFIC unit - so BEWARE!!!

  –If you know 'where to find answers', so do we.

Teaching Method

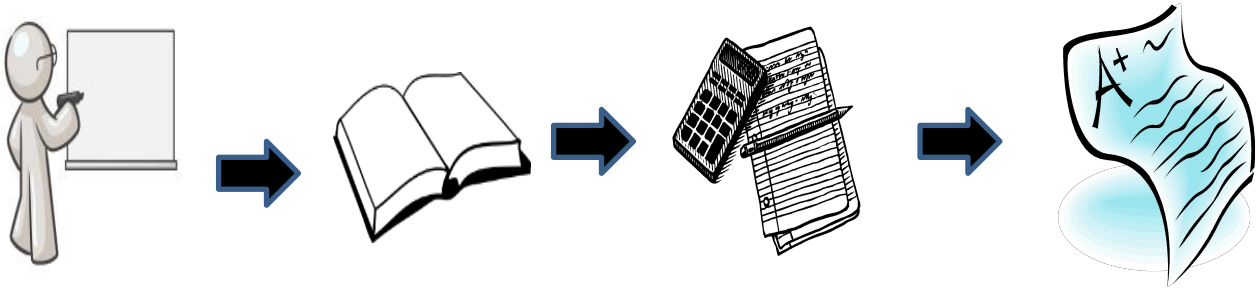
- Your peers help you to understand the concepts through discussion.
- Lecture includes a series of discussions on concepts.
- The lecturer guides the discussion.



Prof Eric Mazur, Harvard University

Traditional Teaching Method

Lecture | Textbook | Homework | Exam



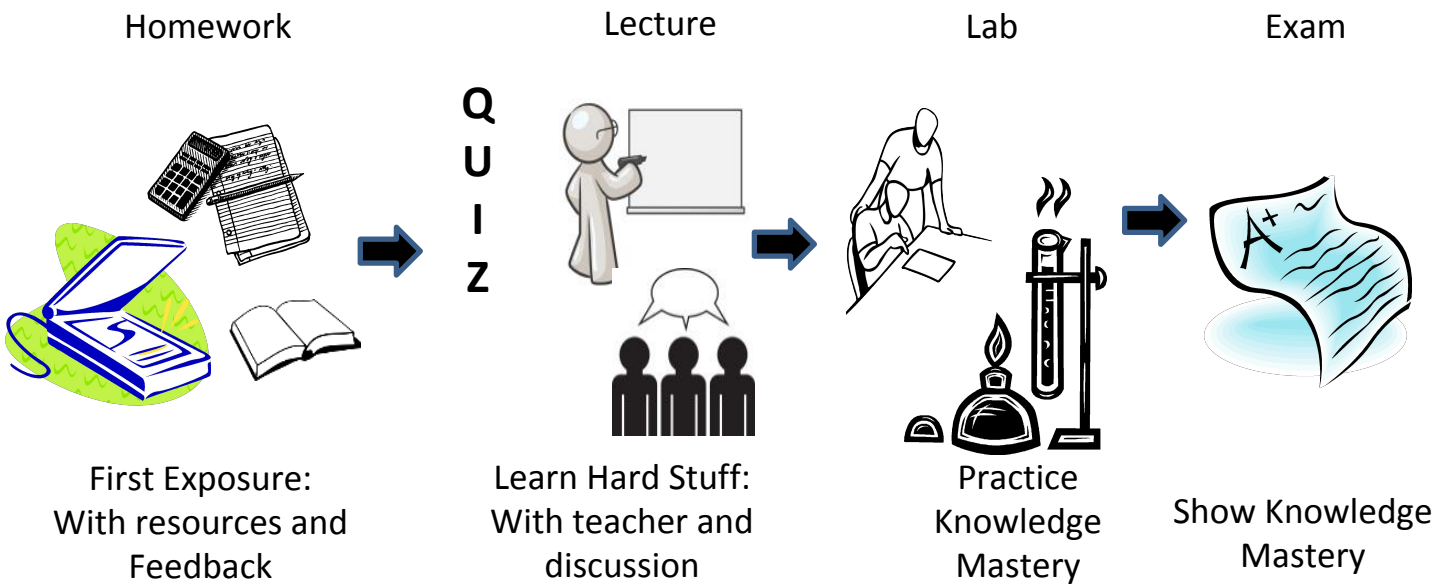
First Exposure

Read Hard Stuff

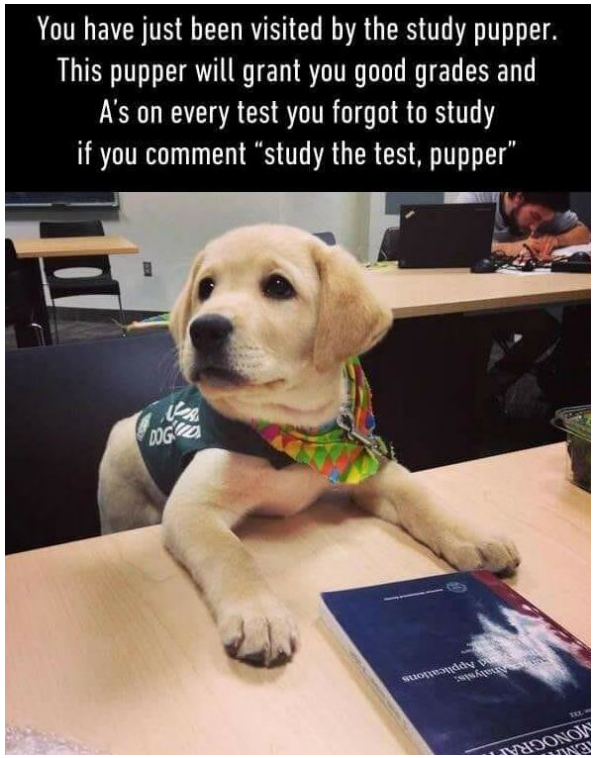
See if You Know Hard Stuff

Show Knowledge Mastery

Peer Instruction – Full Picture



Why Peer Instruction?



- Learn/practice hard concepts in class
- Build and test one's understanding in a supportive environment.
- Develop critical thinking, communication and reflection skills.
- Engage students to take ownership of their learning.

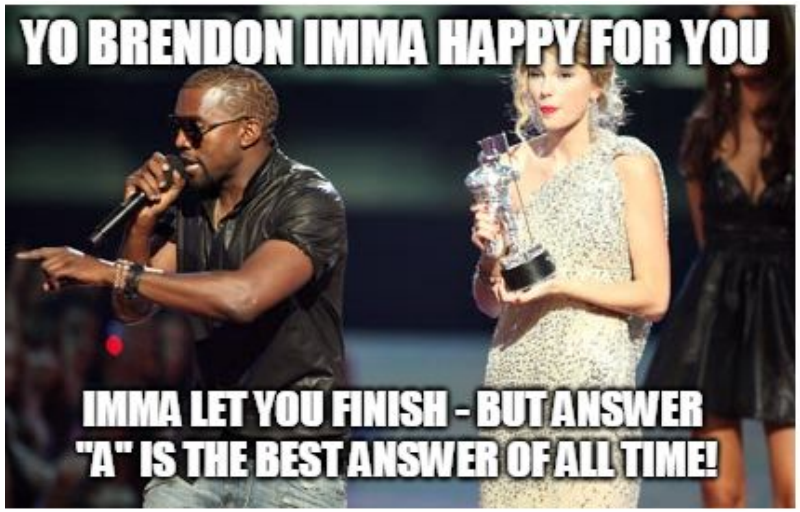
Img src: Facebook

Things are different…

- Pre-lecture activities are crucial.
  - –Your lecture experience will depend on your preparation.
- Attending lectures is very important
- Our lecture slides are NOT your notes!
  - –Create your own notes during pre-lecture reading.
  - –Annotate difficult concepts, revisit the annotation after lecture/tutorials.
  - –It is better not to take notes during lecture. You should be prepared before the lecture, then think, discuss and ask questions during lectures.
  - –After all, if you're not prepared, you can't attempt the quiz (in future)

Discussion Questions – Scenario

- Lecturer shows a question.
- Student answers using the response system.
- If uncertainty arises...
  - –Group discussion (2-3 students) – need to get a consensus.
  - –Class wide discussion.



Img src: imgflip

Why The Scenario?

- Pose carefully designed question
- First: Think for yourself and select answer
  - –Checks your understanding and create an opinion to base your discussion during the group discussion, if needed.
- If needed (e.g. tough question)
  - –Discuss: Analyze problem in teams of 2-3 → Practice analyzing, talking about challenging concepts → Reach consensus
  - –Group vote: Everyone in group votes → You must all vote the same → Convince your group or get convinced by your group.
  - –Class wide discussion.

Using FLUX

- Visit https://flux.qa presenter/dashboard on your Internet-enabled device
- Log in using your Authcate details
- Touch the + symbol
- Enter the code for your lecture.
- Answer questions when they pop up.

These slides with the blue background are Clayton FLUX slides!

[Q1] Clayton students: Which of the following ways are best for discussing a special consideration case in the unit?

a. Moodle forums
b. Email Brendon.Taylor at monash.edu
c. Email Peter.Huynh at monash.edu
d. Contact Donald Trump
e. Email FIT2094-FIT3171.AllCampuses-x@monash.edu
f. Special consideration automatically granted (none of the above)

These slides with the blue background are Clayton FLUX slides!

[Q2] Clayton students: Which student(s) will NOT be guilty for plagiarism and/or collusion?

a. Snoop Lion copies SQL code from StackOverflow for a weekly tute.
b. Kanye West finishes one question in the assignment together with Kim Kardashian.
c. Jack Nicholson answers the quiz for Leo DiCaprio in exchange for a beer and 50 dollars.
d. Dilbert accidentally pasted some SQL code for the assignment on the forums. Garfield pastes this code into his own assignment - and it worked!
e. All of the students above are GUILTY.

These slides with the blue background are Clayton FLUX slides!

[Q3] Clayton students: Do I have to attend my assigned tutorial class, or can I go to any class?

a. As long as you go to at least one tutorial class a week, it doesn't matter.
b. You must attend your assigned tutorial class.
c. Work commitments is a reasonable reason to go to an alternative tutorial class.
d. If I have a timetable clash, it's okay to attend another class.

These slides with the blue background are Clayton FLUX slides!

[Q4] Clayton students: Which database management systems are you most familiar with?

a. Oracle
b. MySQL
c. MS Access
d. SQL Server
e. Other DBMS
f. I am not familiar with any **D**ata**B**ase **M**anagement **S**ystems.

Is it bad to get it WRONG? NO!

Lindsay's anecdote:

"It is better to be WRONG and understand why you are WRONG, rather than, getting the RIGHT answer but NOT knowing WHY it is the RIGHT answer!"

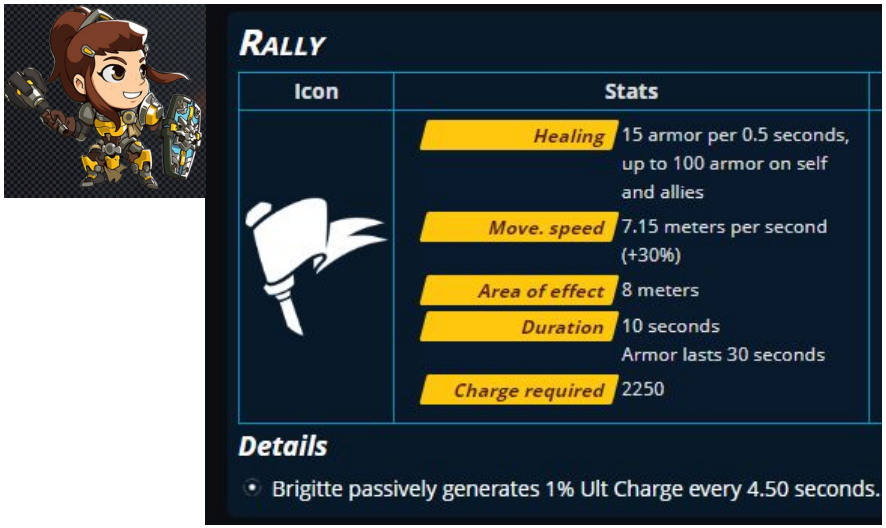
[Marc's anecdote]

Marc's anecdote: "Better get it wrong now and learn from it, rather to get it wrong in the exam and fail it".

Do your best, even if you get it wrong now…
...as long as you learn!

vs

Get it wrong in the exam…



Img src: https://overwatch.gamepedia.com/
Youtube / Tolkien / New Line Cinema

Coffee break - see you in 10 minutes.

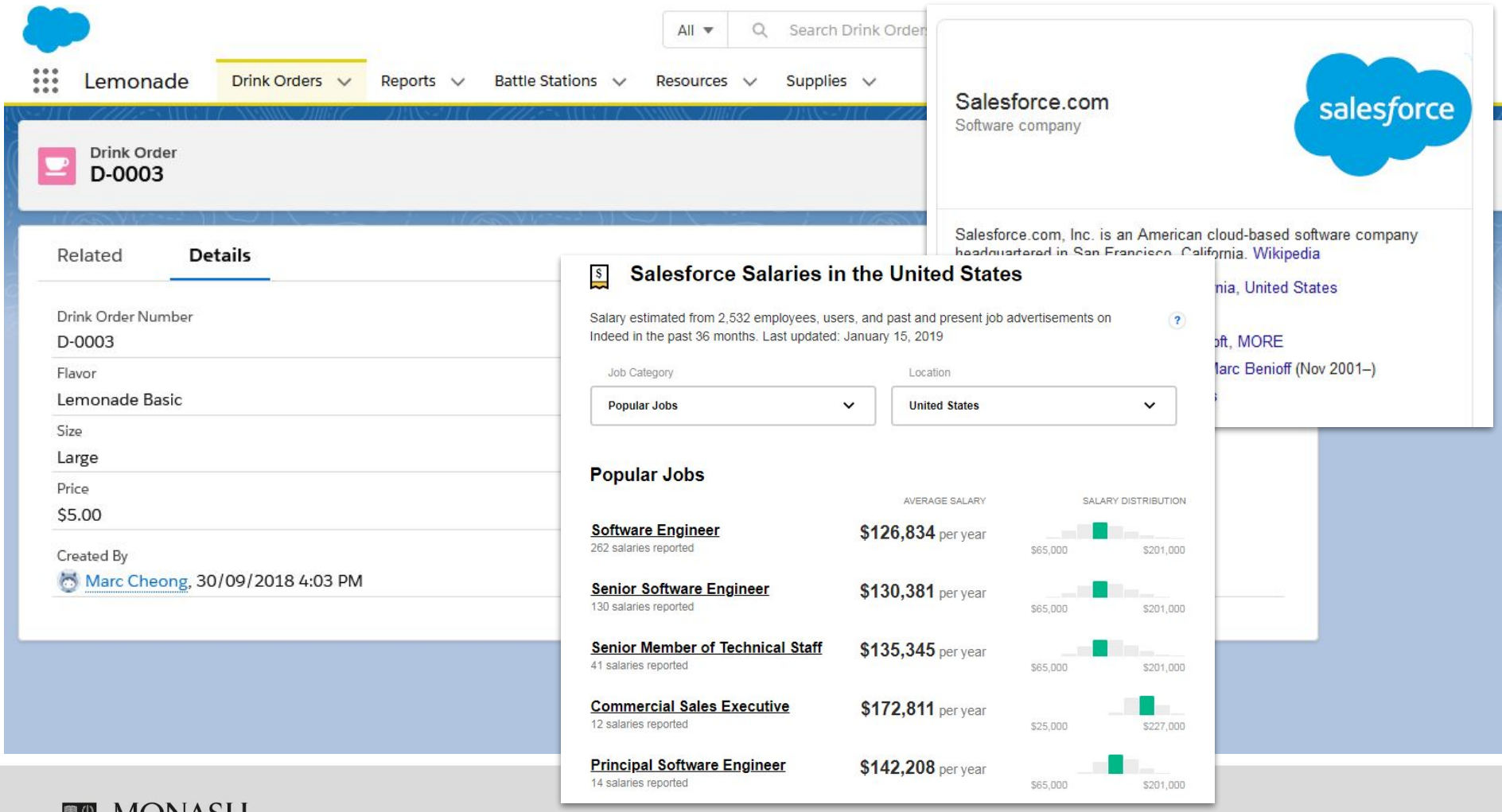
Img src: Jazmin Quaynor @jazminantoinette at Unsplash

[Clayton] Slide relevant to Clayton students only...

Overview (Hour 2)

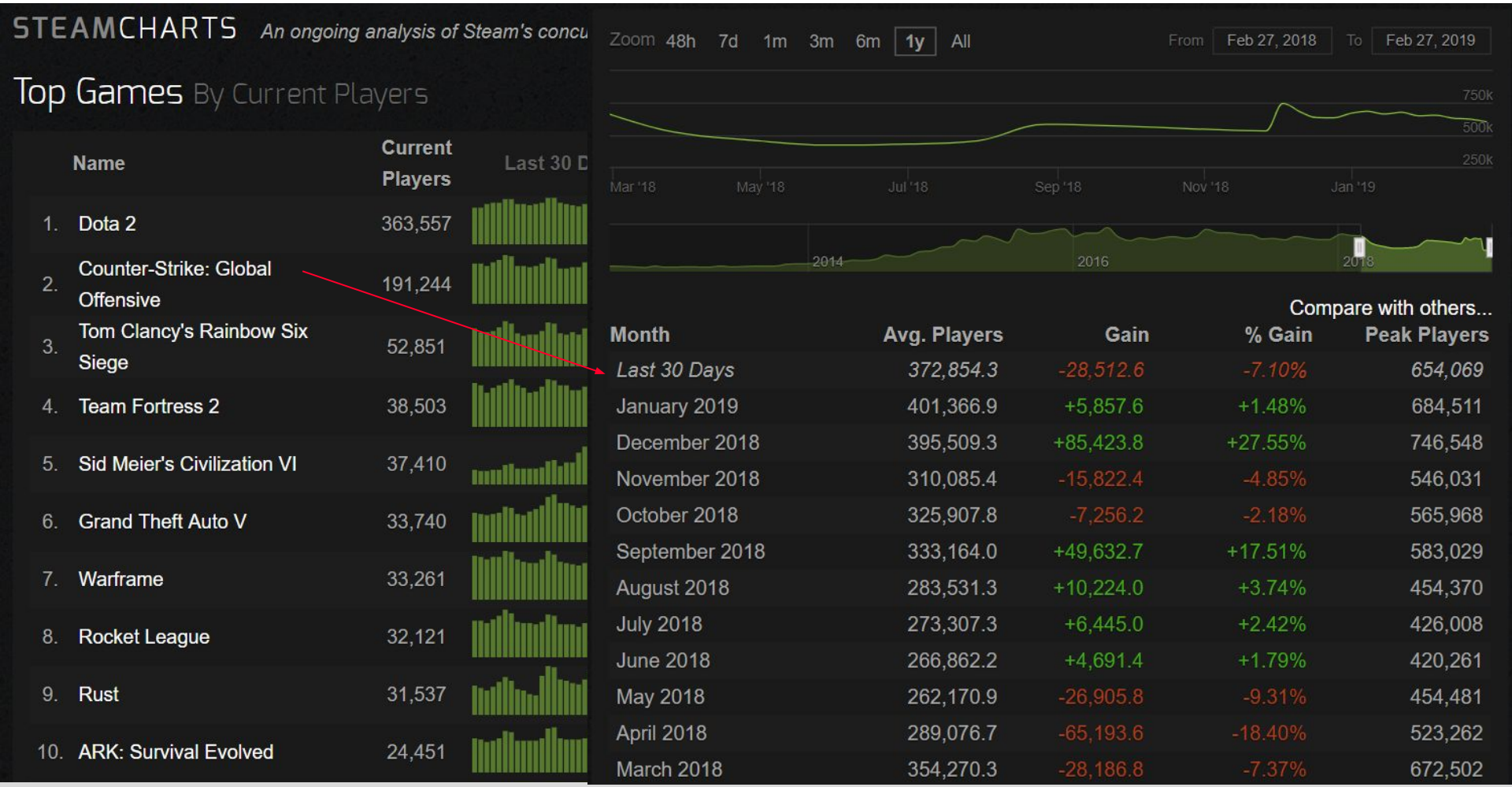
- An overview of relational database management systems (RDBMS)

[Marc's Anecdote]



Img src: Salesforce / Google Info Box / Indeed.com

[Brendon's Anecdote]

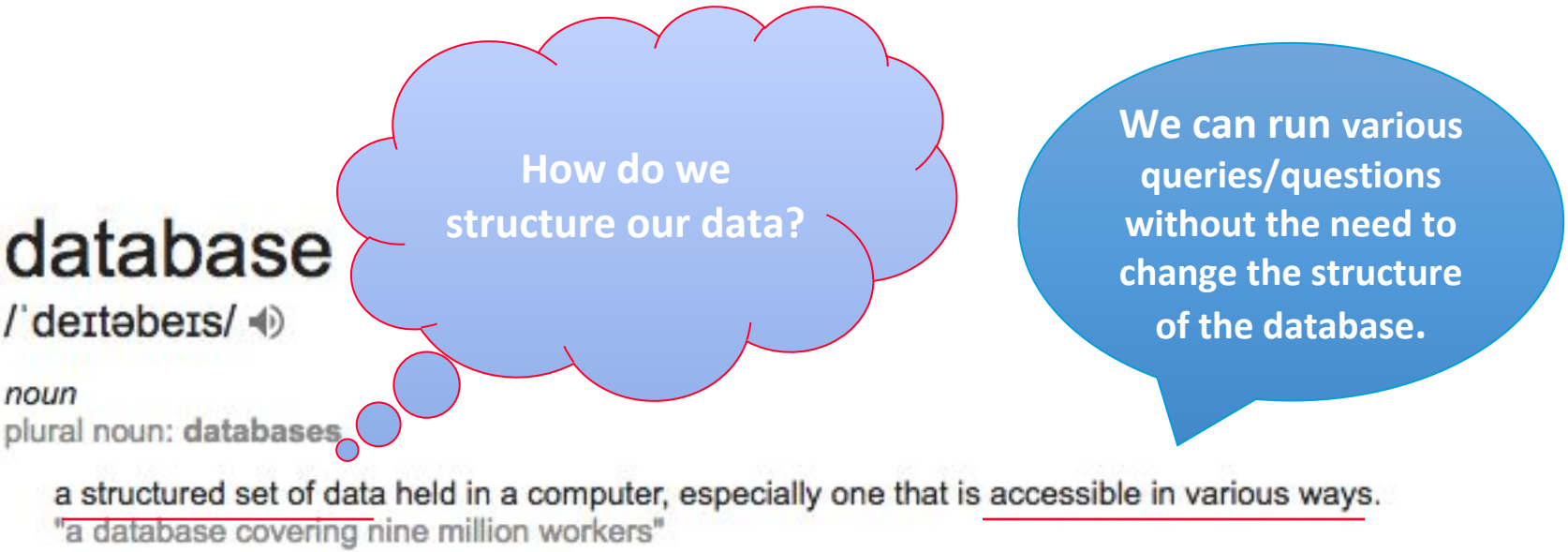


Img src: steamcharts.com

Let's travel back to 1960s

- Relational databases do not exist yet
- Let's create a database to record the information on Monash students
  - What kind of approaches do we have?
  - What kinds of problems are involved?

What is a database?



How do we structure our data?

Use a document format like MS-Word document?

Use a spreadsheet like Excel?

- How easy is it to answer a number of queries?
- What kind of guarantee do we have from the systems on data integrity after a modification
  - (eg deletion, update or insertion of one or more records to the system?

Data Redundancy – a student data spreadsheet

STU_NBR	STU_LNAME	STU_FNAME	STU_DOB	UNIT_CODE	UNIT_NAME	ENROL_YEAF	ENROL_SEM	MARK	GRADE
11111111	Bloggs	Fred	1-Jan-90	FIT1002	Computer Pr	2013	1	66	C
11111111	Bloggs	Fred	1-Jan-90	FIT1004	Database	2013	1	80	HD
11111112	Nice	Nick	10-Oct-94	FIT1001	Computer Sy	2013	1	80	HD
11111112	Nice	Nick	10-Oct-94	FIT1001	Computer Sy	2012	1	35	N
11111114	Sheen	Cindy	25-Dec-96	FIT1001	Computer Sy	2012	1	78	D
11111114	Sheen	Cindy	25-Dec-96	FIT1004	Database	2013	1	60	C
11111113	Wheat	Wendy	5-May-90	FIT1001	Computer Sy	2012	2	65	C
11111113	Wheat	Wendy	5-May-90	FIT1004	Database	2013	1	78	D

What would happen if we delete Fred's enrolment in FIT1002? What happen to the details of FIT1002 information such as its name?

How would you update the mark for Cindy's enrolment in FIT1001? (Imagine the spreadsheet contains thousands of students and each student has 12 enrolment entries).

How would you introduce a new unit, eg FIT2133 Programming in Python into the spreadsheet when no student is enrolled to the unit yet?

Why do we have so many problems in the previous example?

- The structure of the data causes some data management problems or data anomalies.
- The software was not designed to deal with the type of reporting required.

How do we solve it?

STU_NBR	STU_LNAME	STU_FNAME	STU_DOB
11111111	Bloggs	Fred	01/JAN/90
11111112	Nice	Nick	10/OCT/94
11111113	Wheat	Wendy	05/MAY/90
11111114	Sheen	Cindy	25/DEC/96

UNIT_CODE	UNIT_NAME
FIT1002	Computer Programming
FIT1001	Computer Systems
FIT1004	Database

- Keep details of student, unit and enrolment separately, BUT keep the relationships among them in the system.

Relational Model
Relational Database
Relational Database
Management systems

STU_NBR	UNIT_CODE	ENROL_YEAR	ENROL_SEMESTER	MARK	GRADE
11111114	FIT1001	2012	1	78	D
11111111	FIT1002	2013	1	60	C
11111111	FIT1004	2013	1	80	HD
11111112	FIT1001	2012	1	35	N
11111112	FIT1001	2013	1	80	HD
11111113	FIT1001	2012	2	65	C
11111113	FIT1004	2013	1	78	D
11111114	FIT1004	2013	1	60	C

DATABASE

STU_NBR	STU_LNAME	STU_FNAME	STU_DOB
11111111	Bloggs	Fred	01/JAN/90
11111112	Nice	Nick	10/OCT/94
11111113	Wheat	Wendy	05/MAY/90
11111114	Sheen	Cindy	25/DEC/96

Entities/Tables

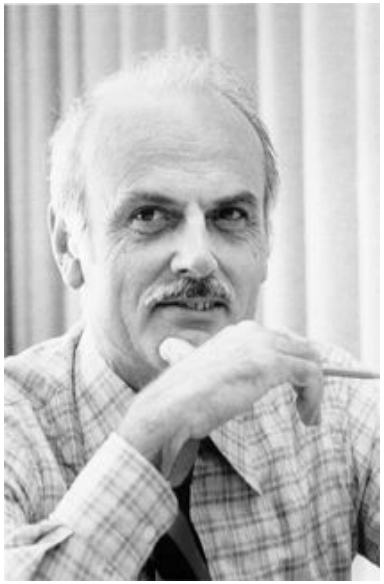
UNIT_CODE	UNIT_NAME
FIT1002	Computer Programming
FIT1001	Computer Systems
FIT1004	Database

STU_NBR	UNIT_CODE	ENROL_YEAR	ENROL_SEMESTER	MARK	GRADE
11111114	FIT1001	2012	1	78	D
11111111	FIT1002	2013	1	60	C
11111111	FIT1004	2013	1	80	HD
11111112	FIT1001	2012	1	35	N
11111112	FIT1001	2013	1	80	HD
11111113	FIT1001	2012	2	65	C
11111113	FIT1004	2013	1	78	D
11111114	FIT1004	2013	1	60	C

A collection of tables and their relationships is a DATABASE

1970: Relational model

- An IBM scientist
- Proposed and developed the relational model
- Also proposed normalisation forms
- Resistance from IBM to implement his model
- Turing award (1981)



E.F Codd
(1923-2003)

- Relational model in week 2
- Normalisation in week 5
- E. F. Codd, "A Relational Model of Data for Large Shared Data Banks", *Comm. Of ACM*, 1970

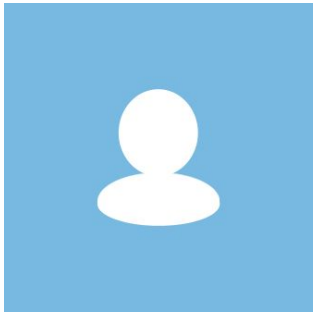
1974: SQL

- Developed at IBM
- Initially called SEQUEL (**S**tructured **E**nglish **QUE**ry **L**anguage)
- Doesn't strictly follow Codd's theory
- Oracle: the first commercially available implementation of SQL in 1979

- SQL in weeks 7, 8, 9 & 10

- D Chamberlin, R Boyce, "SEQUEL: A structured English query language", *ACM SIGFIDET*, 1974

Donald Chamberlin (1944-)



Raymond Boyce
(unknown - 1974)

1976: Conceptual model

- Proposed Entity-Relationship Model (ER diagram)
- A systematic process to design a relational database

- Database design process in week 3 & 4

- Peter Chen, "The entity-relationship model—toward a unified view of data", *ACM TODS*, 1976

Peter Chen (1947 -)

1979: Oracle

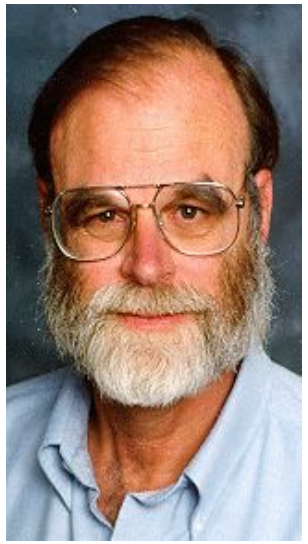
- Inspired by Codd's ideas
- First commercial release in 1979
- Most popular RDBMS
- Introduced PL/SQL in 1988 (Procedural Language/SQL)

Larry Ellison (1944 -)

- Oracle SQL in week 7, 8, 9 & 10

1981: Transactions management

- Introduced transaction management
- Turing award (1998)
- Presumed lost at sea in 2007



Jim Gray (1944 -)

- Transaction management in week 8

- Jim Gray, "The Transaction Concept: Virtues and Limitations ", *VLDB*, 1981

Data Management Today

- Relational databases are still very popular. But …
  - –Social Networks (Facebook, Twitter, Foursquare etc.)
  - –Multimedia data (YouTube, Pinterest, Facebook etc.)
  - –Data streams (Twitter, computer networks)
  - –Spatial data (Road networks, Google Earth, Space etc.)
  - –Textual data
  - –Web data
  - –Big Data
  - –…

https://goo.gl/zMxG3b

In Perspective ...

Early Computers

IBM System/360

IBM

IBM IMS DBMS

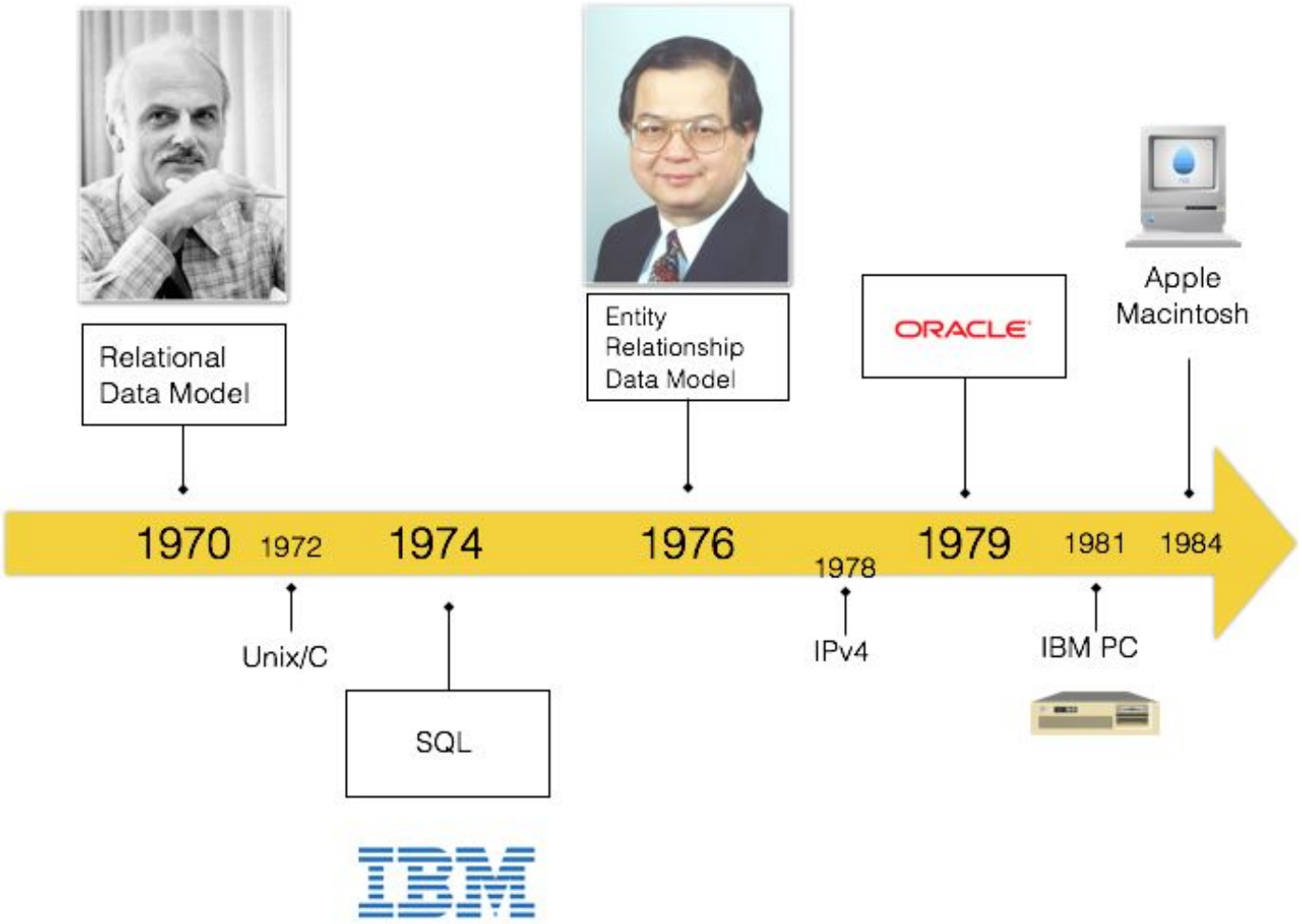
ARPANET

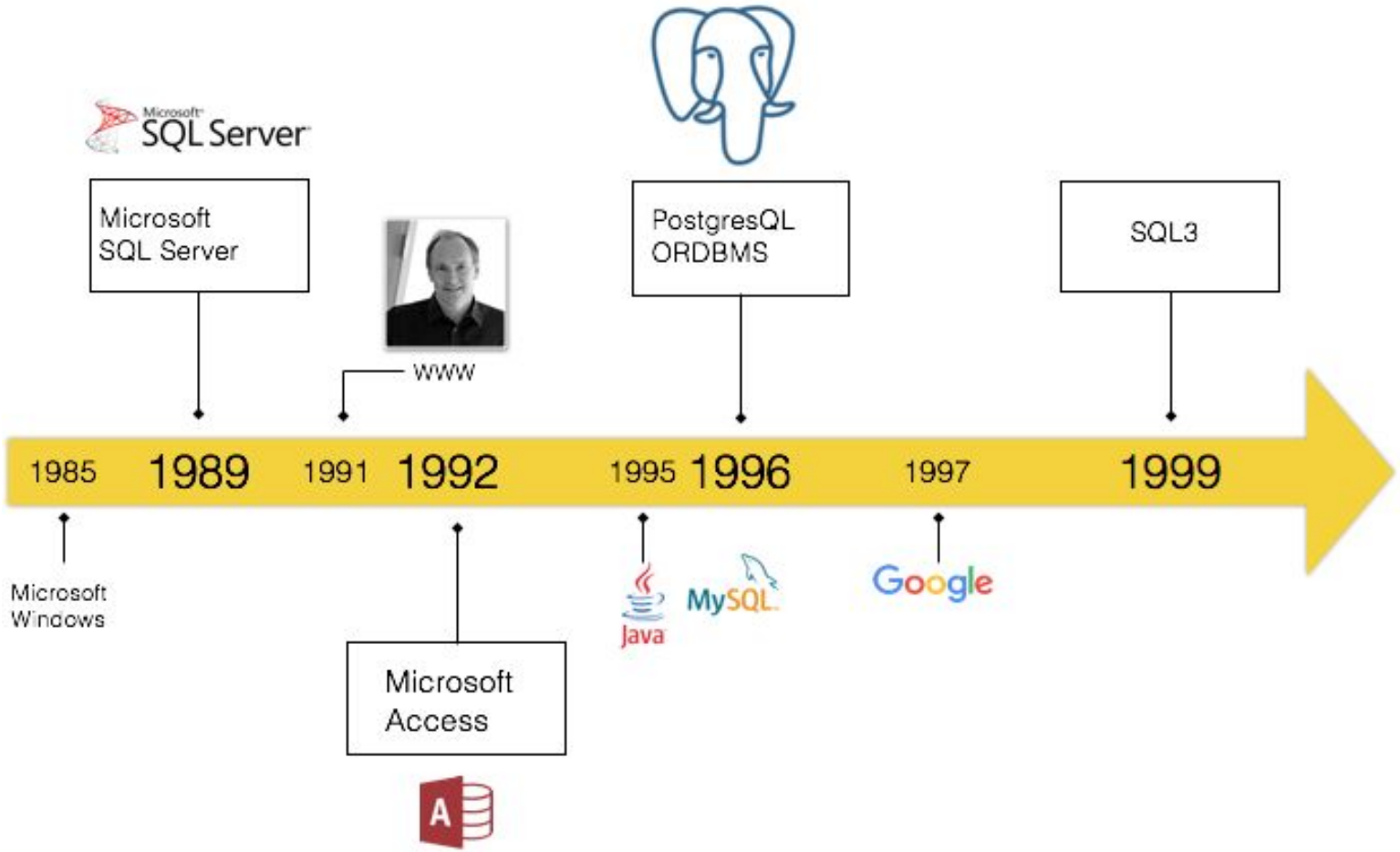
1950s 1961 1964 1966 1969

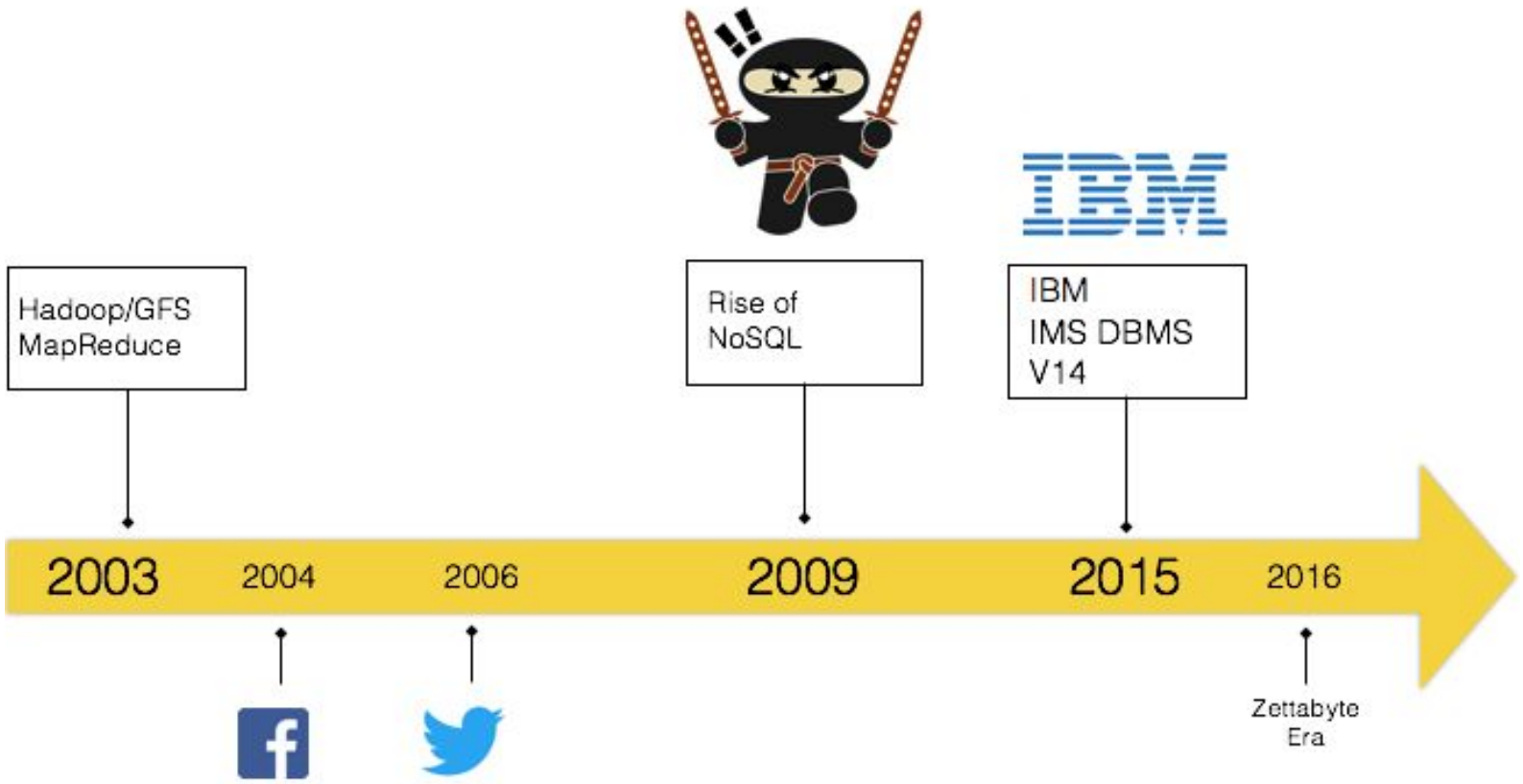
JFK Moon Speech

"*Database*"

Apollo 11


Relational
Data Model
Entity
Relationship
Data Model
ORACLE
Apple
Macintosh
1970
1972
1974
1976
1978
1979
1981
1984
Unix/C
SQL
IPv4
IBM PC
IBM


Microsoft SQL Server
Microsoft SQL Server
www
PostgresQL ORDBMS
SQL3
1985
1989
1991
1992
1995
1996
1997
1999
Microsoft Windows
Microsoft Access
Java
MySQL
Google


Hadoop/GFS
MapReduce
Rise of
NoSQL
IBM
IMS DBMS
V14
2003
2004
2006
2009
2015
2016
Zettabyte
Era

RANK	DBMS	TYPE	INTRODUCED
1	ORACLE	Commercial, Relational DBMS	1979
2	MySQL	Open source, Relational DBMS	1995
3	Microsoft SQL Server	Commercial, Relational DBMS	1989
4	PostgreSQL	Open source, Relational DBMS	1996
5	mongoDB	Open Source, Nosql - Document Store	2009
6	IBM DB2	Commercial, Relational DBMS	1983

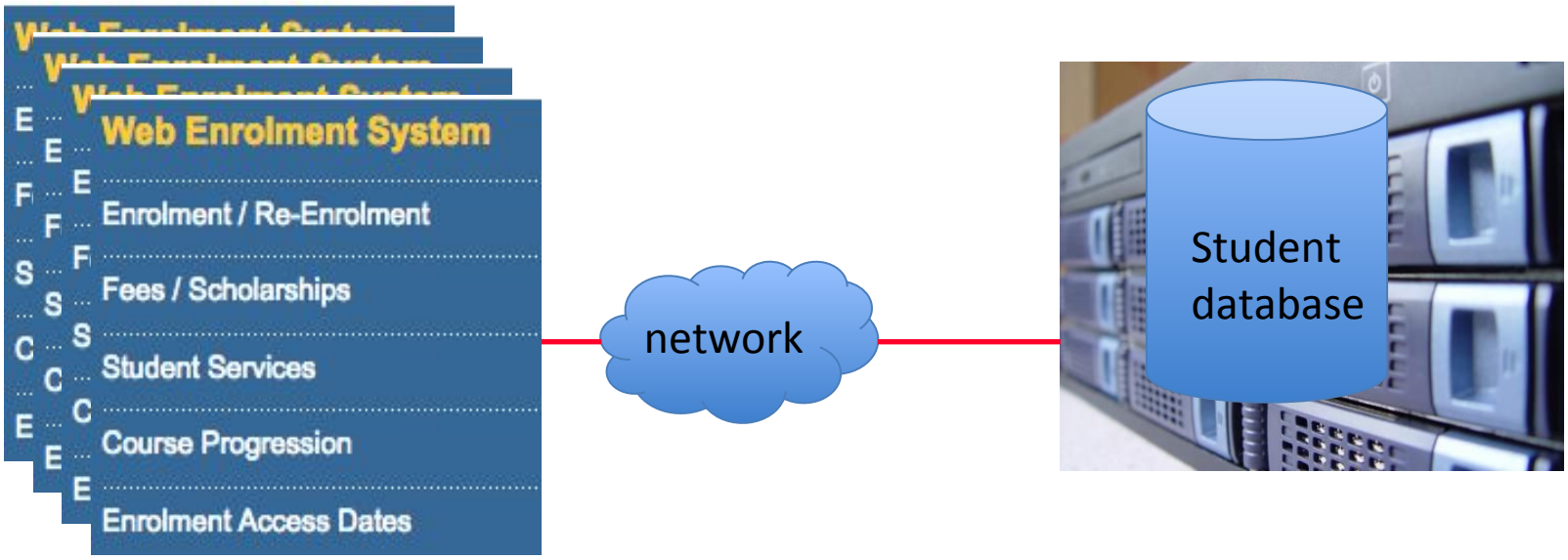
DB-ENGINES July 2018

343 systems in ranking, February 2019

Rank Feb 2019	Rank Jan 2019	Rank Feb 2018	DBMS	Database Model	Score Feb 2019	Score Jan 2019	Score Feb 2018
1.	1.	1.	Oracle	Relational, Multi-model	1264.02	-4.82	-39.26
2.	2.	2.	MySQL	Relational, Multi-model	1167.29	+13.02	-85.18
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1040.05	-0.21	-81.98
4.	4.	4.	PostgreSQL	Relational, Multi-model	473.56	+7.45	+85.18
5.	5.	5.	MongoDB	Document	395.09	+7.91	+58.67
6.	6.	6.	IBM Db2	Relational, Multi-model	179.42	-0.43	-10.55
7.	7.	↑ 8.	Redis	Key-value, Multi-model	149.45	+0.43	+22.43
8.	8.	↑ 9.	Elasticsearch	Search engine, Multi-model	145.25	+1.81	+19.93
9.	9.	↓ 7.	Microsoft Access	Relational	144.02	+2.41	+13.95
10.	10.	↑ 11.	SQLite	Relational	126.17	-0.63	+8.89
11.	11.	↓ 10.	Cassandra	Wide column	123.37	+0.39	+0.59
12.	↑ 13.	↑ 17.	MariaDB	Relational, Multi-model	83.42	+4.60	+21.77
13.	↓ 12.	13.	Splunk	Search engine	82.81	+1.39	+15.55
14.	14.	↓ 12.	Teradata	Relational	75.97	-0.22	+2.98
15.	15.	↑ 18.	Hive	Relational	72.29	+2.38	+17.23
16.	16.	↓ 14.	Solr	Search engine	60.96	-0.52	-2.91
17.	17.	↓ 16.	HBase	Wide column	60.28	-0.12	-1.43
18.	18.	↑ 19.	FileMaker	Relational	57.79	+0.64	+3.43
19.	19.	↑ 20.	SAP HANA	Relational, Multi-model	56.55	-0.09	+9.19
20.	↑ 21.	↓ 15.	SAP Adaptive Server	Relational	55.75	+0.71	-7.74

Img src: https://db-engines.com/en/ranking, retrieved Feb 2019

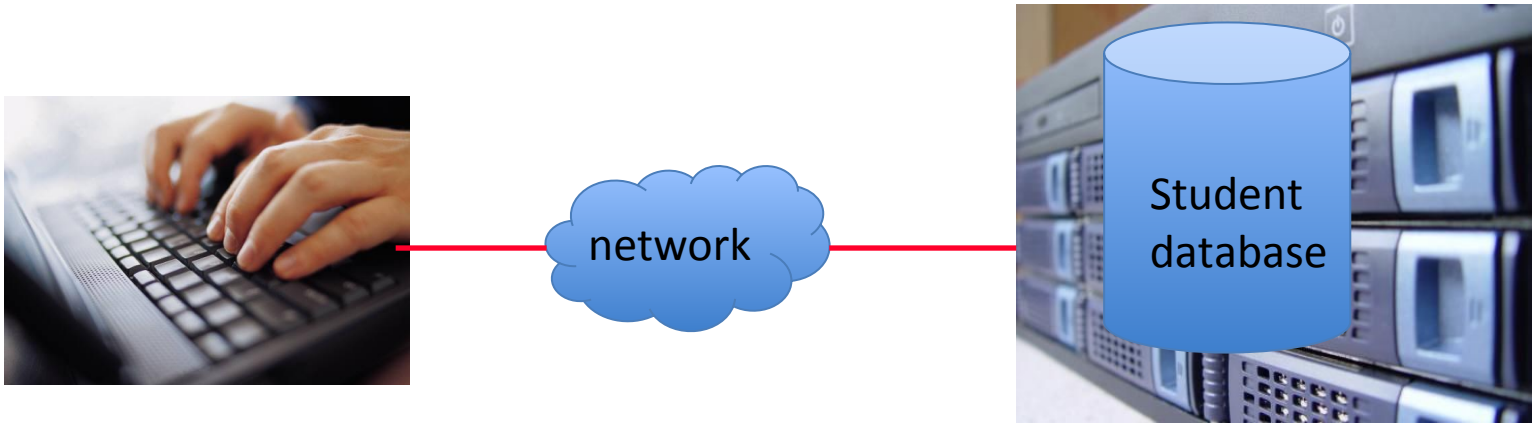
Relational database systems in action: End-users' view



Front end application (client)

Student Database is implemented in an Oracle DBMS (server)

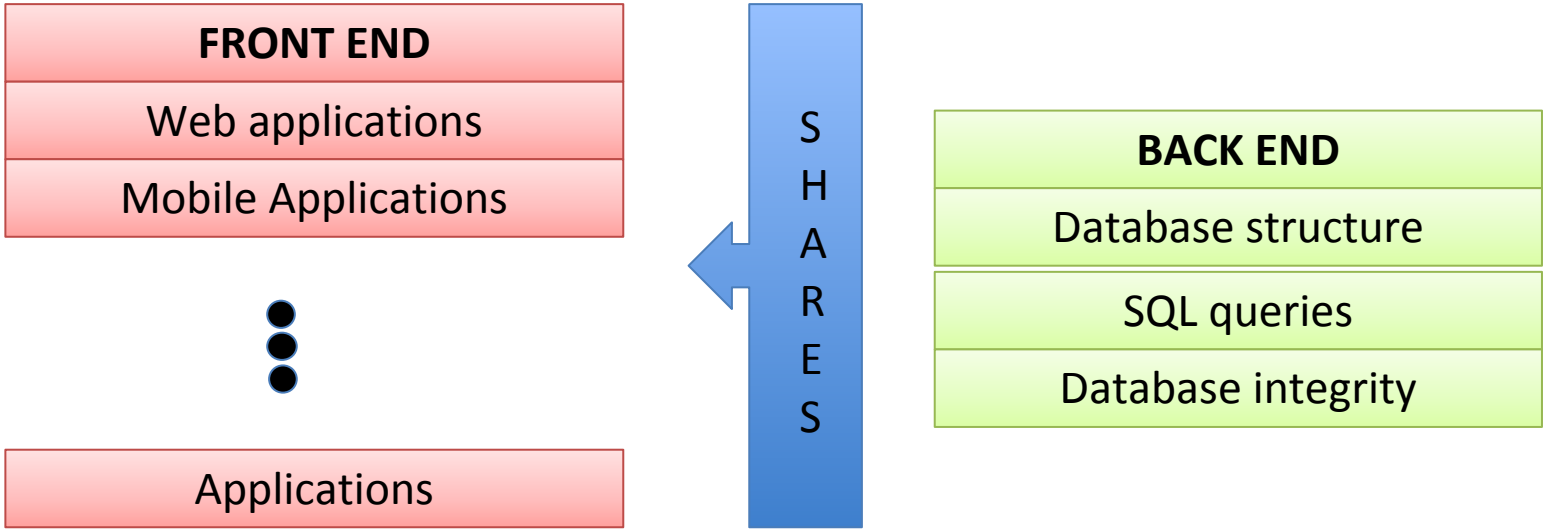
Database Systems in Action
Developers' View



Development environment (client, eg SQL Developer, Integrated Development Environment for web scripting)

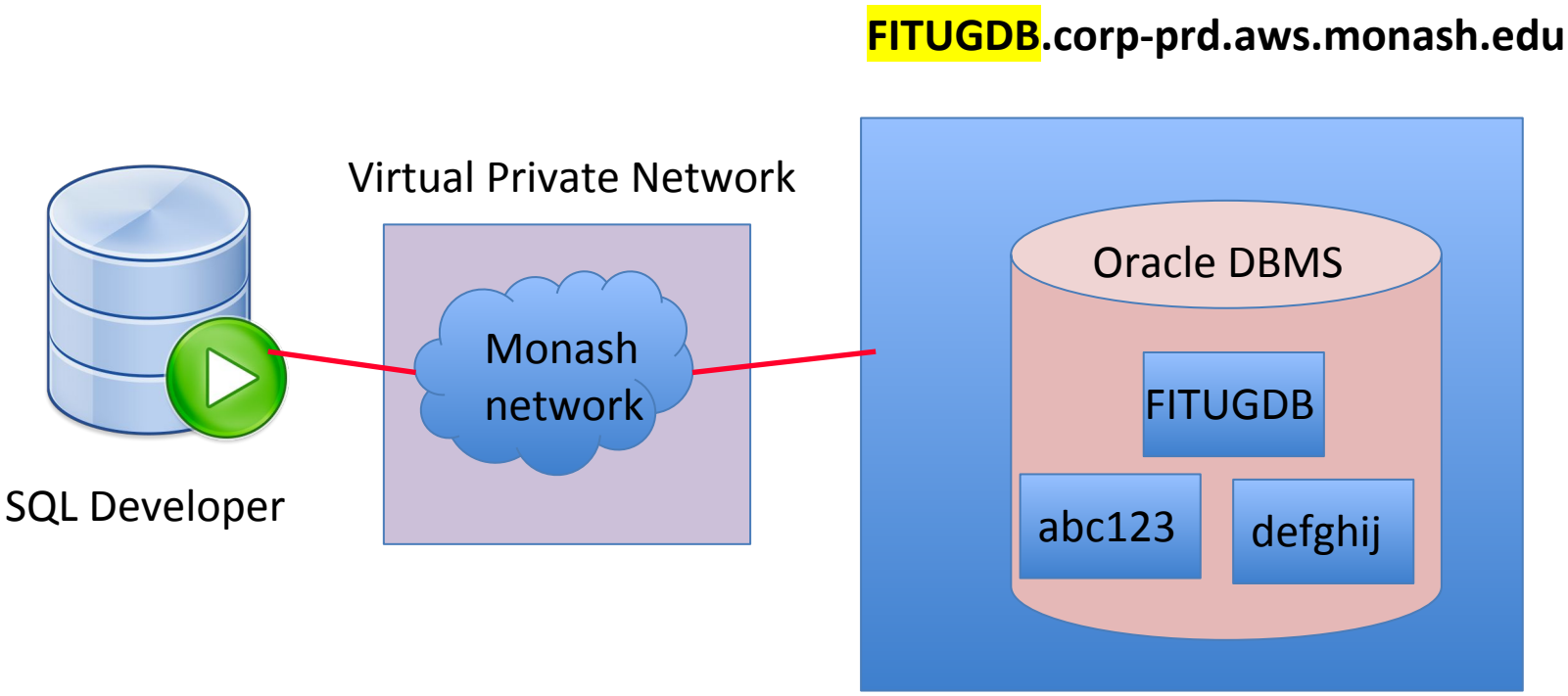
Student Database (server)

Developing Application with Database

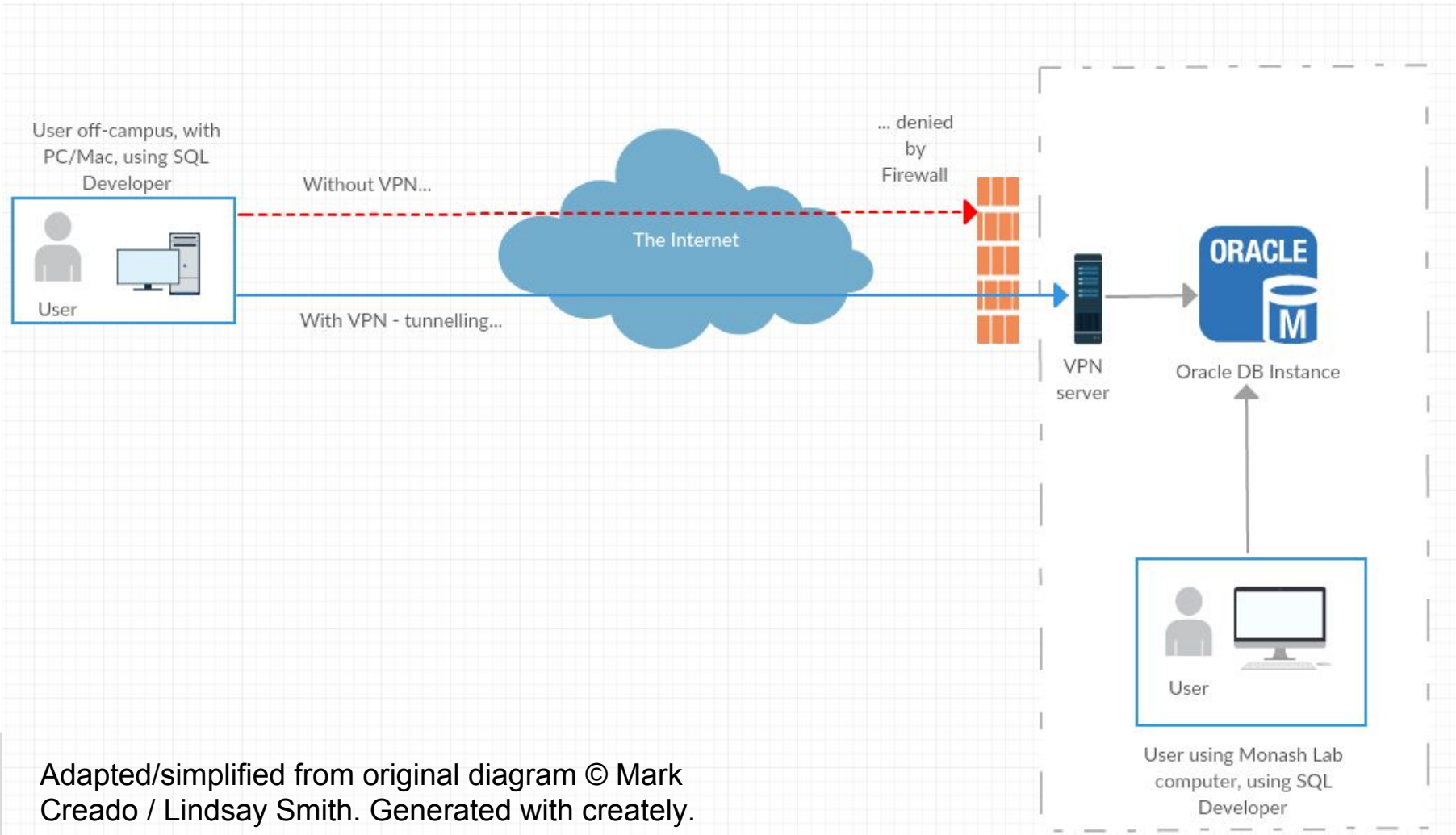


In this unit, we will concentrate on building the back end. Database Designer.

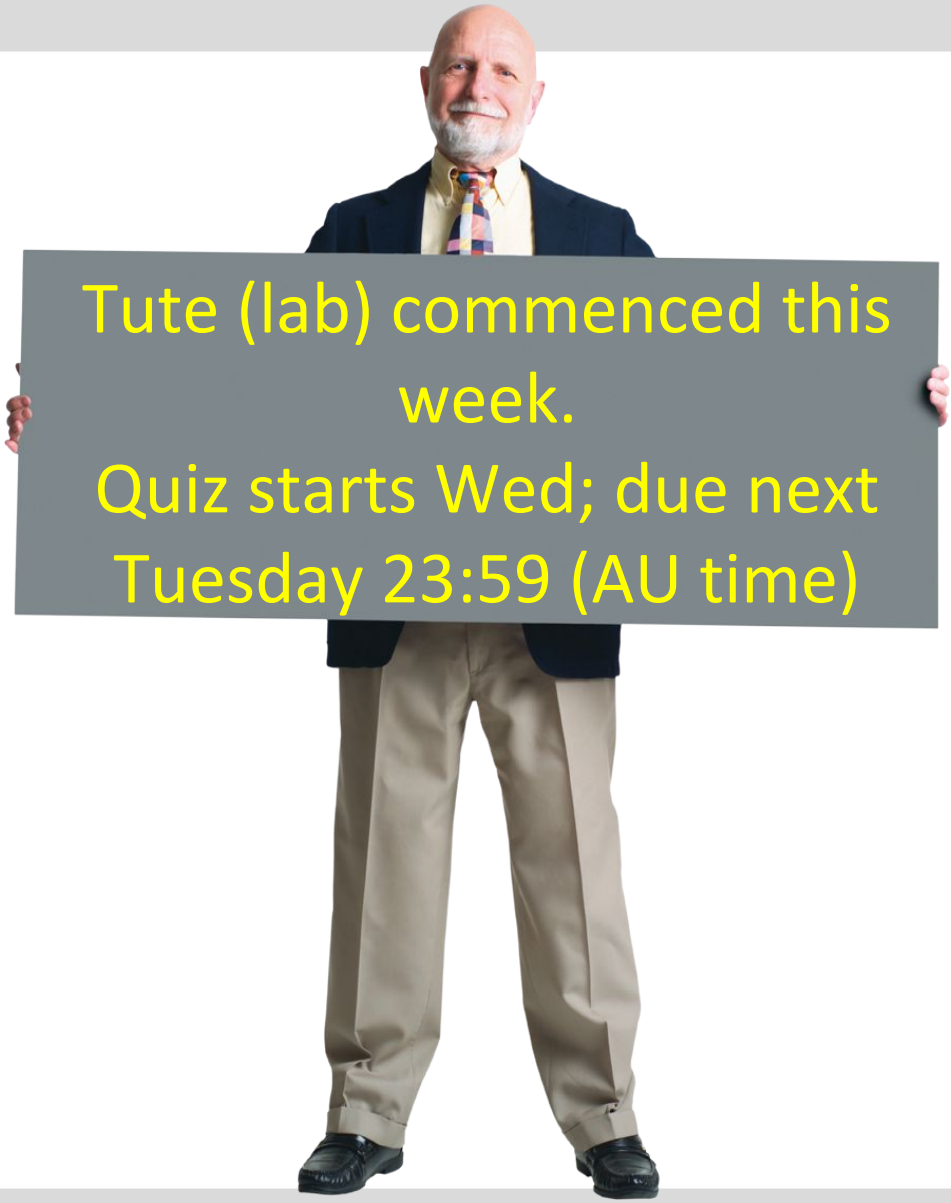
Our Database Systems Environment



Our Database Systems Environment - VPN usage



Adapted/simplified from original diagram © Mark Creado / Lindsay Smith. Generated with creately.


Tute (lab) commenced this week.
Quiz starts Wed; due next Tuesday 23:59 (AU time)