MONASH University

MONASH
INFORMATION
TECHNOLOGY

Week 1 - Introduction
FIT9132 Introduction to Databases

GROUP OF EIGHT AUSTRALIA

Your FIT9132 Teaching Team - Caulfield Campus

Chief Examiner & Lecturer

Lindsay Smith

Lecturer & Head Tutor

Dwi Rahayu

Lecturer & Tutor

Manoj Kathpalia

To contact the lecturing team for FIT9132 Administration matters - assignment extensions, absences, class issue etc please email: **fit9132.allcampuses-x@monash.edu**

Overview

- Unit Guide
  - –During the semester your first contact should be your tutor, be sure to obtain their email address from Moodle so you can email them if necessary
  - –**Note** the FIT9132 Email requirements:
    "When you contact your tutor (or lecturer) via email, please ensure you clearly include your full name, unit code and lab number as part of every email you send. This will ensure we can respond as quickly and accurately as possible."
    email which does not comply will **not be responded to**
- Moodle
- Teaching Method (Peer Instruction in Lecture)
- A summary of topics to be studied

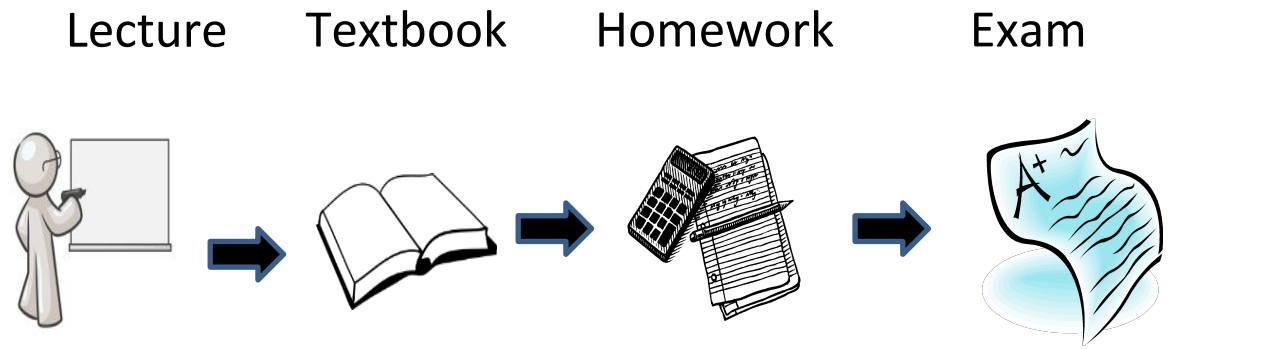
Teaching Method

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

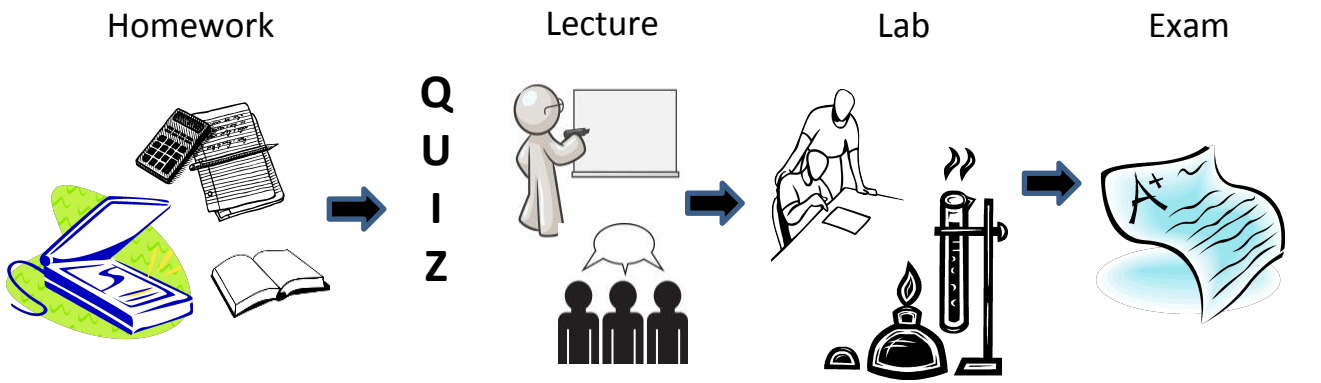
Peer Instruction

Prof Eric Mazur, Harvard University

Traditional Teaching Method



Peer Instruction – Full Picture



Discussion Questions – Scenario

- Lecturer shows a question.
- Student answers using the response system. (no discussion – individual vote).
- If uncertainty
  - –Group discussion (2-3 students) – need to get a consensus.
  - –Student answers using the response system (group vote – everyone in the group still needs to vote).
  - –Class wide discussion.

Why The Scenario?

- Pose carefully designed question
  - –Solo vote: 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
    - •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.

Let's Practice

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.

Multiple choice questions

Q1: 1 + 1 = ?

Hint: There are 10 types of people in this world. Those who understand binary and those who don't.

a. 2
b. 10
c. 11
d. Not sure

Multiple choice questions

Q2: If the following equations are true,

5 + 3 = 28
9 + 1 = 810
8 + 6 = 214
5 + 4 = 19

what is 3 + 2?

a. 5
b. 15
c. 11
d. 55

Text-based poll

Q3: Write the name of your favourite fruit.

Q4. What database management systems are you most familiar with?

a. Oracle
b. MySQL
c. MS Access
d. SQL Server
e. others
f. I am not familiar with any database management systems.

Is it bad to get it WRONG?
NO

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!

Why Peer Instruction?

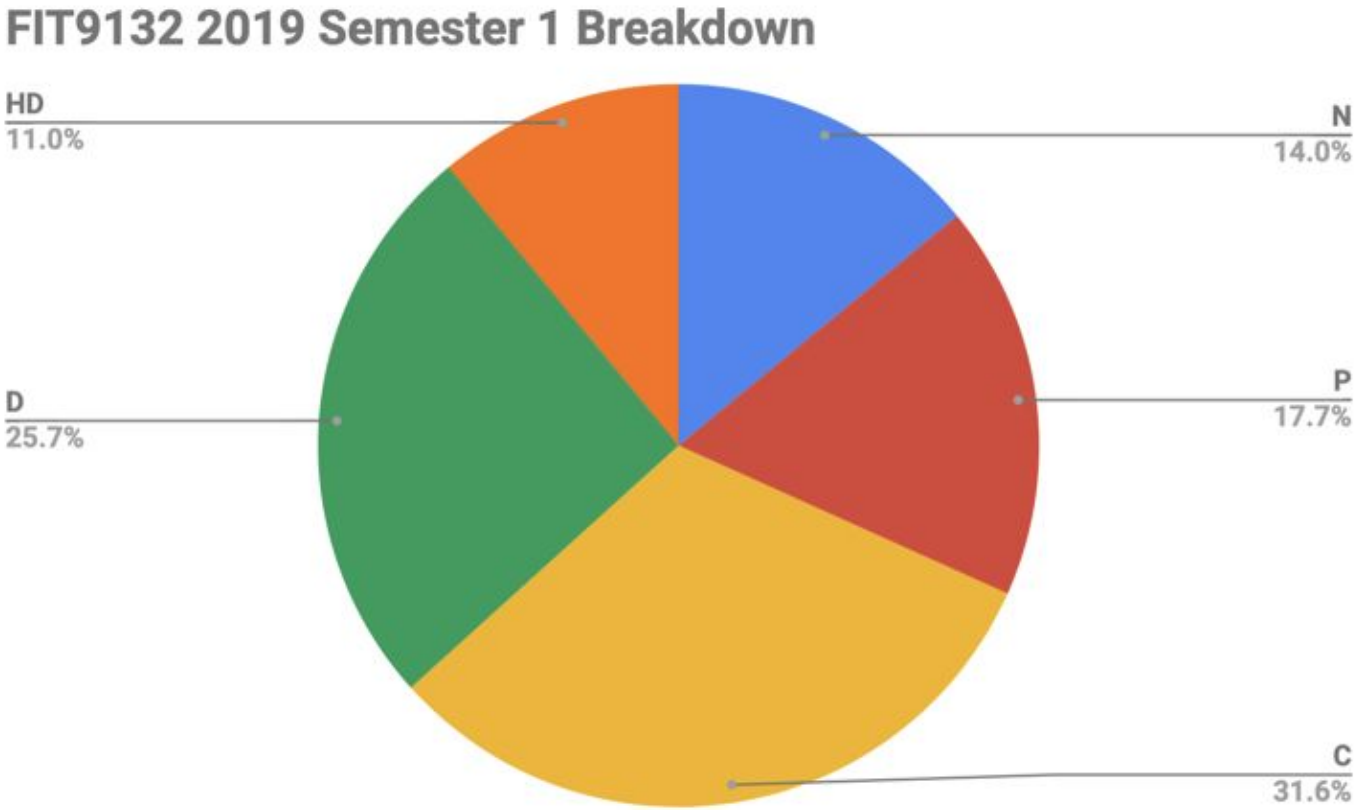
- Learn/practice hard concepts in class
- Build and test one's understanding in a supportive environment.
- Develop critical thinking, communication and reflection skills.
- Engage students to take ownership of their learning.

Things are different…

- Pre-lecture activities are crucial.
  - –Your lecture experience will depend on your preparation.
- Attending lectures is very important
- My 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.

Study Program

Week	Activities	Assessment	
0		No formal assessment or activities are undertaken in week 0	
1	Introduction to database		
2	**PART I: Database Design** Database Design I - Conceptual Model	Pre-lecture Quiz Questions due weekly prior to the lecture (Weeks 2 to 11),	5%
3	Relational Model		
4	Database Design II - Logical Design		
5	Database Design III - Normalisation	Assignment 1 - Conceptual Model due	10%
6	Database Implementation - DDL		
7	**PART II: The SQL Database Language** SQL I - Basic		
8	Update, Delete and Transaction Management	Assignment 2 - Database Design due	15%
9	SQL II - Intermediate		
10	Triggers and Oracle PL/SQL		
11	SQL III - Advanced		
12	Topics on Big Data	Assignment 3 - SQL due	20%



Enrolment: 844 students
Average Grade: 63% C

MONASH University

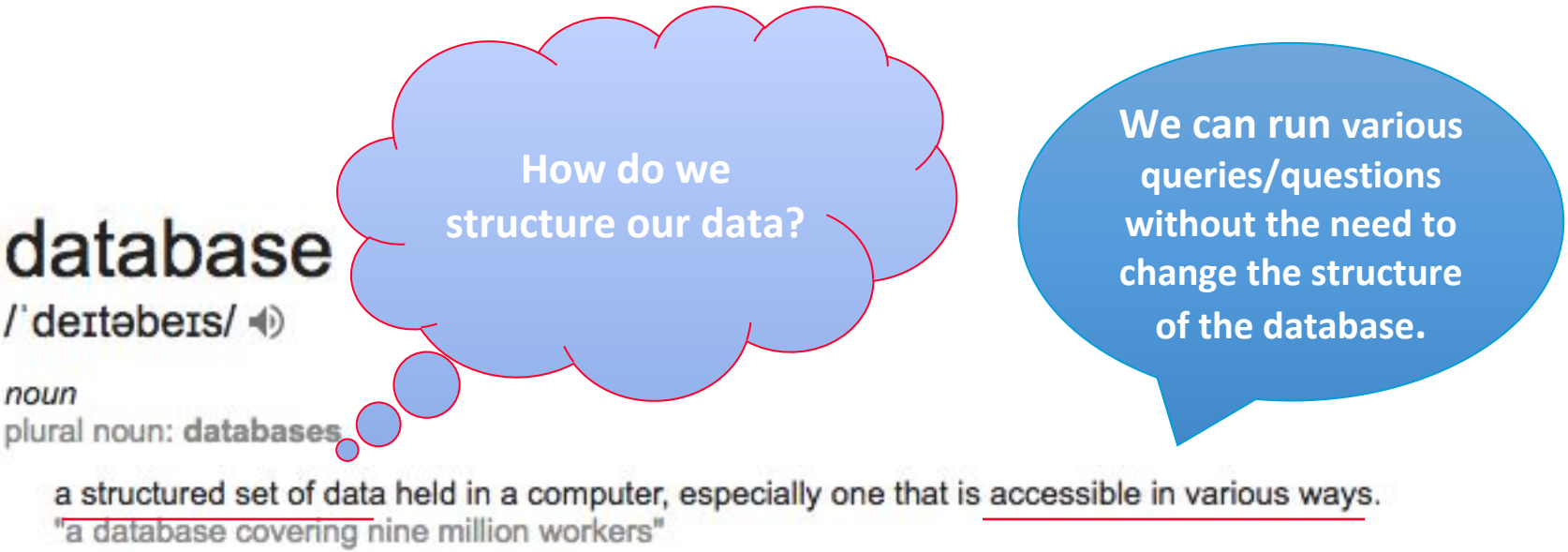
Overview

- An overview of relational database management systems (RDBMS)

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 FIT9133 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

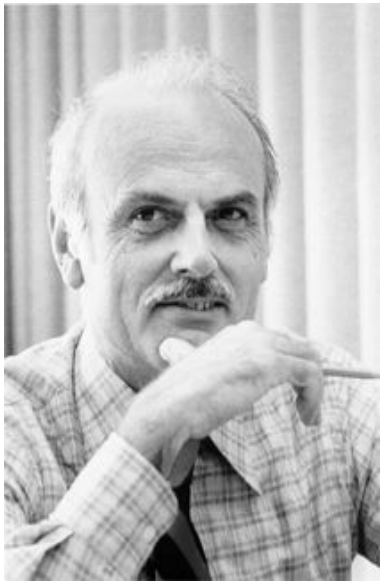
UNIT_CODE	UNIT_NAME
FIT1002	Computer Programming
FIT1001	Computer Systems
FIT1004	Database

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 3
- Normalisation in week 5
- E. F. Codd, "A Relational Model of Data for Large Shared Data Banks", *Comm. Of ACM*, 1970

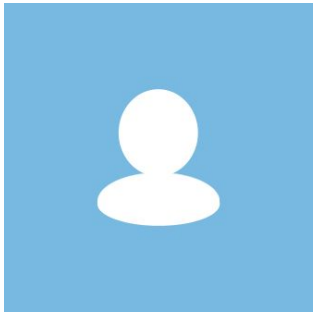
1974: SQL

- 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 2 & 4

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

Peter Chen (1947 -)

1979: Oracle

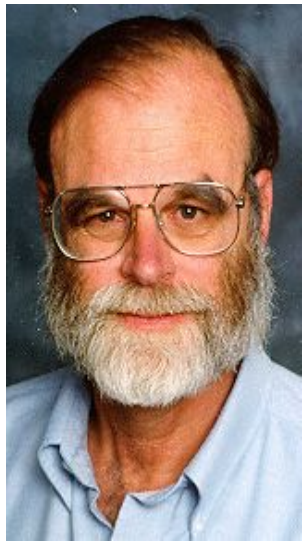
- Inspired by Codd's ideas
- First commercial release in 1979
- Most popular RDBMS
- Introduced PL/SQL in 1988 (Procedural Language/SQL)

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

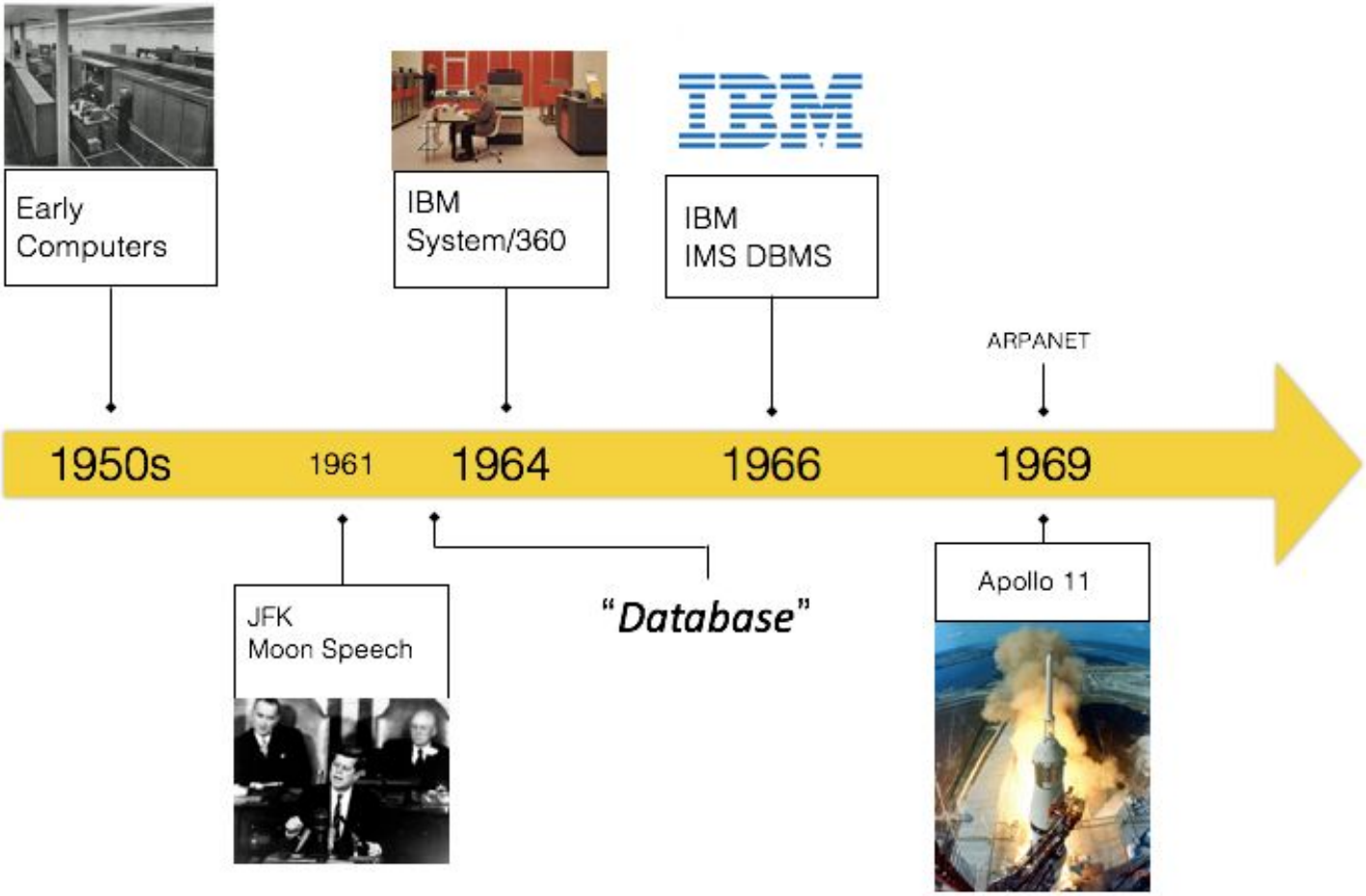
- Jim Gray, "The Transaction Concept: Virtues and Limitations ", *VLDB*, 1981

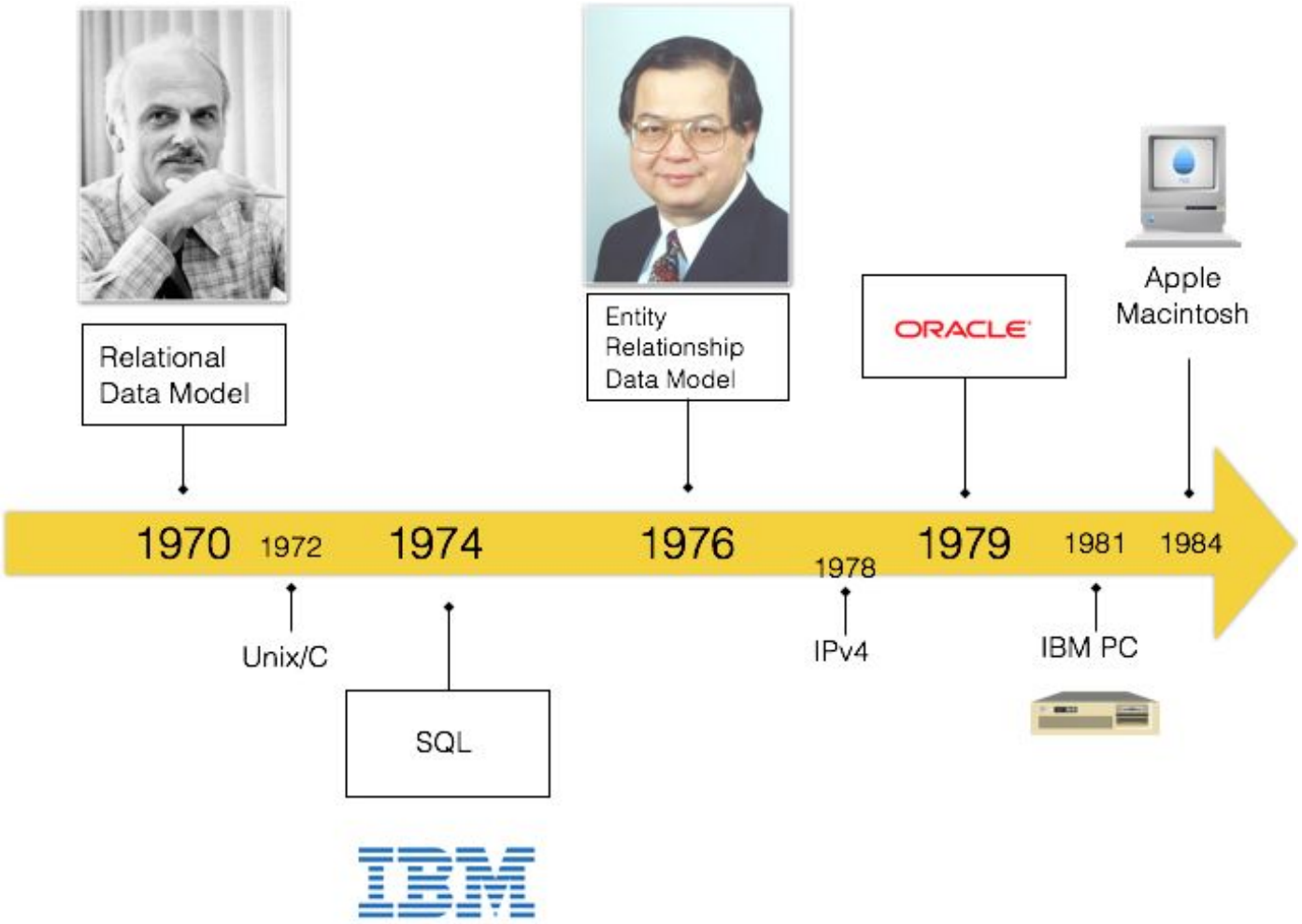
Data Management Today

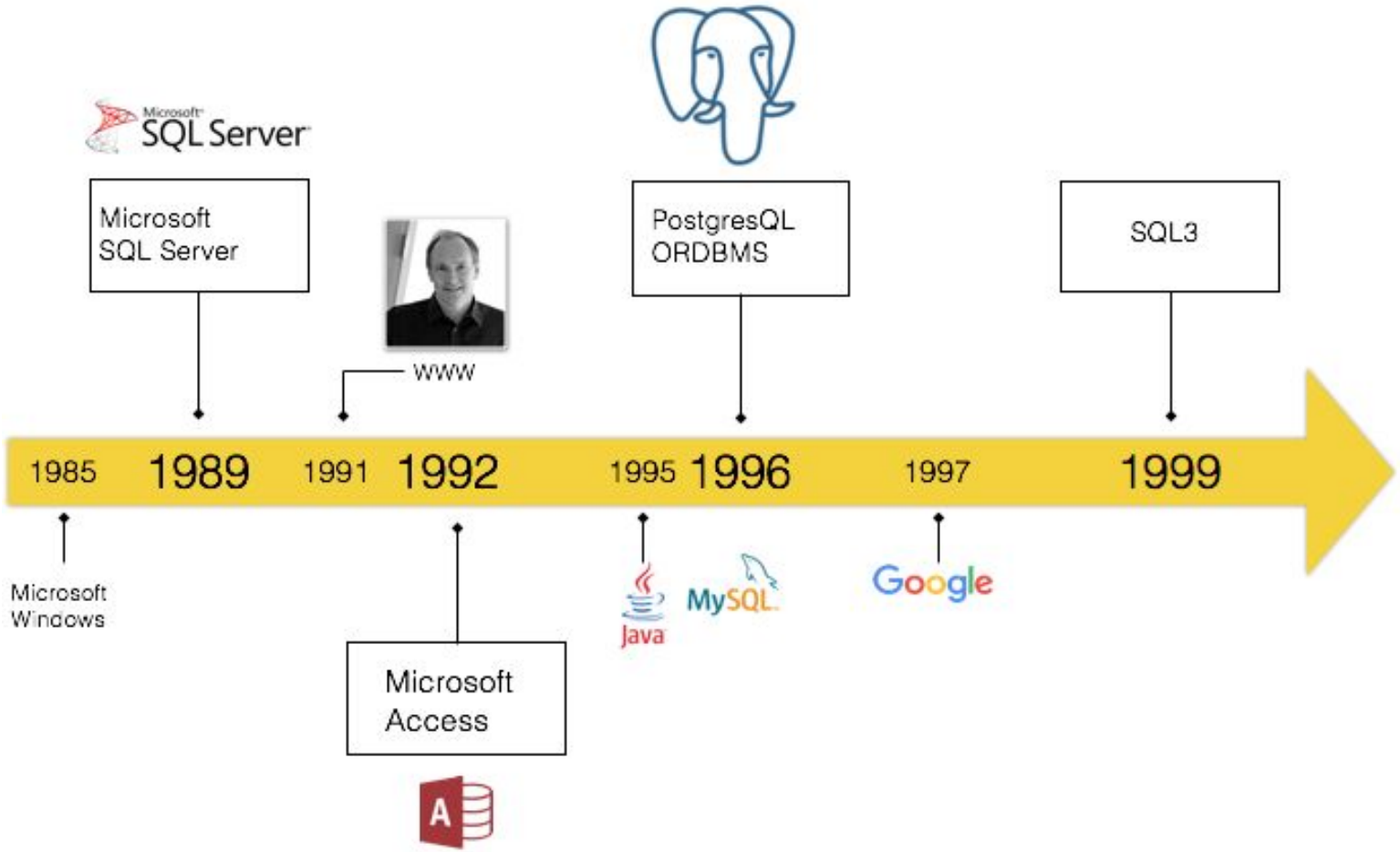
- Relational databases are still very popular. But …
  - –Social Networks (Facebook, Twitter, Foursquare etc.)
  - –Multimedia data (YouTube, Pinterest, Facebook etc.)
  - –Data streams (Twitter, computer networks)
  - –Spatial data (Road networks, Google Earth, Space etc.)
  - –Textual data
  - –Web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

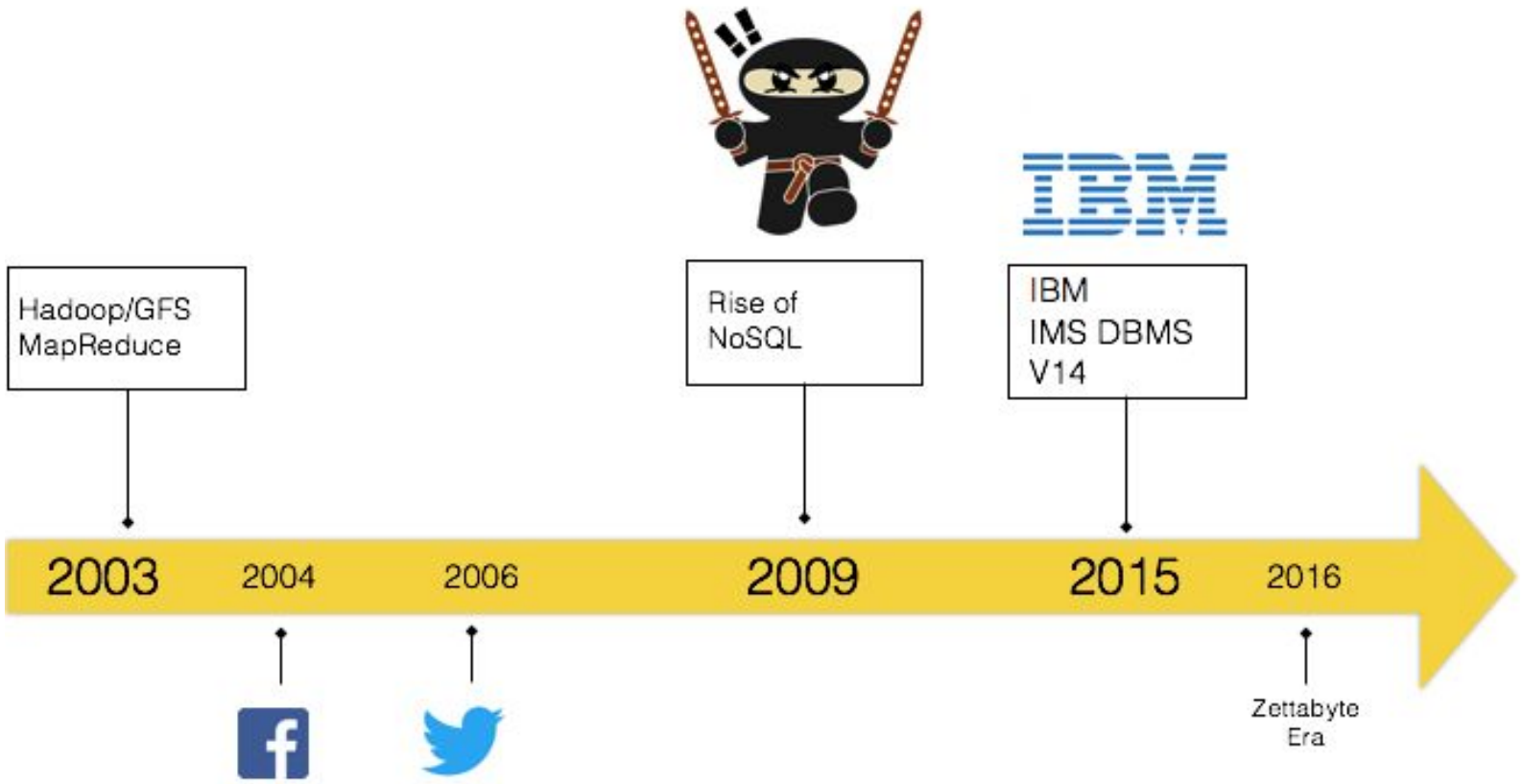
https://goo.gl/zMxG3b

In Perspective ...




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

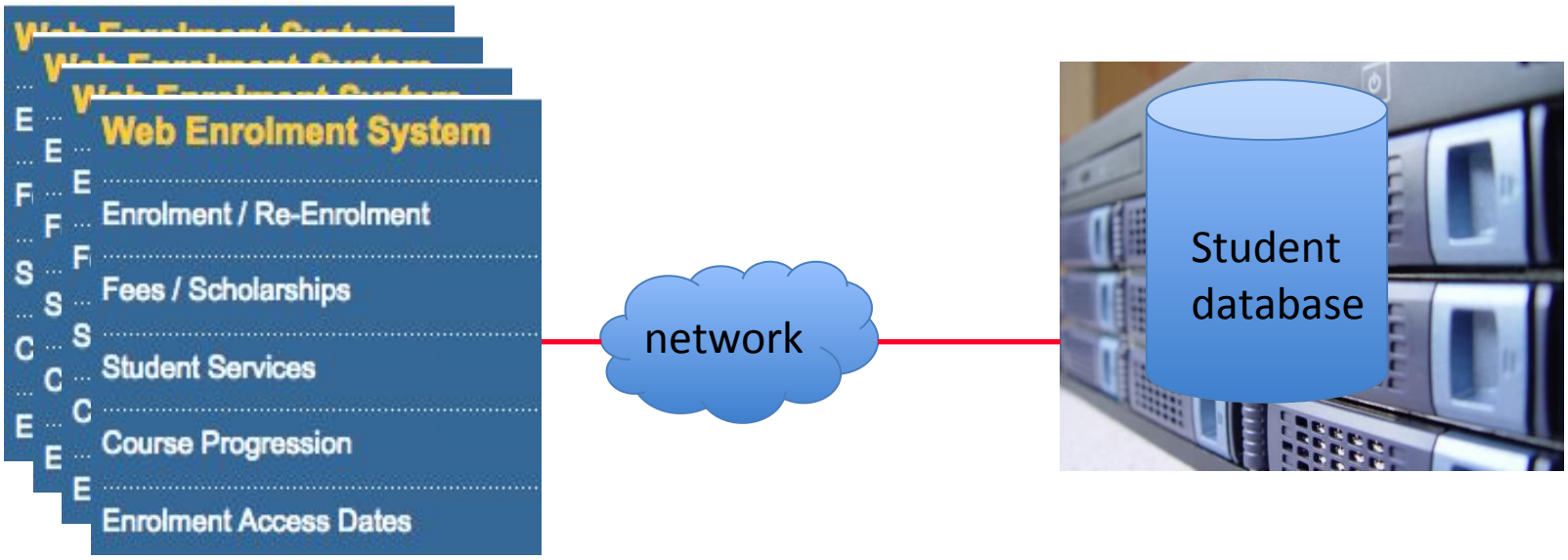

Microsoft SQL Server
Microsoft SQL Server
www
PostgresQL ORDBMS
SQL3
1985
1989
1991
1992
1995
1996
1997
1999
Microsoft Windows
Microsoft Access
Java
MySQL
Google


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

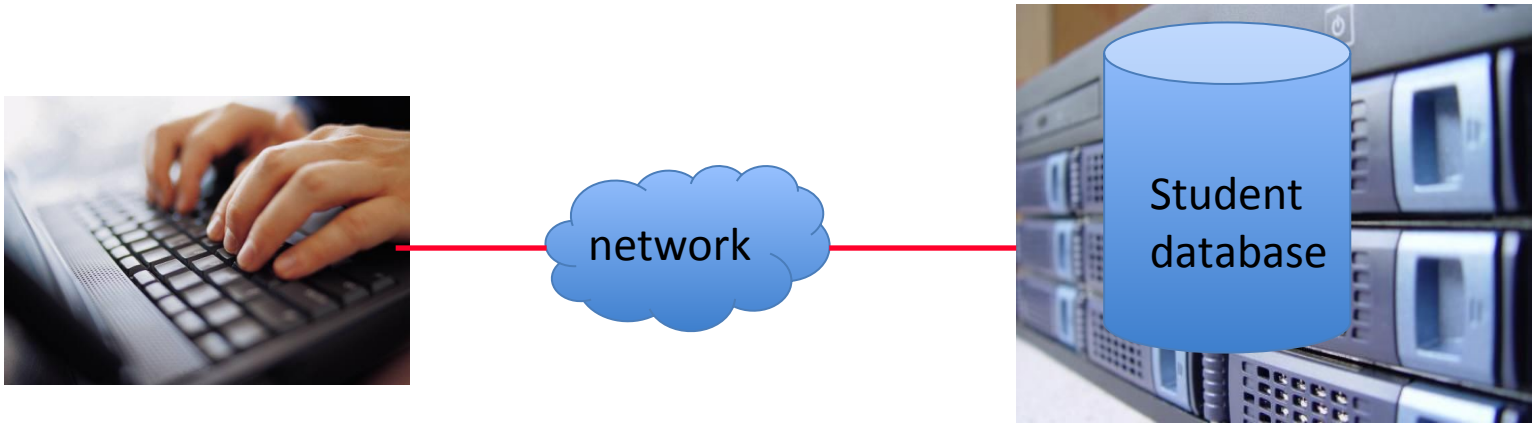
Relational database systems in action: End-users' view



Front end application (client)

Student Database is implemented in an Oracle DBMS (server)

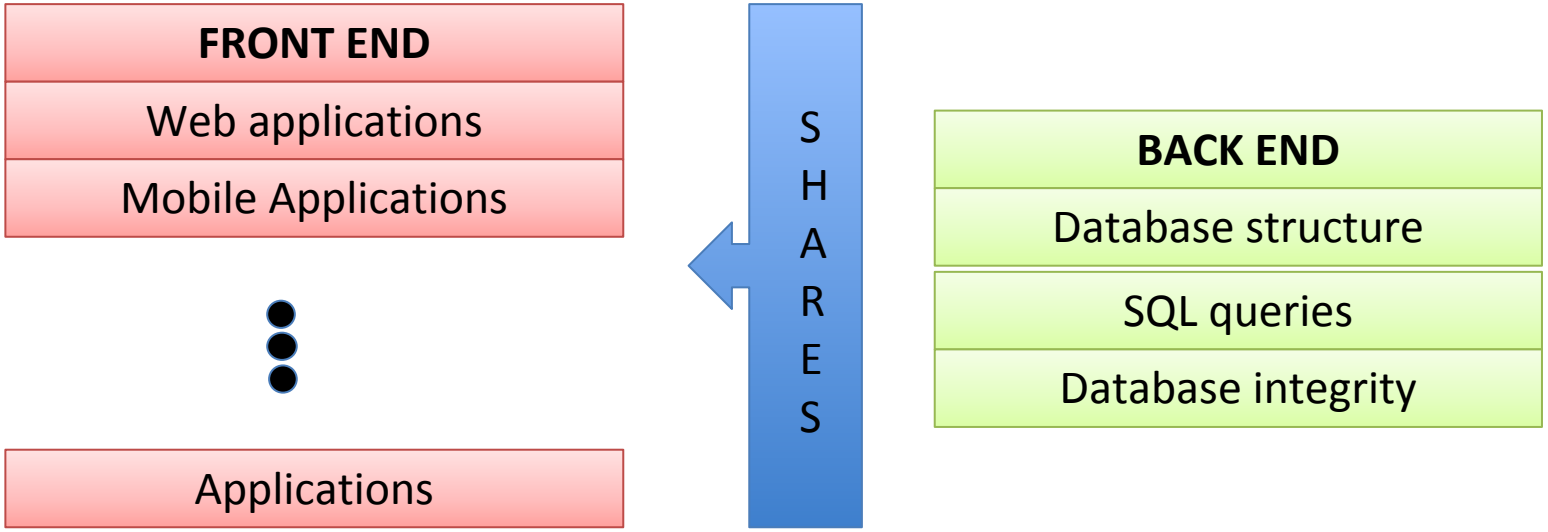
Database Systems in Action
Developers' View



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

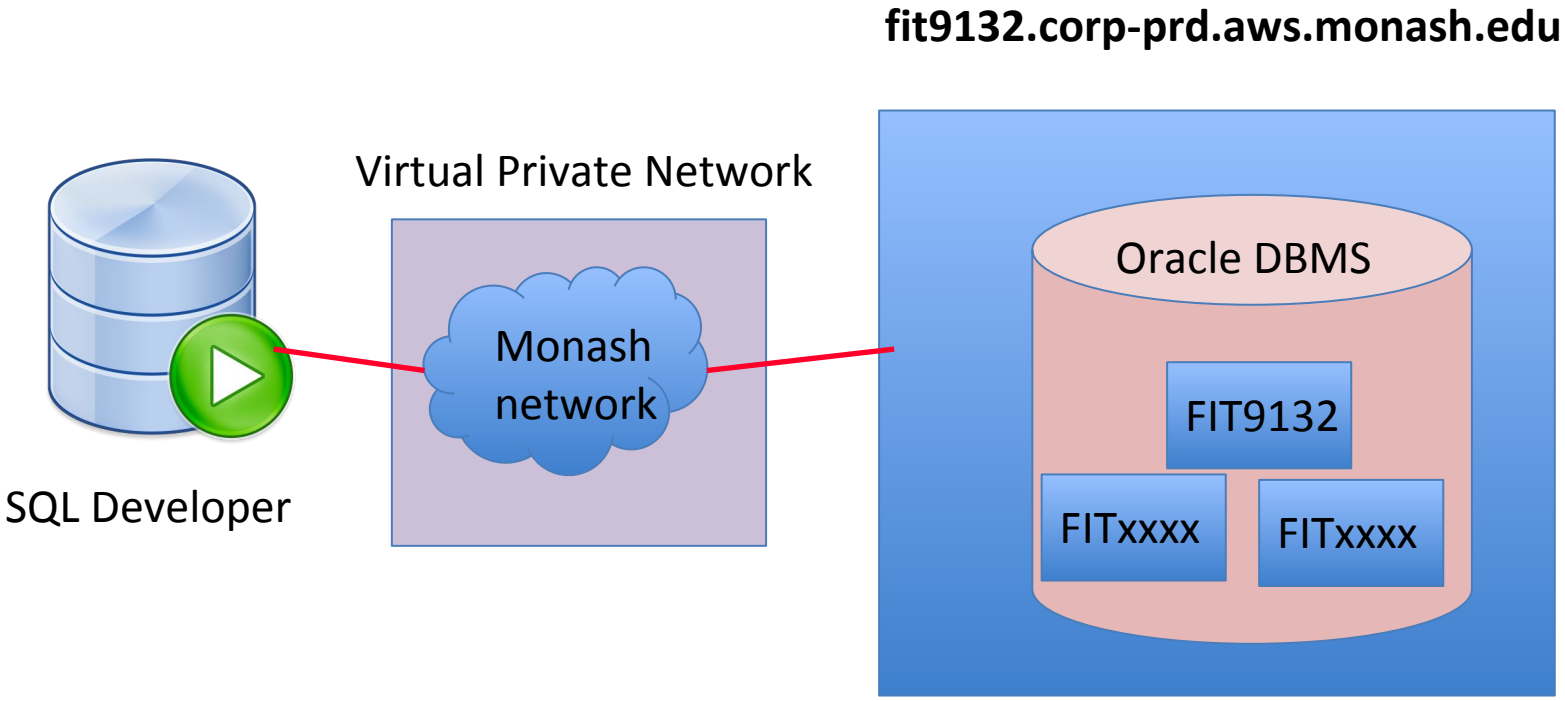
Student Database (server)

Developing Application with Database



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

Our Database Systems Environment




Tutorials/Labs
start this week