

SFWRENG 3DB3

DATABASE SYSTEMS

FALL 2021

Fei Chiang (fchiang@mcmaster.ca)

1

Administration

2

- Instructor: Fei Chiang
- Course website: MS Teams + Avenue
 - MS Teams: Lecture recordings, tutorial slides, lecture slides, assignments, policies
 - Avenue: Assignment submission and grading
- Lectures:
 - Mon/Wed/Thurs at 1:30pm-2:20pm
- Tutorials:
 - (T01) Tues: 1:30-2:20pm
 - (T02) Mon: 12:30-1:20pm
 - (T03) Wed: 11:30am – 12:20pm
 - Note: Tutorials start next week (week of Sept 13th)

2

Admin (cont'd)

3

- Teaching Assistants:
 - Morteza Alipour Langouri (alipoum@mcmaster.ca)
 - Levin Noronha (noroni@mcmaster.ca)
 - Lucia Cristiano (cristial@mcmaster.ca)
 - Office hours: listed on course info sheet
- Textbook: Database Management Systems (3rd edition) by R. Ramakrishnan, J. Gehrke.
- Bookstore:
 - https://campusstore.mcmaster.ca/cgi-mcm/ws/txsub.pl?wsTERMG1=214&wsDEPTG1=SFWRENG&wsCOURSEG1=3DB3&wsSECTIONG1=DAY%20C01&crit_cnt=1

3

Grading

4

	Asg1	Asg2	Asg3	Total
Assignments	12%	14%	14%	40%
Midterm	20%		20%	
Final Exam	40%		40%	

Midterm: Thurs Oct. 21, 2021 during lecture time.

4

Lectures, Tutorials, Office Hours

5

- Microsoft Teams
- Channel for:
 - Lectures
 - Each tutorial
 - Each TA and instructor office hour
- Q & A

5

Assignments

6

- Will be posted on Avenue
- Submit through Avenue
- Late policy:
 - Marked with a late penalty of 20% per day
 - No assignments will be accepted beyond 5 days past the due date
 - Do not wait until deadline to raise problems
- Re-marking
 - Within 7 days of returning the assignment

6

Plagiarism

7

- You are encouraged to talk to your fellow students, but submitted work must be based on your own ideas and conclusions.
- Plagiarism and cheating are serious academic offenses, and will be handled accordingly.
- When you submit assignments with your name on it, you are certifying that you have completed the work for that assignment yourself.
- Will use Avenue for assignment submission

7

Questions

8

- If something is unclear, please do ask questions in class!
- E-mail is preferred contact method (fchiang@mcmaster.ca)
- Office hours: Mon 11:30 – 12:30pm
- Feedback is encouraged. If something is concerning you, please let me know early!

8

Topics

- Relational Model
- Database Design
- E-R Model
- Transactions
- SQL
- Concurrency
- Views, Indexes, Constraints
- Advanced Topics
- Relational Algebra
- Data Mining
- Sustainability

We will be using IBM DB2

"Accessing
DB2 Servers"
in Avenue

9

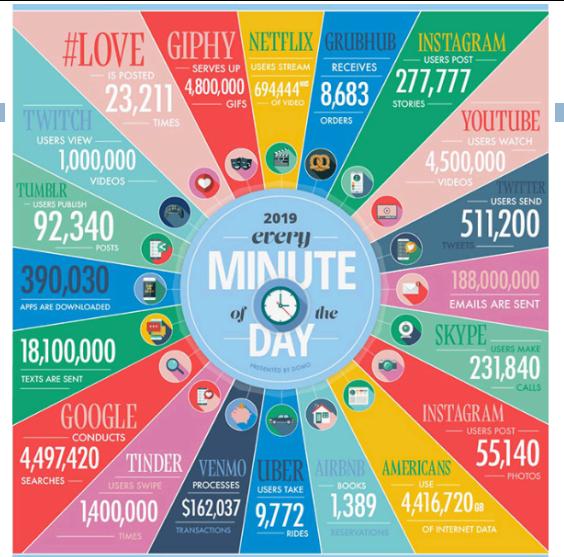
2.5M Terabytes

(or equivalently 2.5 quintillion bytes)

Amount of data
generated every day

5M Terabytes of
data generated up
to 2003

10



11

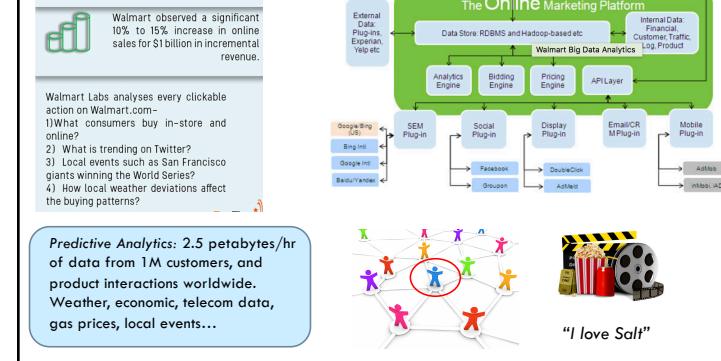
Walmart

The analysis covers millions of products and 100's of millions customers from different sources.

Walmart observed a significant 10% to 15% increase in online sales for \$1 billion in incremental revenue.

- Walmart Labs analyses every clickable action on Walmart.com-
- 1) What consumers buy in-store and online?
 - 2) What is trending on Twitter?
 - 3) Local events such as San Francisco giants winning the World Series?
 - 4) How local weather deviations affect the buying patterns?

Predictive Analytics: 2.5 petabytes/hr of data from 1M customers, and product interactions worldwide. Weather, economic, telecom data, gas prices, local events...



13

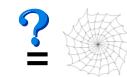
What Is a Database System?

14

- Database:
a very large, integrated collection of data.
- Models a real-world enterprise
 - Entities (e.g., teams, games)
 - Relationships
(e.g., Barack Obama received the Nobel Peace Prize)
- A Database Management System (DBMS) is a software system designed to **store, manage, and facilitate access to** data.

Credit: Renee J. Miller

14



Is the WWW a DBMS?

15

- Fairly sophisticated search available
 - crawler indexes pages on the web
 - Keyword-based search for pages
- But, currently
 - data is mostly unstructured and untyped
 - search only:
 - can't modify the data
 - can't get summaries, complex combinations of data
 - few guarantees provided for freshness of data, consistency across data items, fault tolerance, ...
 - Web sites (e.g. e-commerce) typically have a DBMS in the background to provide these functions.

15

Search vs. Query

16

- What if you wanted to find out how to donate to help victims of hurricane Dorian?
- Search for “**hurricane Dorian donations**” in your search engine.



[Hurricane Dorian: Disaster Relief & Donations | American Red Cross](#)
<https://www.redcross.org/donate/hurricane-dorian-donations> •
 Help people affected by Hurricane Dorian. Your donation enables the Red Cross to prepare for, respond to and help people recover from this disaster... Contributions to the American National Red Cross are tax-deductible to the extent permitted by law.

[How You Can Help the Bahamas Recover From Hurricane Dorian](#)
[https://www.miaminewstimes.com/news/miami-groups-launch-fundraiser...>](https://www.miaminewstimes.com/news/miami-groups-launch-fundraiser...)
 6 hours ago · Donations are also being accepted online at the Hurricane Dorian Relief Fund organized by The Smathers Trust, a Miami-based nonprofit that ...

[Here's how to donate, help Bahamians recover from Dorian - WPLG](#)
[https://www.local10.com/weather/hurricane/heres-how-to-donate-h...>](https://www.local10.com/weather/hurricane/heres-how-to-donate-h...)
 17 mins ago · Videos show Category 5 Hurricane Dorian's fury in Bahamas... has set up an online fund to donate money for their hurricane relief efforts.

[Donations begin to help Bahamians post Hurricane Dorian - YouTube](#)
[https://www.youtube.com/watch...>](https://www.youtube.com/watch...)
 9 hours ago · The South Florida community is beginning to collect relief supplies for the people of the Bahamas as Hurricane Dorian continues to cause ...

[Hurricane Dorian : Charity Navigator](#)
[https://www.charitynavigator.org...>](https://www.charitynavigator.org...)
 Hurricane Dorian, the first major hurricane of the 2019 storm season, ... Designated donations made from this page will be applied to charity programs per each ...

[Hurricane Dorian: How You Can Help - Donate Blood - The Blood...](#)
[https://bloodconnection.org/hurricane...>](https://bloodconnection.org/hurricane...)
 3 hours ago · As evacuations begin for those in the path of Hurricane Dorian, The ... the areas where TBC is unable to collect donations during the storm.

[Hurricane Dorian: Unused supplies can be donated to charitable ...](#)
[https://www.newspress.com/story/weather/hurricane/2019/09/01/h...>](https://www.newspress.com/story/weather/hurricane/2019/09/01/h...)
 23 hours ago · With this region out of the cone, residents are urged to donate some of ... Hurricane Dorian. Surplus of storm supplies provides opportunity to ...

[Investigate before you donate to Hurricane Dorian disaster relief |...](#)
[https://kron4news.com/news/consumer/investigate-before-you-donate-to-hu...>](https://kron4news.com/news/consumer/investigate-before-you-donate-to-hu...)
 3 days ago · Those of us who are not in the path of a major hurricane can only imagine the fear and hope as we watch people prepare for the worst. We've ...

16

Search

17

- Based on keyword matching
 - Our search matches warnings about donations to legitimate places
 - Ranking of results
 - Popularity or reputation
- Web documents
 - Limited structure

17

Search vs. Query

18

- “Search” : returns stored documents “as-is”



08.29.2019 | Emergency Management

**Hurricane Dorian is on the way and
Verizon is ready**



We're ready to support V Teamers

We're also ready to help employees in times of need through the [VtoV Employee Relief Fund](#). The charity provides grants for Verizon employees displaced from their homes due to a natural or personal emergency - such as fire, flood, severe weather or domestic violence. VtoV is entirely supported by employee donations and the Verizon Foundation's generous matching gift program, with 100% of all donations going to help employees.

18



Is a File System a DBMS?

19

- Thought Experiment 1:

- You and your project partner are editing the same file.
- You both save it at the same time.
- Whose changes survive?

A) Yours B) Partner's C) Both D) Neither E) ???

- Thought Experiment 2:

- You're updating a file.
- The power goes out.
- Which of your changes survive?

Q: How do you write programs over a subsystem when it promises you only “???” ?
A: Very, very carefully!!

A) All B) None C) All Since last save D) ???

19

Why Use a DBMS?



- Data independence and efficient access.
- Reduced application development time.
- Data integrity and security.
- Concurrent access, recovery from crashes.

20

Why Study Databases??

21

- Shift from computation to information
 - always true for corporate computing
 - Web made this point for personal computing
 - more and more true for scientific computing
- Strong need for DBMS
 - **Corporate:** retail swipe/clickstreams, “customer relationship mgmt”, “supply chain mgmt”, “data warehouses”, Big Data, etc.
 - **Scientific:** digital libraries, Human Genome project, Sloan Digital Sky Survey, physical sensors
- DBMS encompasses much of CS in a practical discipline
 - OS, languages, theory, machine learning, logic
 - Yet traditional focus on real-world apps



21

What's the intellectual content?

22

- representing information
 - data modeling
- languages and systems for querying data
 - complex queries with real semantics
 - over massive data sets
- concurrency control for data manipulation
 - controlling concurrent access
 - ensuring *transactional semantics*
- reliable data storage
 - maintain data semantics even if you pull the plug



Describing Data: Data Models

23

- A data model is a collection of concepts for describing data.
- A schema is a description of a particular collection of data, using a given data model.
- The relational data model is the most widely used model today.
 - Main concept: relation, basically a table with rows and columns.
 - Every relation has a schema, which describes the columns, or fields.

22

23

Data Independence

24

- Applications insulated from how data is structured and stored.
- Logical data independence: Protection from changes in *logical structure* of data.
- Physical data independence: Protection from changes in *physical structure* of data.
- Q: Why is this particularly important for DBMS?

Because rate of change of DB applications is slow. More generally:
 $dapp/dt \ll dplatform/dt$

24