

SFWRENG 3DB3
INTRO TO DATABASES
FALL 2024

Fei Chiang (fchiang@mcmaster.ca)

Administration

2

- Instructor: Fei Chiang
- Course website: Avenue
 - ▣ Tutorial slides, lecture slides, assignments, policies
 - ▣ Assignment submission and grading
- Lectures:
 - ▣ Mon/Wed/Thurs at 10:30am - 11:20am
- Tutorials start next week
- Teaching Assistants:
 - ▣ Office hours: TBA, see course info sheet
- Textbook: Database Management Systems (3rd edition) by R. Ramakrishnan, J. Gehrke.

Grading

3

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

Midterm: Thurs. Oct. 24, 2024 during lecture time.

Assignments

4

- 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

Plagiarism

5

- 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

Generative AI

6

- Foundation models for brainstorming
- All submitted content, analysis, conclusions must be developed independently by you
- LLMs are prone to generate false conclusions, misinformation
- Any use without citation will be considered academic dishonesty

Questions

7

- If something is unclear, please do ask questions in class!
- E-mail is preferred contact method (fchiang@mcmaster.ca)
- Office hours: Thurs 11:30am – 12:30pm in ITB 122
- Feedback is encouraged. If something is concerning you, please let me know early!

Topics

8

- Relational Model
- E-R Model
- SQL
- Views, Indexes, Constraints
- Relational Algebra
- Database Design
- Transactions
- Concurrency
- Advanced Topics
 - ▣ Data Mining
 - ▣ ML 101

We will be using IBM DB2
se3db3.cas.mcmaster.ca

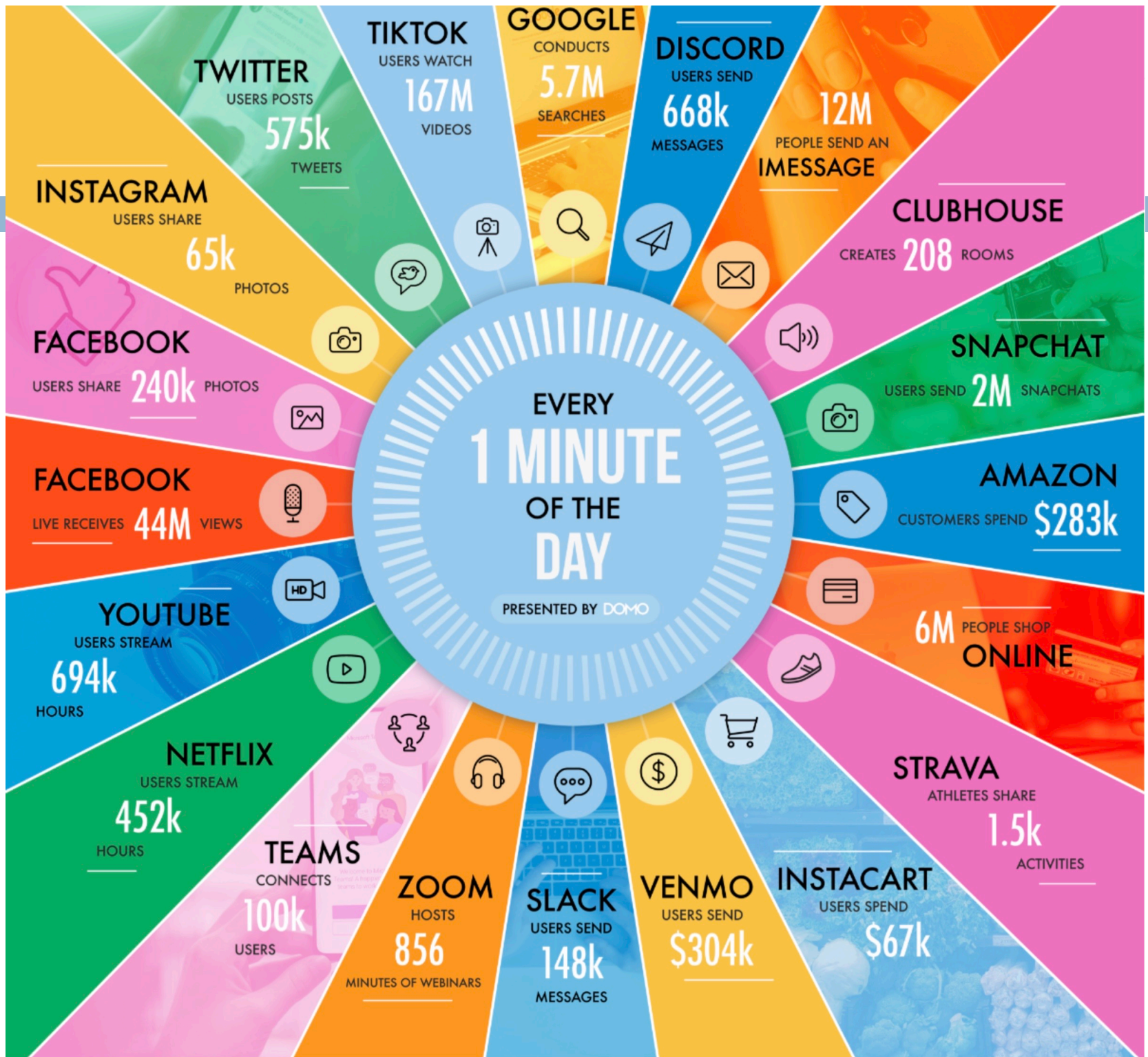
**Accessing
DB2 Servers
(see Avenue
doc)**

2.5M Terabytes
(or equivalently 2.5 quintillion bytes)

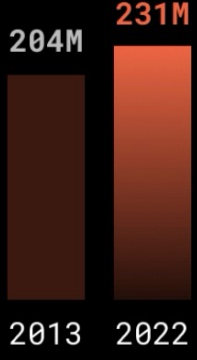
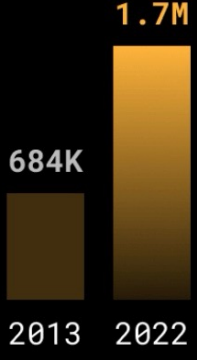
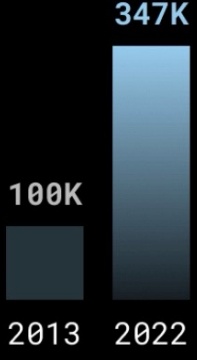
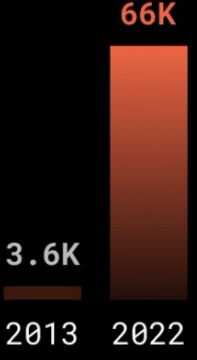
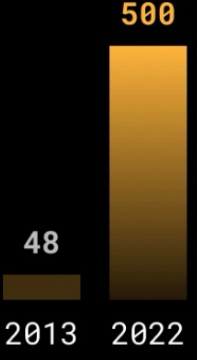
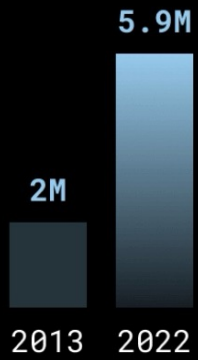
Amount
generate

By 2025, it's estimated to be
463 exabytes (463M TB)

tes of
ted up
3



Data Never Sleeps 1.0 vs. Data Never Sleeps 10.0



GOOGLE
USER QUERIES

YOUTUBE
HOURS UPLOADED

INSTAGRAM
PHOTOS SHARED

TWITTER
TWEETS SHARED

FACEBOOK
CONTENT SHARED

EMAILS
EMAILS SENT

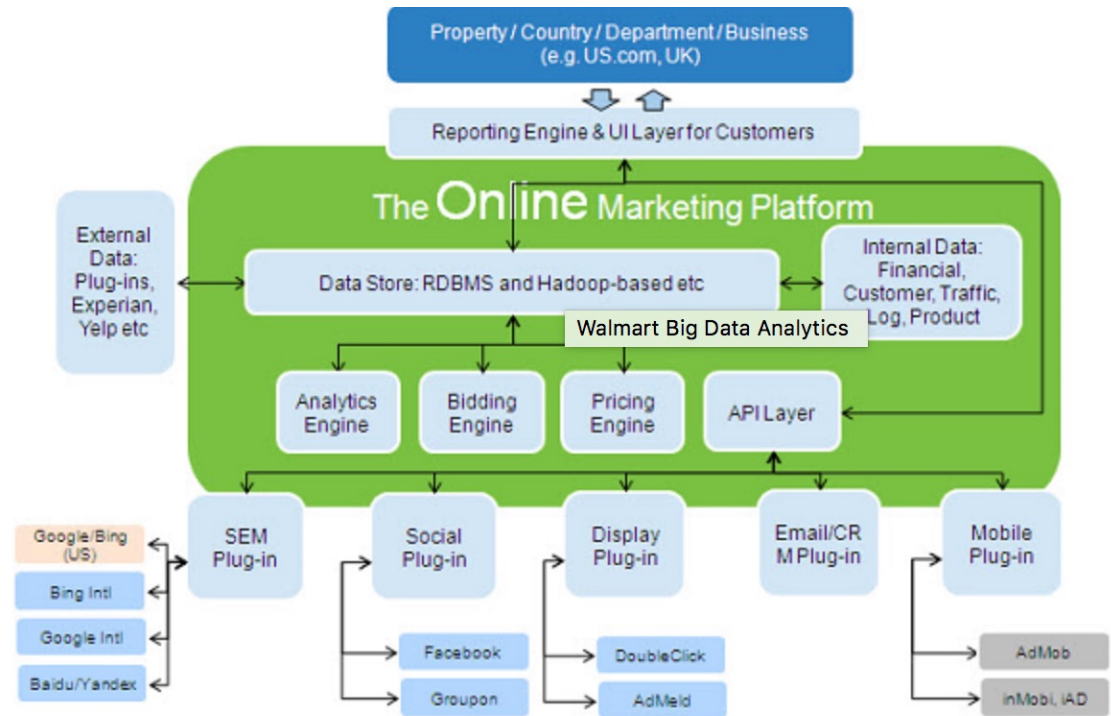
The analysis covers millions of products and 100's of millions customers from different sources.



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...

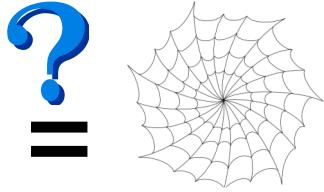


"I love Barbie"

What Is a Database System?

13

- 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.



Is the WWW a DBMS?

14

- 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.

Search vs. Query

15

- What if you wanted to find out how to donate to help victims of the Jasper wildfire?
- Search for “**Jasper wildfire donation**” in your search engine.



 Edmonton Humane Society
<https://www.edmontonhumanesociety.com> :


[Edmonton Humane Society: Home](#)

It takes approximately 7 million dollars to keep the Edmonton Humane Society operational for one year and 40% of these funds come from generous donations made ...

 The Anglican Journal
<https://anglicanjournal.com> > Archives :

[September 2024](#)

Donate. September 2024 Issue. Download Issue (PDF) ... Jasper wildfire destroys heritage church. A historic Anglican church in ...

 CBC
<https://www.cbc.ca> > news > canada > edmonton :


[Edmonton - CBC News](#)

Jasper wildfire latest disaster in climate claim deluge facing Canada's insurance sector ... donor. Nick Murray. Politics | September 1 · Ma-Me-O Beach ...

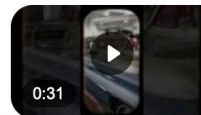
 Wikipedia
<https://en.wikipedia.org> > wiki > Jasper,_Alberta :

[Jasper, Alberta](#)

Donate · About Wikipedia · Disclaimers · Wikipedia. Search. Jasper, Alberta. Article ... "Buildings in Jasper in ashes as 'monster' wildfire spans 36,000 hectares ...

 YouTube · Jack Carter Chevrolet Buick GMC
180+ views · 1 week ago :

[Donating a 1962 Buick Special Convertible for Jasper Wildfire ...](#)



Jack Carter Chev is proud to **donate** a 1962 #Buick Special Convertible to help drive change in support of the #Jasper Wildfires relief ...

 Heritage Park
<https://heritagepark.ca> :

[Heritage Park Homepage | Heritage Park](#)

Heritage Park Offering Free Admission to Jasper Wildfire Evacuees | Click Here for More Details ... Donate. Book an Event. Book an Event. Back. Christmas and ...

 MSN
<https://www.msn.com> > en-us > video > animals > grizzl... :

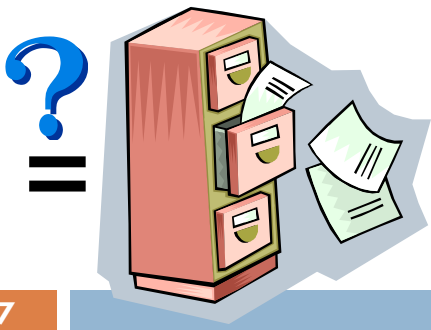
[Grizzly Bear And Cubs Survive Devastating Jasper Wildfire](#)

Jul 31, 2024 — Grizzly Bear And Cubs Survive Devastating Jasper Wildfire. Posted ... donation

Search

16

- Based on keyword matching
 - ▣ Our search matches relevant, and sometimes, less relevant pages
 - ▣ Ranking of results
 - ▣ Popularity or reputation
- Web documents
 - ▣ Limited structure



Is a File System a DBMS?

17

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) ???

Why Use a DBMS?



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

Complex Sub-system for Data Management

19

- 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

- 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.

Data Independence

21

- 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:
 $dapp/dt \ll dplatform/dt$

Concurrency Control

22

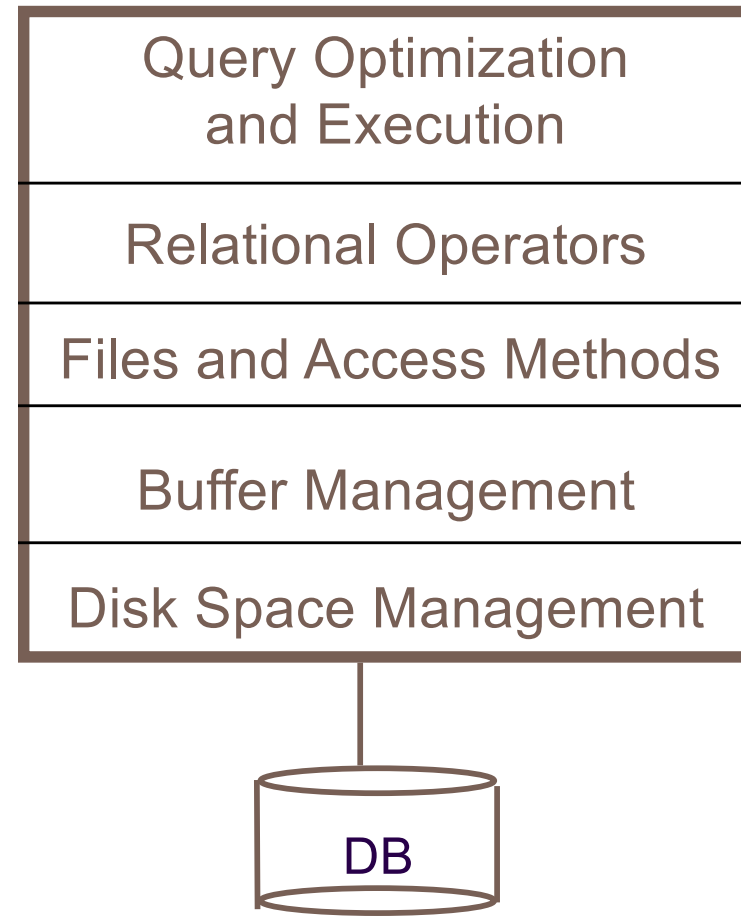
- Concurrent execution of user programs: key to good DBMS performance.
 - ▣ Disk accesses frequent
 - ▣ Keep the CPU working on several programs concurrently.
- Interleaving actions of different programs: trouble!
 - ▣ e.g., account-transfer & print statement at same time
- DBMS ensures such problems don't arise.
 - ▣ Users/programmers can pretend they are using a single-user system. (called “**Isolation**”)
 - ▣ Thank goodness! Don't have to program “very, very carefully”.

Structure of a DBMS

23

These layers must consider concurrency control and recovery

- A typical DBMS has a layered architecture.
- The figure does not show the concurrency control and recovery components.
- Each system has its own variations.



Summary

24

- DBMS used to maintain, query large datasets.
 - ▣ can manipulate data and exploit *semantics*
- Other benefits include:
 - ▣ Data independence,
 - ▣ quick application development,
 - ▣ data integrity and security,
 - ▣ recovery from system crashes,
 - ▣ concurrent access.
- Levels of abstraction provide data independence
 - ▣ Key when $d_{app}/dt \ll d_{platform}/dt$
- In this course we will explore:
 - ▣ How to be a sophisticated user of DBMS technology
 - ▣ What goes on inside the DBMS