CSIE30600/CSIEB0290 Database Systems

Lecture 1: Introduction

Outline

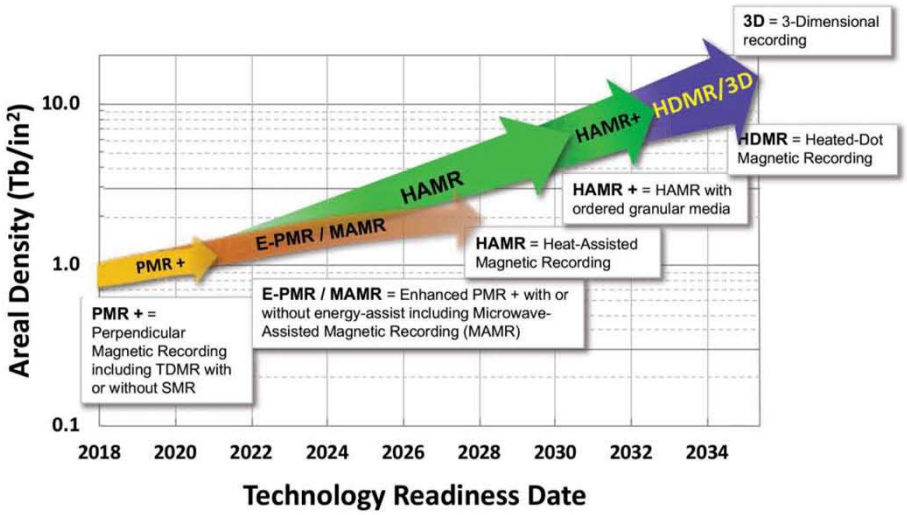
- Data is ubiquitous
- Basic definitions
- Types of databases and database applications
- Typical DBMS functionalities
- Example of a database (UNIVERSITY)
- Main characteristics of the database approach
- Database users
- Advantages of using the database approach
- History of database systems
- Extending database capabilities
- When not to use databases

Data vs Technological Advances

- **Five** classes of technological advances are changing our relationship with data:
- More storage space—allows us to keep more data
- Faster processor (and memory) speeds—allows us to access and process more data
- Better networking—allows us to share data more efficiently
- Different sensors—allows us to access all kinds of data at real-time
- Clever processing methods (AI & machine learning)—allows us to process data more intelligently

Introduction 3

HDD Density Growth



(https://ieeexplore.ieee.org/document/9918580, over 200TB HDDs by mid-2030's)

Introduction 4

HDD Cost (Historical)

Hard drive cost per GB over time

date	capacity	cost	$/GB
1957	3.75 MB	$34,500	$9.2 million/GB
1989	40 MB	$1,200	$30,000/GB
1995	1 GB	$850	$850/GB
2004	250 GB	$250	$1/GB
2011	2 TB	$70	$0.035/GB
2018	4 TB	$75	$0.019/GB

Disk Price 2025

Price per TB	Price	Capacity	Warranty	Form Factor	Technology	Condition	Affiliate Link
$6.663	$20	3 TB	3 months	Internal	SAS	Used	MDD MAXDIGITALDATA 3TB 7200RPM 128MB Cache SAS 6Gb/s 3.5inch Internal Enterprise Hard Drive (MD3TSAS12872E) - [NOT a SATA HDD] (Renewed)
$7.495	$15	2 TB	3 months	Internal	SAS	Used	MDD MAXDIGITALDATA 2TB 7200RPM 128MB Cache SAS 6Gb/s 3.5inch Internal Enterprise Hard Drive (MD2TSAS12872E) - [NOT a SATA HDD] (Renewed)
$7.760	$31	4 TB		Internal	SAS	Used	HGST Ultrastar 7K6000 HUS726040ALS210 4 TB 3.5 Internal Hard Drive
$7.837	$24	3 TB	5 years	Internal	SAS	Used	MDD MAXDIGITALDATA 3TB 7200RPM 128MB Cache SAS 6Gb/s 3.5inch Internal Enterprise Hard Drive (MD3TSAS12872E) [NOT a SATA HDD] (Renewed)
$8.182	$33	4 TB	5 years	Internal	SAS	Used	MDD MAXDIGITALDATA 4TB 7200RPM 128MB Cache SAS 12Gb/s 3.5inch Internal Enterprise Hard Drive (MD4TSAS12872E) - [NOT a SATA HDD] (Renewed)
$8.330	$25	3 TB		Internal	SAS	Used	HUS724030ALS640 HGST 3TB 7.2K 6G LFF SAS Hard Drive
$8.488	$34	4 TB	3 months	Internal	SAS	Used	MDD MAXDIGITALDATA 4TB 7200RPM 128MB Cache SAS 12Gb/s 3.5inch Internal Enterprise Hard Drive (MD4TSAS12872E) - [Not a SATA HDD] (Renewed)
$8.999	$90	10 TB	3 months	Internal	SAS	Used	MDD 10TB 7200RPM 256MB Cache SAS 12.0Gb/s 3.5inch Internal Enterprise Hard Drive (MDD10TSAS25672E) - [NOT a SATA HDD] (Renewed)
$9.000	$36	4 TB	3 months	Internal	SAS	Used	Dell 4TB 3.5" 7.2K NL SAS 12Gbs HDD (0F9W8) (Certified Refurbished)
$9.214	$129	14 TB		Internal 3.5"	HDD	Used	HGST WD Ultrastar DC HC530 14TB SATA 6Gb/s 3.5-Inch Data Center HDD - WUH721414ALE604 0F31152 (Renewed)
$9.333	$28	3 TB	1 year	Internal	SAS	Used	Seagate ST3000NM0023 Constellation Es.3 3tb Sas 6g
$9.333	$28	3 TB		Internal	SAS	Used	HGST Ultrastar HUS723030ALS640 3TB 7200RPM SAS 6Gbps 3.5" 64MB Cache Enterprise Internal Bare or OEM Hard Disk Drives
$9.990	$10	1 TB	3 months	Internal	SAS	Used	MDD MAXDIGITALDATA 1TB 7200RPM 128MB Cache SAS 6Gb/s 3.5inch Internal Enterprise Hard Drive (MD1TSAS12872E) - [NOT a SATA HDD] (Renewed)
$10.00	$80	8 TB	3 months	Internal	SAS	Used	MDD MAXDIGITALDATA 8TB 7200RPM 128MB Cache SAS 12.0Gb/s 3.5inch Internal Enterprise Hard Drive (MD8TSAS12872E) - [NOT a SATA HDD]
$10.00	$140	14 TB	3 months	Internal 3.5"	HDD	Used	MDD (MD14TSATA25672NAS) 14TB 7200 RPM 256MB Cache SATA 6.0Gb/s 3.5" Internal NAS Hard Drive - 5 Years Warranty (Renewed)
$10.00	$140	14 TB	3 years	Internal 3.5"	HDD	Used	MDD 14TB 7200RPM 256MB Cache SATA 6.0Gb/s 3.5inch Internal Hard Drive for Surveillance Storage (MD14TGSA25672DVR) - 3 Years Warranty (Renewed)
$10.00	$140	14 TB	3 years	Internal 3.5"	HDD	Used	MDD 14TB 7200RPM 256MB Cache SATA 6.0Gb/s 3.5inch Internal Hard Drive for Surveillance Storage (MD14TGSA25672DVR) - 3 Years Warranty (Renewed)
$10.36	$145	14 TB	5 years	Internal 3.5"	HDD	Used	MDD (MDD14TSATA25672E) 14TB 7200 RPM 256MB Cache SATA 6.0Gb/s 3.5" Internal Enterprise Hard Drive - 5 Years Warranty (Renewed)
$10.43	$146	14 TB	3 years	Internal 3.5"	HDD	Used	MDD (MDD14TSATA25672DVR) 14TB 7200RPM 256MB Cache SATA 6.0Gb/s 3.5inch Internal Surveillance Hard Drive - 3 Years Warranty (Renewed)
$10.61	$148	14 TB	5 years	Internal 3.5"	HDD	Used	MDD (MDD14TSATA25672E) 14TB 7200RPM 256MB Cache SATA 6.0Gb/s 3.5inch Enterprise Hard Drive - 5 Years Warranty (Renewed)
$10.64	$149	14 TB	3 years	Internal 3.5"	HDD	Used	Basicnology NAS HDD 14TB 7200RPM SATA 6Gb/s 128MB Cache 3.5inch Internal NAS Hard Drive (BG14TSA256N) - 3 Years Warranty (Renewed)
$10.64	$149	14 TB	3 years	Internal 3.5"	HDD	Used	Basicnology Surveillance HDD 14TB 7200RPM SATA 6Gb/s 128MB Cache 3.5inch Internal Surveillance Hard Drive (BG14TSA256S) - 3 Years Warranty (Renewed)

(https://diskprices.com/, Last updated at 2025-02-19)

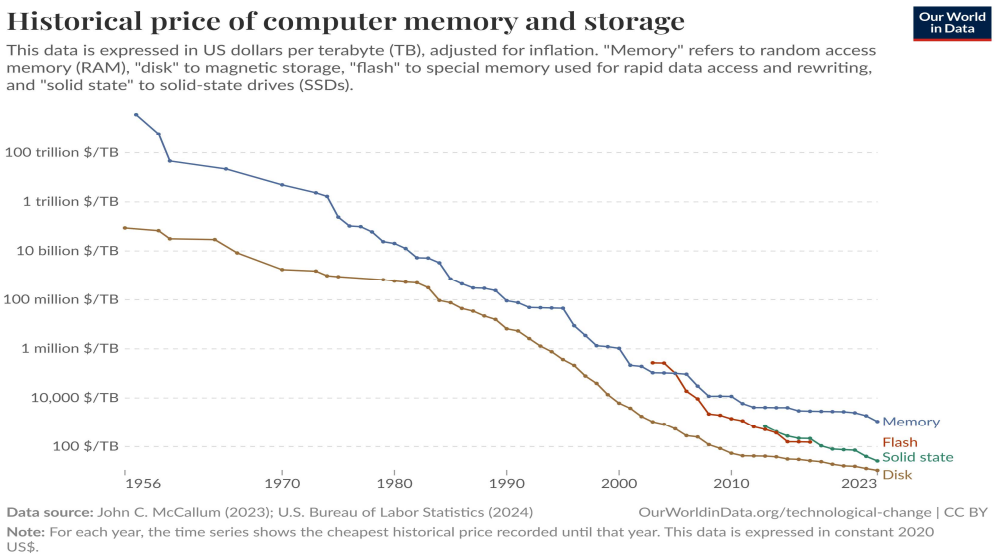
Disk Price 2025 (new)

Price per TB	Price	Capacity	Warranty	Form Factor	Technology	Condition	Affiliate Link
$12.50	$75	6 TB		Internal	SAS	New	Seagate Enterprise Capacity 3.5 HDD 6TB 7200RPM 12Gb/s SAS 128 MB Cache Internal Bare Drive ST6000NM0034
$13.75	$110	8 TB	2 years	Internal 3.5"	HDD	New	Seagate BarraCuda 8TB Internal Hard Drive HDD – 3.5 Inch Sata 6 Gb/s 5400 RPM 256MB Cache for Computer Desktop PC (ST8000DMZ04/004)
$13.75	$110	8 TB		Internal	SAS	New	Seagate Enterprise Capacity 3.5 | ST8000NM0075 | 8TB 7200 RPM SAS 12Gb/s 256MB Cache | 512e | Internal Hard Disk Drive
$14.00	$280	20 TB	3 years	External 3.5"	HDD	New	Seagate Expansion 20TB External Hard Drive HDD - USB 3.0, with Rescue Data Recovery Services (STKP20000400)
$14.64	$205	14 TB		External 3.5"	HDD	New	Seagate Expansion 14TB External Hard Drive HDD - USB 3.0, with Rescue Data Recovery Services (STKP14000400)
$14.85	$148	10 TB	5 years	Internal	SAS	New	Seagate Enterprise Capacity 3.5 HDD V.6 (Helium) Hard Drive 10 TB SAS 12Gb/S (ST10000NM0226)
$15.00	$60	4 TB		Internal 3.5"	HDD	New	HGST Deskstar HDS724040ALE640 0F14681 7K4000 4TB 7.2K RPM 64MB SATA Internal HDD
$15.00	$60	4 TB		Internal 3.5"	HDD	New	HGST Deskstar 3.5-Inch 4TB 7200 RPM SATA III 6Gbps 64MB Cache Internal Hard Drive (0F14681) (HDS724040ALE640)
$15.00	$120	8 TB	2 years	Internal 3.5"	HDD	New	Western Digital 8TB WD Blue PC Internal Hard Drive HDD - 5640 RPM, SATA 6 Gb/s, 256 MB Cache, 3.5" - WD80EAAZ
$15.39	$123	8 TB	2 years	Internal 3.5"	HDD	New	Seagate BarraCuda Internal Hard Drive 8TB SATA 6Gb/s 256MB Cache 3.5-Inch (ST8000DM004),Mechanical Hard Disk
$15.50	$248	16 TB	2 years	External 3.5"	HDD	New	Seagate Expansion Desktop 16TB, External Hard Drive, USB 3.0, 2 Year Rescue Services (STKP16000400)
$15.62	$250	16 TB	2 years	External 3.5"	HDD	New	WD 16TB Elements Desktop External Hard Drive, USB 3.0 for plug-and-play storage - WDBWLG0160HBK-NESN
$15.71	$220	14 TB	5 years	Internal 3.5"	HDD	New	Seagate IronWolf Pro ST14000NE0008 14 TB Hard Drive - 3.5" Internal - SATA (SATA/600)
$15.75	$63	4 TB		Internal	SAS	New	Seagate 4TB Enterprise Capacity SAS 12Gb s 512n 3.5" Internal Hard Drive Model ST4000NM0025
$15.83	$190	12 TB	5 years	Internal 3.5"	HDD	New	Toshiba MG Series Enterprise 12TB 3.5" SATA 6Gbit/s Internal HDD 7200RPM 550TB/year 24/7 Operation. MG07ACA12TE
$15.83	$190	12 TB	5 years	Internal	SAS	New	Seagate ST12000NM002G 12TB 12Gb/s SAS HDD
$15.87	$127	8 TB		Internal 3.5"	HDD	New	Seagate BarraCuda Pro 8TB Internal Hard Drive Performance HDD – 3.5 Inch SATA 6 Gb/s 5400 RPM 256MB Cache for Computer Desktop PC Laptop. (ST8000DM004)

(https://diskprices.com/, Last updated at 2025-02-19)

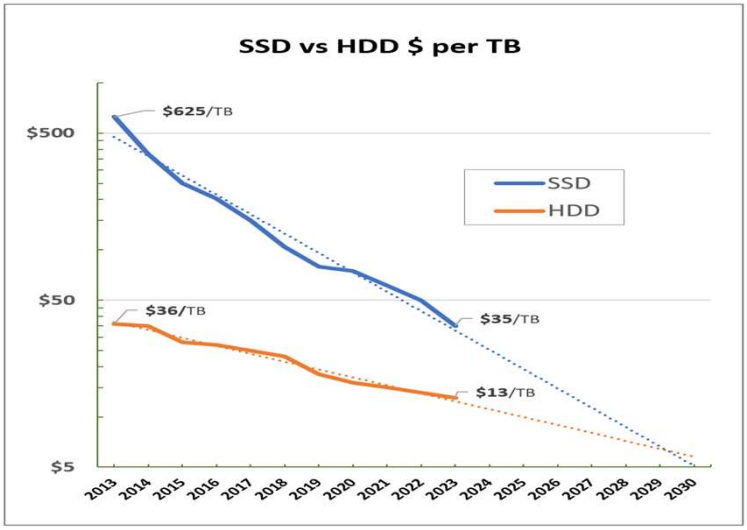
Course Information 7

Cost of Memory & Storage



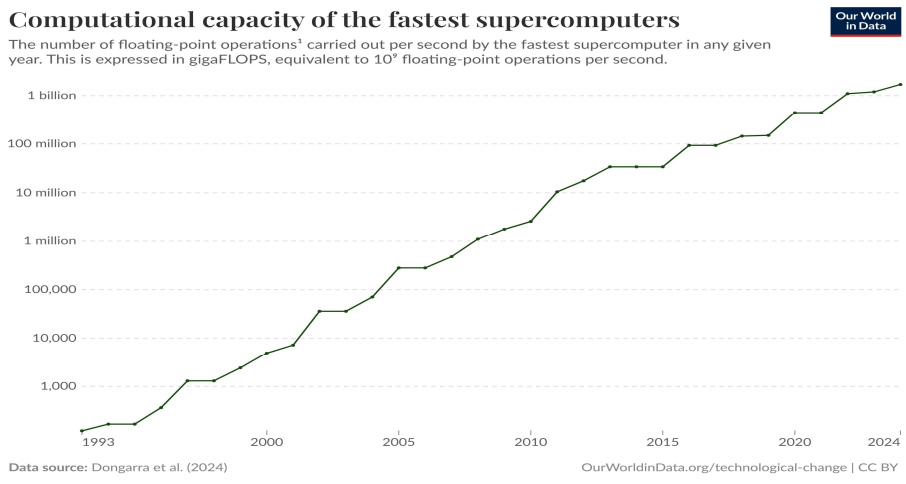
Introduction 8

SSD vs HDD



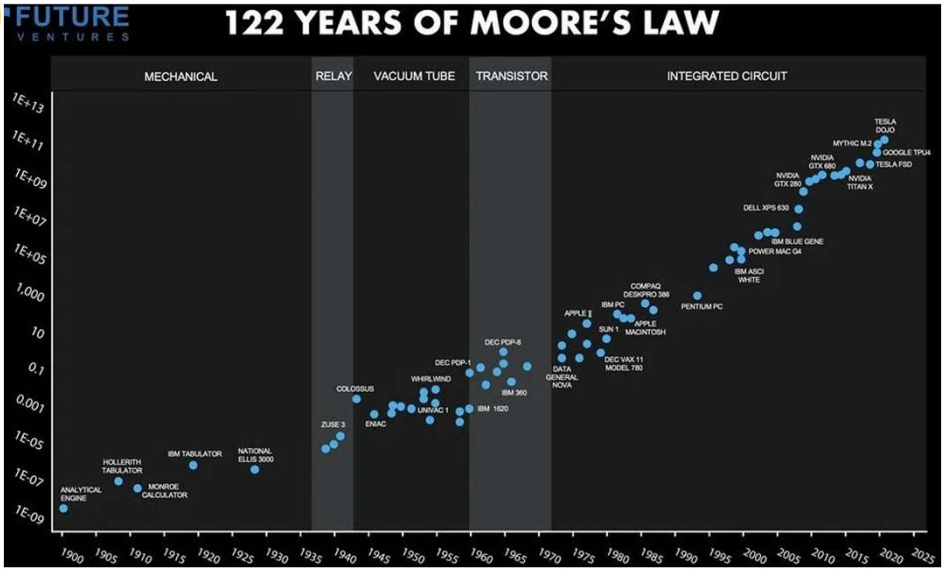
(https://www.stellarinfo.co.in/blog/the-future-of-hard-drives-exploring-the-innovations-in-storage-media/)

Computing Power

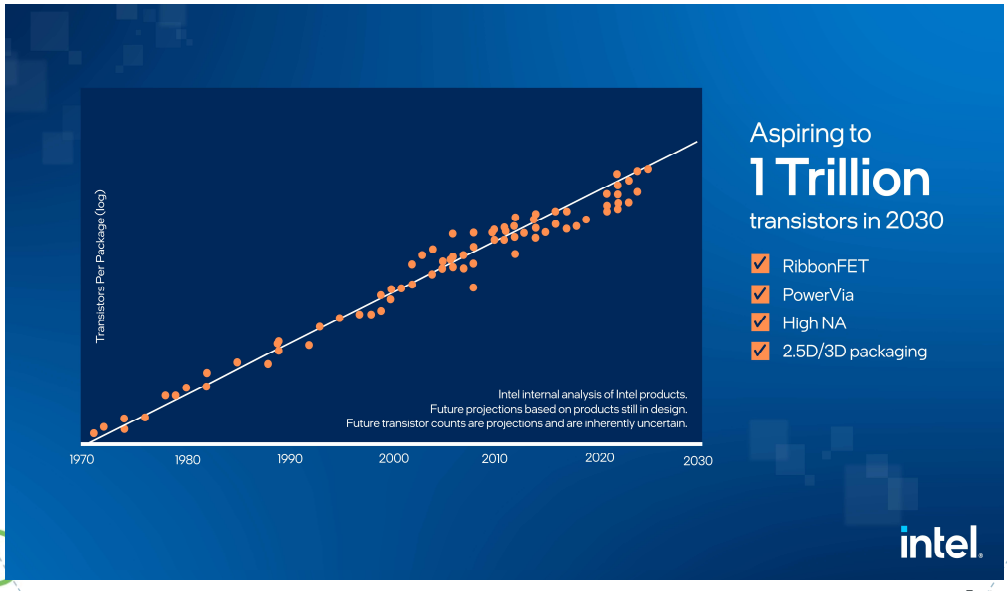


(https://ourworldindata.org/grapher/supercomputer-power-flops)

Moore's Law Continue?

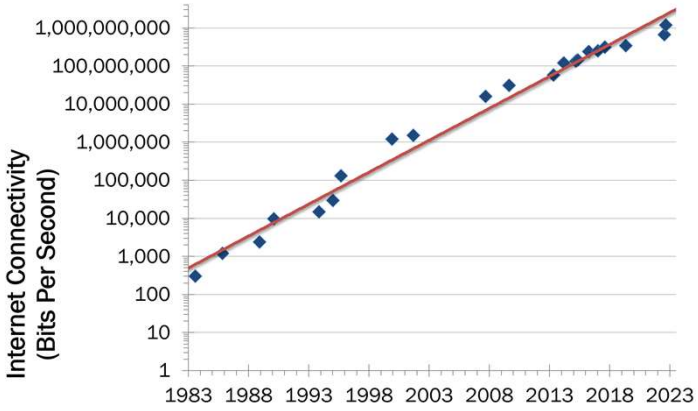


Moor's Law at Intel



Network Bandwidth Growth

Nielsen's law: A high-end user's connection speed grows by **50% per year**.



Introduction 13

Data Everywhere

- Airline flight management system
- Financial data
- Commercial store (eg, WalMart) data
- Department of Motor Vehicles
- Surveillance video
- University student records
- Baseball results
- Web sites
- Medical records
- ...

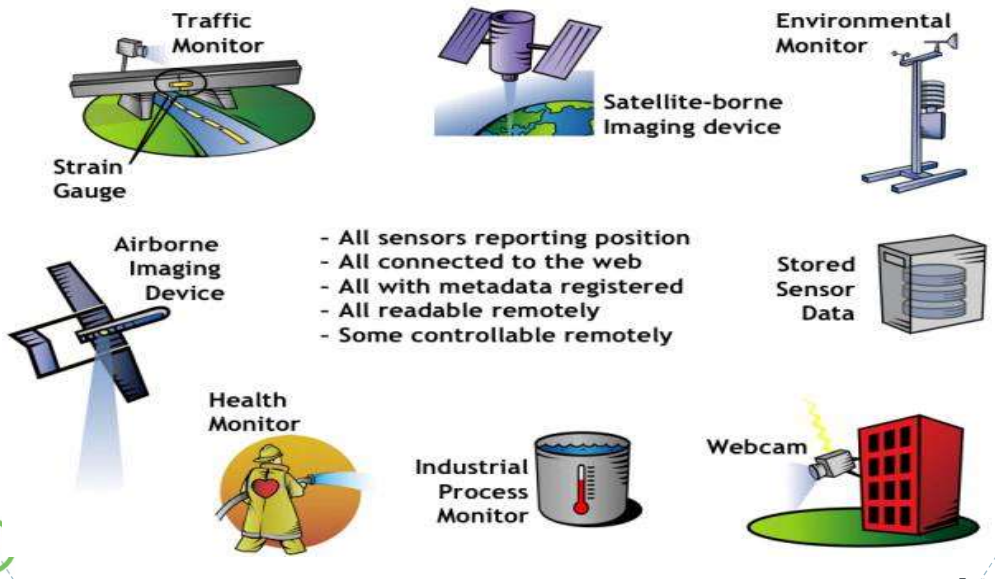
Introduction 14

New Types of Data

- Social Networks (FB, IG, Twitter, Linked-In, …) capturing a lot of information about people and relationships-posts, tweets, photos, videos, ...
- Search Engines - Google, Bing, Yahoo : collect their own repository of web pages.
- Cloud services – huge amount of data now resides on the cloud.
- Internet of Things (IoT) generate real-time sensing and streaming data.
- All of the above constitutes new types of data.

Introduction 15

Data from IoT Sensors



Introduction 16

Effective Management Required

- Effective management can make an organization's data a valuable asset(資產).
- Ineffective policies can make an organization's data a liability(負債).
- Big data analytics is becoming the gold mine of the 21st century.
- The paradigm has been extended from database systems to data science.

Introduction 17

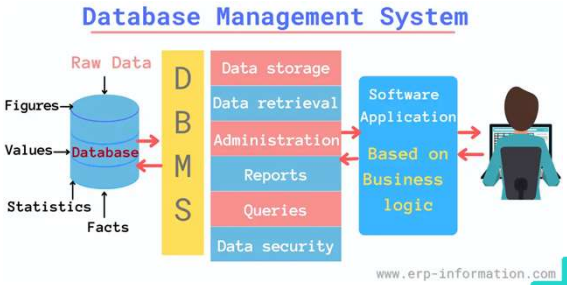
Basic Concepts

- Data: Known facts (recordable) with an implicit meaning.
- Database: Collection of interrelated data
- Mini-World or Universe of Discourse (UoD): Some part of the real world about which data is stored in a database.
- Database Management System (DBMS): A collection of programs to facilitate the creation and management of databases.
- Database System = DBMS + Database
- A database system contains information about a particular **enterprise** and provides an **environment** that is both **convenient** and **efficient** to use.

Introduction 18

Database Management System (DBMS)

- DBMS is:
  - A collection of software programs
  - General purpose
- DBMS enables users to:
  - Define DB
  - Construct DB
  - Change (or update) DB
  - Query the data in DB
  - Share DB
- DBMS maintains the integrity of DB



Introduction 19

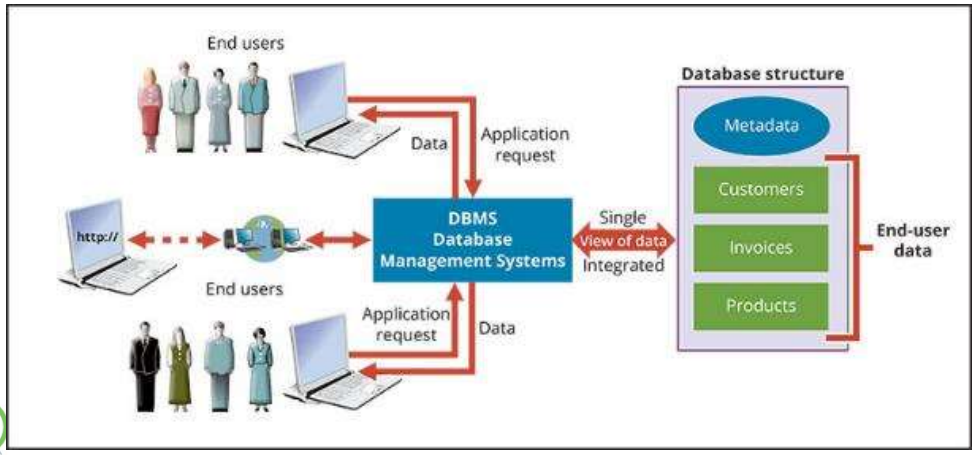
Role and Advantages of DBMS

- The DBMS presents the end user with a single, integrated view of the data in the database
- A DBMS provides the following advantages:
  - Improved data sharing
  - Improved data security
  - Better data integration
  - Minimized data inconsistency
  - Improved data access
  - Improved decision making
  - Increased end-user productivity

Course Information20

Role of DBMS

- The DBMS manages the interaction between the end users and the database.



Main Goals of DB Course

- To learn why, when, and where DBs are useful.
- To understand how to use a DBMS
  - How to design and create DB, data models, SQL, ...
- To study how a DBMS works
  - Properties of disks and files, software to manage reading/writing, algorithms to answer user queries, transaction management, ...
- To explore new concepts in data science/analytics
  - Big data, NoSQL/NewSQL/Distribugted SQL
  - Data science, data analytics
  - Streaming data management
  - AI & ML

DB Engine Ranking

424 systems in ranking, February 2025

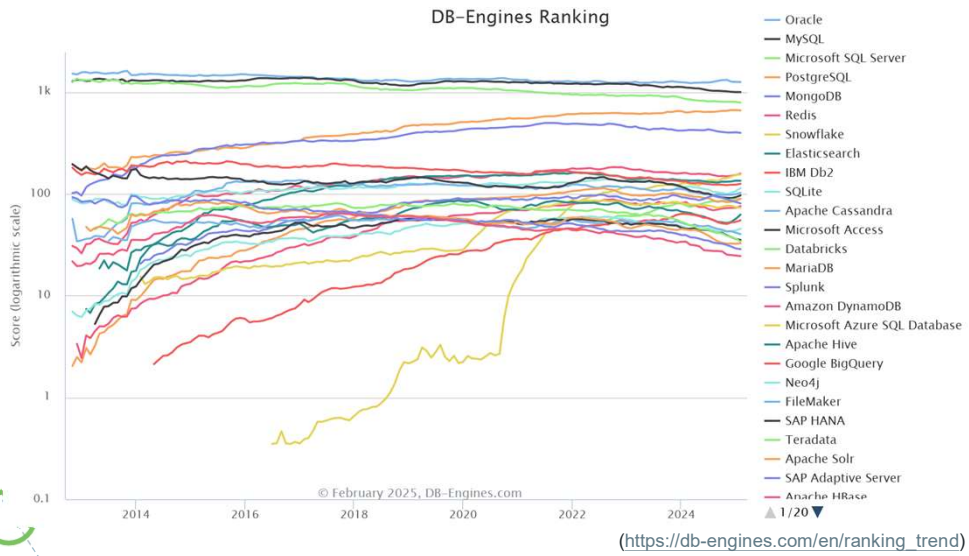
Rank Feb 2025	Rank Jan 2025	Rank Feb 2024	DBMS	Database Model	Score Feb 2025	Score Jan 2025	Score Feb 2024
1.	1.	1.	Oracle	Relational, Multi-model	1254.82	-3.93	+13.38
2.	2.	2.	MySQL	Relational, Multi-model	999.99	+1.84	-106.67
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	786.87	-11.69	-66.70
4.	4.	4.	PostgreSQL	Relational, Multi-model	659.62	-3.79	+30.21
5.	5.	5.	MongoDB	Document, Multi-model	396.63	-5.87	-23.73
6.	↑ 7.	6.	Redis	Key-value, Multi-model	157.91	+4.55	-2.80
7.	↓ 6.	↑ 9.	Snowflake	Relational	155.58	+1.68	+28.13
8.	8.	↓ 7.	Elasticsearch	Multi-model	134.63	-0.29	-1.11
9.	9.	↓ 8.	IBM Db2	Relational, Multi-model	125.44	+2.46	-6.79
10.	10.	10.	SQLite	Relational	113.82	+7.13	-3.47
11.	11.	↑ 12.	Apache Cassandra	Wide column, Multi-model	102.58	+3.39	-6.69
12.	12.	↓ 11.	Microsoft Access	Relational	96.54	+3.84	-16.63
13.	13.	↑ 17.	Databricks	Multi-model	90.03	+2.19	+13.13
14.	14.	↓ 13.	MariaDB	Relational, Multi-model	89.50	+3.92	-7.73
15.	15.	↓ 14.	Splunk	Search engine	80.56	-2.53	-11.09
16.	↑ 17.	↓ 15.	Amazon DynamoDB	Multi-model	75.58	+2.58	-7.31
17.	↓ 16.	↓ 16.	Microsoft Azure SQL Database	Relational, Multi-model	72.76	-1.02	-6.80
18.	18.	18.	Apache Hive	Relational	62.49	+5.61	-3.32
19.	19.	19.	Google BigQuery	Relational	54.89	+1.85	-8.74
20.	20.	↑ 22.	Neo4j	Graph	45.34	+1.65	-1.27

https://db-engines.com/en/ranking

Introduction 23

DB Engines Ranking Trend

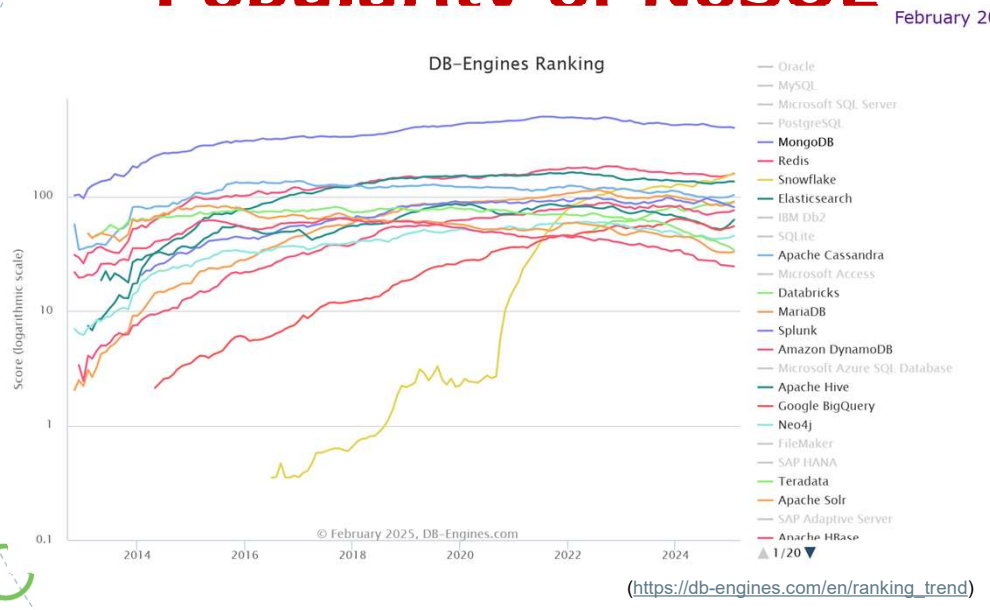
February 2025



(https://db-engines.com/en/ranking_trend)

Introduction 24

Popularity of NoSQL



(https://db-engines.com/en/ranking_trend)

Introduction 25

DataBases and Applications

- **Traditional** Applications: Numeric and Textual DBs
- More Recent Applications:
  - Multimedia Databases (images, audio, video, …)
  - Geographic Information Systems (GIS)
  - Data Warehouses
  - Real-time and Active Databases
  - Many other applications
- **New Trends**: cloud DB, big data, IoT, AI/ML
- First part: focuses on traditional applications
- *A number of recent applications are described later in the class and book.*

Introduction 26

Database System Environment

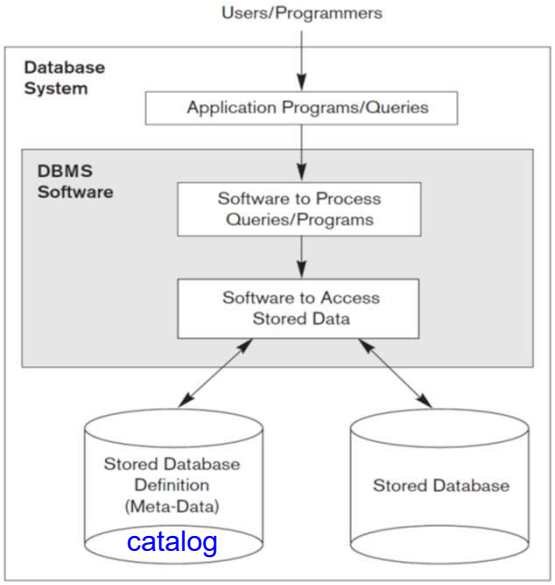
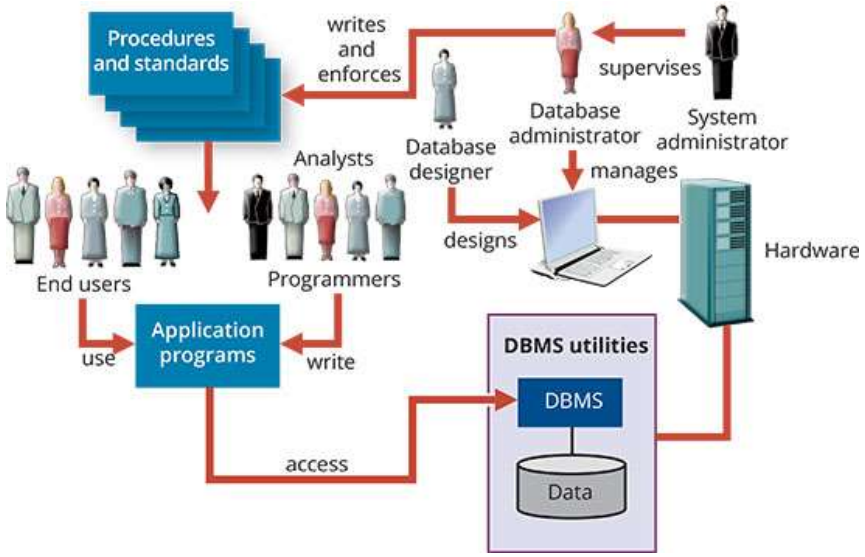
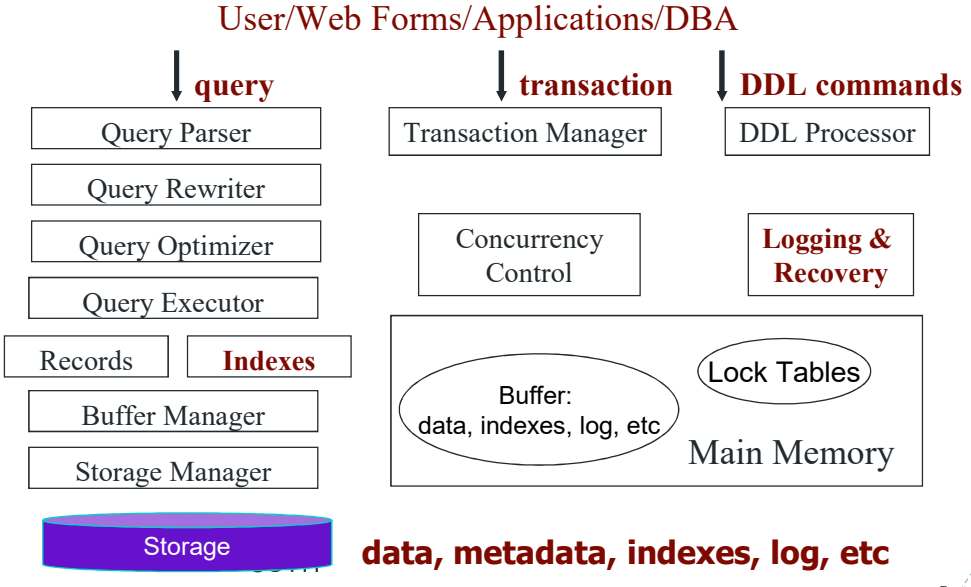


Figure 1.1
A simplified database system environment.

Database System and Users



DBMS Architecture



Typical DBMS Functionality

- Define a particular DB in terms of data **types**, **structures**, and **constraints**
- Construct or Load the initial DB **contents** on a secondary storage medium
- Manipulating the DB:
  - Retrieval: Querying, generating reports
  - Modification: Insertions, deletions and updates
  - Generally known as **CRUD**
  - Accessing the database with Web/cloud/mobile applications
- Processing and sharing by concurrent **users** and **applications** – yet, keeping all data **valid** and **consistent**

Typical DBMS Functionality

- Additional features:
  - Protection or Security measures to prevent unauthorized access
  - "Actively" take internal actions on data
  - Presentation and Visualization of data
  - Maintaining the database and associated programs over the lifetime of the applications
    - Called database, software, and system maintenance

Application Activities

- Applications interact with a DB by generating
  - Queries: that access different parts of data and formulate the result of a request
  - Transactions: that may read some data and "update" certain values or generate new data and store that in the database
- Applications must not allow unauthorized users to access data
- Applications must keep up with changing user requirements against the database

Database Applications

- **Banking**: all transactions
- **Airlines**: reservations, schedules, …
- **Universities**: registration, grades, …
- **Sales**: customers, products, purchases, …
- **Online retailers**: orders, customized recommendations
- **Manufacturing**: production, inventory, supply chain, …
- **Human resources**: employee records, salaries, tax, …
- **Internet & Web**: search, data management, …
- **Social media**: content management, relationships, …
- DBs are closely related to ALL aspects of our lives.

Introduction 33

Purpose of DB Systems

- In the early days, database applications were built directly on top of **file systems**
- **Drawbacks** of using file systems :
  - **Data redundancy** and **inconsistency**
    - Multiple file formats, duplication in different files
  - **Difficulty** in **accessing** data
    - Need to write a new program for each new task
  - **Data isolation** — multiple systems and locations
  - **Integrity** problems
    - Integrity constraints (e.g. account balance > 0) become "buried" in code rather than being stated explicitly
    - Hard to add new constraints or change existing ones

Introduction 34

Purpose of DB Systems

- Drawbacks of using file systems (cont.)
  - Atomicity of updates
    - Failures may leave DB in an inconsistent state
    - Example: Transfer of funds from one account to another
  - Concurrent access by multiple users
    - Concurrent accessed needed for performance
    - Uncontrolled accesses can lead to inconsistencies
      - Example: reading and updating at the same time
  - Security problems
    - Hard to provide user access to some, but not all, data
- DB systems offer solutions to **ALL** the problems above.

Introduction 35

Types of Databases

- A **single-user database** (one user at a time)
  - A **desktop database** is single-user database on a PC
- A **multiuser database** (multiple users at the same time)
  - A **workgroup database** supports a small number of users or a specific department
  - An **enterprise database** supports many users across many departments
- Classification by **location**
  - A centralized database (data located at a single site)
  - A distributed database (data distributed across different sites)
  - A cloud database is created and maintained using cloud data services

Course Information 36

Types of Databases

- Classification by data type
  - General-purpose databases contain a wide variety of data used in multiple disciplines
  - Discipline-specific databases contain data focused on specific subject areas
  - An operational database is designed to support a company's day-to-day operations
  - An analytical database stores historical data and business metrics used exclusively for tactical or strategic decision making and is comprised of two main components:
    - The **data warehouse** stores data in a format optimized for decision support
    - **Online analytical processing (OLAP)** is a set of tools for retrieving, processing, and modeling data from the data warehouse

Introduction 37

Types of Databases

- **Business intelligence** describes a comprehensive approach to capture and process business data to generate information that support **decision making**
- Databases can be classified to reflect the degree to which the data is structured
  - Unstructured data exists in its original (raw) state
  - Structured data is the result of formatting unstructured data to facilitate storage and use
  - Semistructured data has already been processed to some extent
- **Extensible Markup Language (XML)** is a language used to represent data elements in textual format
  - An XML database supports the storage and management of unstructured XML data

Course Information38

Types of Databases

- **Social media** refers to web and mobile technologies that enable "anywhere, anytime, always on" human interaction
  - Data is captured about end users and consumers that requires the use of specialized database systems
- The term **NoSQL** (Not only SQL) is a new generation of DBMS that is not based on the traditional relational database model
  - These databases are designed to handle an unprecedented volume of data, variety of data types and structures, and velocity of data operations of new business requirements

Introduction 39

DB Example

- Mini-world for the example:
  - UNIVERSITY environment.
- Some mini-world *entities*:
  - STUDENTs
  - COURSEs
  - SECTIONs (of COURSEs)
  - DEPARTMENTs
  - INSTRUCTORs
  - ...

Introduction 40

DB Example

- Some mini-world *relationships*:
  - SECTIONs *are of specific* COURSEs
  - STUDENTs *take* SECTIONs
  - COURSEs *have prerequisite* COURSEs
  - INSTRUCTORs *teach* SECTIONs
  - COURSEs *are offered by* DEPARTMENTs
  - STUDENTs *major in* DEPARTMENTs
- Note: Entities and relationships are typically expressed in a conceptual data model, such as the Entity-Relationship model (to be discussed later in class)

DB Example

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Figure 1.2
A database that stores student and course information.

Main Characteristics

- Self-describing: A DBMS catalog (meta-data) stores the description of the database. (next slide)
- Program-data Independence: Allows changing storage structures w/o changing DBMS access programs.
- Data abstraction: Data models hide storage details and present the users with a conceptual view of the DB.
- Multiple views: Each user may see a different view of the DB.
- Data sharing: among multiple users
- Transactions, concurrent access , recovery , OLTP

A Simplified DB Catalog

RELATIONS

Relation_name	No_of_columns
STUDENT	4
COURSE	4
SECTION	5
GRADE_REPORT	3
PREREQUISITE	2

COLUMNS

Column_name	Data_type	Belongs_to_relation
Name	Character (30)	STUDENT
Student_number	Character (4)	STUDENT
Class	Integer (1)	STUDENT
Major	Major_type	STUDENT
Course_name	Character (10)	COURSE
Course_number	XXXXNNNN	COURSE
....	....	.....
....	....	.....
....	....	.....
Prerequisite_number	XXXXNNNN	PREREQUISITE

Note: Major_type is defined as an enumerated type with all known majors.
XXXXNNNN is used to define a type with four alphabetic characters followed by four numeric digits.

Figure 1.3
An example of a database catalog for the database in Figure 1.2.

DB Users

- Users may be divided into
  - Those who use and control the DB content, and those who design, develop and maintain DB applications (called "actors on the scene" 幕前使用者), and
  - Those who design and develop the DBMS software and related tools, and the computer systems operators (called "workers behind the scene"幕後工作者).

DB Users

- Actors on the scene
  - DB administrators:
    - Authorizing access to the DB, coordinating and monitoring its use, acquiring software and hardware resources, controlling its use and monitoring efficiency of operations.
  - DB designers:
    - Define the content, the structure, the constraints, and functions or transactions against the DB. They must communicate with the end-users and understand their needs.

DB Administrator (DBA)

- Coordinates all the activities of the DB system; must have a good understanding of the enterprise's information resources and needs.
- DBA's duties:
  - Schema definition
  - Storage structure and access method definition
  - Schema and physical organization modification
  - Granting user authority to access the database
  - Specifying integrity constraints
  - Acting as liaison with users
  - Monitoring performance and responding to changes

Introduction 47

End-users

- Actors on the scene (continued)
  - End-users: use the data for queries, reports and some of them update the content.
- End-users can be categorized into:
  - Casual: access DB occasionally when needed
  - Naïve or Parametric: they make up a large section of the end-user population.
    - They use previously well-defined functions in the form of "canned transactions" against the DB.
    - Examples are bank-tellers or reservation clerks who do this activity for an entire shift of operations.

Introduction 48

End-users (cont.)

- Sophisticated:
  - Business analysts, scientists, engineers, others thoroughly familiar with the system capabilities.
  - Many use tools in the form of software packages that work closely with the stored DB.
- Stand-alone:
  - Mostly maintain personal DBs using ready-to-use packaged applications.
  - An example is a tax program user that creates its own internal DB.
  - Another example is a user that maintains an address book

System Analysts and Application Programmers

- System analysts: Analyze problem, determine the requirements of the users, develop specifications.
- Application programmers: Design and implement specification, testing, debugging, maintaining software. Also known as software developers or software engineers.
- Business analysts: People who can analyze vast amounts of business data and real-time data for better decision making, planning, advertising, marketing etc.

Users behind the Scene

- DB designers – design the DB systems for end users
- DBMS designers – design DBMS and tools for building DBs
- Tool designers – Design and implement tools that facilitate building of applications and allow using DB effectively (eg. modeling and designing DBs, performance monitoring, prototyping, test data generation, user interface creation, simulation etc.)
- Operators and maintenance personnel.

Advantages of Using DBs

- Controlling redundancy in data storage and in development and maintenance efforts.
  - Sharing of data among multiple users.
- Restricting unauthorized access to data.
- Providing persistent storage for program objects
  - In object-oriented DBMS
- Providing storage structures (e.g. indexes) for efficient data access
- Providing optimization of queries for efficient processing

Advantages of Using DBs(cont.)

- Providing backup and recovery services.
- Providing multiple interfaces to different users.
- Representing complex relationships among data.
- Enforcing integrity constraints on the DB.
- Drawing inferences and actions from the stored data using deductive and active rules
- Finding actionable insights from large data sets
- …

Additional Implications

- Potential for enforcing standards:
  - This is very crucial for the success of DB applications in large organizations.
  - **Standards** refer to data item names, display formats, screens, report structures, meta-data (description of data), Web page layouts, etc.
- Reduced application development time:
  - Incremental time to add each new application is reduced.

Additional Implications (cont.)

- Flexibility to change data structures:
  - DB structure may evolve as new requirements are defined.
- Availability of current information:
  - Extremely important for on-line transaction systems such as airline, hotel, car reservations.
- Economies of scale:
  - Wasteful overlap of resources and personnel can be avoided by consolidating data and applications across departments.

History of DB Systems

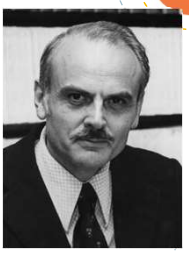
- **Pre-1960s**
  - File processing systems
  - Redundancy and inconsistency between files
  - Incompatibility between access programs
  - Data isolation
  - Concurrent access anomalies
  - Security and integrity problems
  - Hard to share

DB History (cont.)

- **The '60s**
  - Charles Bachman designed the 1st DBMS Integrated Data Store (the 1st Turing Award winner w/o a PhD, in 1973)
  - Three-level architecture (more about this in next lecture)
  - CODASYL, DBTG, and the network model
  - Hierarchical model and the IMS system

DB History (cont.)

- **The '70s**
  - Edgar F. Codd (1970): The Relational model (Codd won the 1981 Turing Award)
  - Provide a sound theoretical base.
  - 1975, 1st ACM SIGMOD international conference
  - 1975, 1st VLDB international conference
  - Peter Chen 陳品山 (1976): The Entity-relationship model
  - System R (IBM), INGRES (UC-Berkely), System 2000 (UT-Austin)
  - SQL, QUEL



DB History (cont.)

The '80s

- Commercial relational DBMS (DB2, ORACLE, SYBASE, INFORMIX, ...)
- DBMS on PC's (DBASE, PARADOX, ...)
- Transaction management (James Gray won the 1999 Turing Award)
- Standards (SQL standardized in the late 1980s)

DB History (cont.)

The '90s

- New applications (Web, CAD/CAM, CASE, office automation, science and engineering, VLSI, ...)
- Demand for new DBMS technologies
- Object-oriented DBs, Parallel/Distributed DBs, Active/Deductive DBs, Multimedia DBs, Mobile DBs, Temporal/Real-time DBs, Spatial DBs(such as GIS), ...
- The emergence of ERP (Enterprise Resource Planning) and MRP (Material Requirements Planning) packages
- Data Warehousing and data mining
- DBMS in the Internet/Web and E-commerce applications

DB History (cont.)

The 2000s and beyond

- XML, XQuery and the Semantic Web
- Data Stream Management Systems (DSMS)
  - Sensor databases, IoT data management
  - Network traffic analysis
  - RFID data management
  - …
- Mobile Data Management (MDM)
- Cloud/Fog/Edge Databases
- AIDB, MLDB (DB + AI&ML)

Extending DB Capabilities

New functionalities are added to DBMSs in new areas:

- Scientific Applications –Physics, Chemistry, Biology -Genetics
- Earth and Atmospheric Sciences and Astronomy
- XML (eXtensible Markup Language), JSON
- Image Storage and Management
- Audio and Video Data Management
- Data Warehousing and Data Mining –a very major area for future development using new technologies
- Spatial Data Management and Location Based Services
- Time Series and Historical Data Management

The above gives rise to new research and development in DB systems.

New Trends

- First decade of the 21st century has seen tremendous growth in user generated data and automatically collected data from APPs, search engines, sensors...
- Social Media platforms such as Facebook, IG and Twitter are generating millions of transactions a day and businesses are interested to tap into this data to "understand" the users.
- Cloud Storage and Backup is making unlimited amount of storage available to users and applications
- IoT provides real-time sensing and streaming data
- AI&ML need to process huge amount of data

Big Data, NoSQL, NewSQL, Distributed SQL

- New data storage, management and analysis technologies were necessary for huge volume of data in PB a day (10^{15} bytes or 1000 TBs) – "Big Data".
- Hadoop and MapReduce approach to distributed data as well as the Google File System have given rise to Big Data technologies. Further enhancements by Spark.
- NoSQL/NewSQL/Distributed SQL systems for rapid search and retrieval, processing huge graphs, and other forms of data with flexible models of transaction processing across the globe.
- **True value of data** can now be revealed !!

When NOT to Use a DBMS

- **Disadvantages** of database systems:
  - Increased costs
  - Management complexity
  - Maintaining currency
  - Vendor dependence
  - Frequent upgrade/replacement cycles
  - Data loss/corruption
  - Security threat
  - Cost of staff/user training

When NOT to Use a DBMS

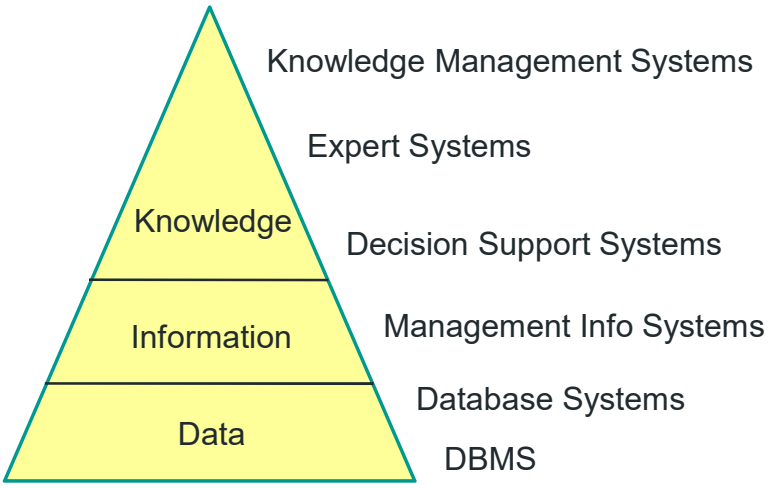
- Main inhibitors (costs) of using a DBMS:
  - High initial investment and possible need for additional hardware. (Better with cloud)
  - The generality that a DBMS provides for defining and processing data
  - Overhead for providing security, concurrency control, recovery, and integrity functions.
- When a DBMS may be unnecessary:
  - If the database and applications are simple, well defined, and not expected to change.
  - If access to data by multiple users is not required.

When NOT to use a DBMS

- When a DBMS may be infeasible:
  - E.g.: In embedded systems where a general purpose DBMS may not fit in available storage
- When no DBMS may suffice:
  - If there are hard real-time requirements that may not be met because of DBMS overhead
  - If the database system is not able to handle the complexity of data because of modeling limitations
  - If the database users need special operations not supported by the DBMS

Introduction 67

Related Systems



Introduction 68

Summary

- Types of databases and database applications
- Basic concepts and definitions
- Typical DBMS functionality
- Example of a database (UNIVERSITY)
- Main characteristics of the database approach
- Database users
- Advantages of using the database approach
- Database history and new trends
- When NOT to use databases
- Related systems