Introduction to Database Systems

CSci 150 - Fundamentals of Database Systems

Rodney Maniego Jr.
Instructor I ●●● DCST
College of Engineering and Technology
Visayas State University - Main Campus

Version: 20230814

Course Outcome

Describe data and database concepts.

Learning Outcomes

- Identify and explain the concepts about data and databases.
- Identify and explain the users, characteristics, and applications of databases.

Outline:

- File Systems
- Database Systems
- Database Management Systems
  - DBMS Characteristics
  - DBMS ACID Properties
  - Evolution of DBMS (pre-1960s to present)
  - DBMS Types
- Database Applications

Database Designer

Responsible for designing the structure and layout of a database based on the requirements.

Database Administrator

Responsible for the management, maintenance, and performance of an organization's databases.

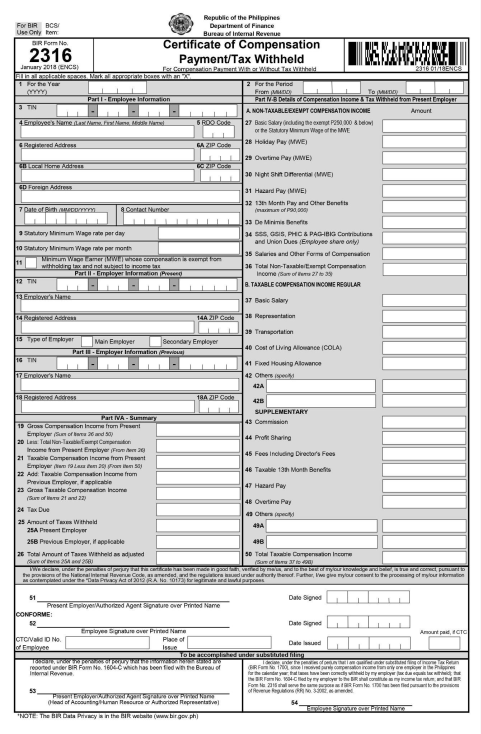
File Systems

What is Data?

It refers to raw, unprocessed facts, figures, symbols, or its representations, it may convey no meaning by its own.

When one or more data is transformed, interpreted, and contextualized, it becomes **information**.

#1 / File Systems



For BIR Use Only BCS/ Item:

Republic of the Philippines
Department of Finance
Bureau of Internal Revenue

BIR Form No. **2316** January 2018 (ENCS)

Certificate of Compensation Payment/Tax Withheld
For Compensation Payment With or Without Tax Withheld

2316 01/18ENCS

Fill in all applicable spaces. Mark all appropriate boxes with an "X".

1 For the Year (YYYY)
2 For the Period From (MM/DD) To (MM/DD)

Part I - Employee Information
3 TIN
4 Employee's Name (Last Name, First Name, Middle Name)
5 RDO Code
6 Registered Address
6A ZIP Code
6B Local Home Address
6C ZIP Code
6D Foreign Address
7 Date of Birth (MM/DD/YYYY)
8 Contact Number
9 Statutory Minimum Wage rate per day
10 Statutory Minimum Wage rate per month
11 Minimum Wage Earner (MWE) whose compensation is exempt from withholding tax and not subject to income tax

Part II - Employer Information (Present)
12 TIN
13 Employer's Name
14 Registered Address
14A ZIP Code
15 Type of Employer Main Employer Secondary Employer

Part III - Employer Information (Previous)
16 TIN
17 Employer's Name
18 Registered Address
18A ZIP Code

Part IVA - Summary
19 Gross Compensation Income from Present Employer (Sum of Items 36 and 50)
20 Less: Total Non-Taxable/Exempt Compensation Income from Present Employer (From Item 36)
21 Taxable Compensation Income from Present Employer (Item 19 Less Item 20) (From Item 50)
22 Add: Taxable Compensation Income from Previous Employer, if applicable
23 Gross Taxable Compensation Income (Sum of Items 21 and 22)
24 Tax Due
25 Amount of Taxes Withheld
25A Present Employer
25B Previous Employer, if applicable
26 Total Amount of Taxes Withheld as adjusted (Sum of Items 25A and 25B)

Part IV-B Details of Compensation Income & Tax Withheld from Present Employer

A. NON-TAXABLE/EXEMPT COMPENSATION INCOME — Amount
27 Basic Salary (including the exempt P250,000 & below) or the Statutory Minimum Wage of the MWE
28 Holiday Pay (MWE)
29 Overtime Pay (MWE)
30 Night Shift Differential (MWE)
31 Hazard Pay (MWE)
32 13th Month Pay and Other Benefits (maximum of P90,000)
33 De Minimis Benefits
34 SSS, GSIS, PHIC & PAG-IBIG Contributions and Union Dues (Employee share only)
35 Salaries and Other Forms of Compensation
36 Total Non-Taxable/Exempt Compensation Income (Sum of Items 27 to 35)

B. TAXABLE COMPENSATION INCOME REGULAR
37 Basic Salary
38 Representation
39 Transportation
40 Cost of Living Allowance (COLA)
41 Fixed Housing Allowance
42 Others (specify)
42A
42B

SUPPLEMENTARY
43 Commission
44 Profit Sharing
45 Fees Including Director's Fees
46 Taxable 13th Month Benefits
47 Hazard Pay
48 Overtime Pay
49 Others (specify)
49A
49B
50 Total Taxable Compensation Income (Sum of Items 37 to 49B)

I/We declare, under the penalties of perjury that this certificate has been made in good faith, verified by me/us, and to the best of my/our knowledge and belief, is true and correct, pursuant to the provisions of the National Internal Revenue Code, as amended, and the regulations issued under authority thereof. Further, I/we give my/our consent to the processing of my/our information as contemplated under the *Data Privacy Act of 2012 (R.A. No. 10173) for legitimate and lawful purposes.

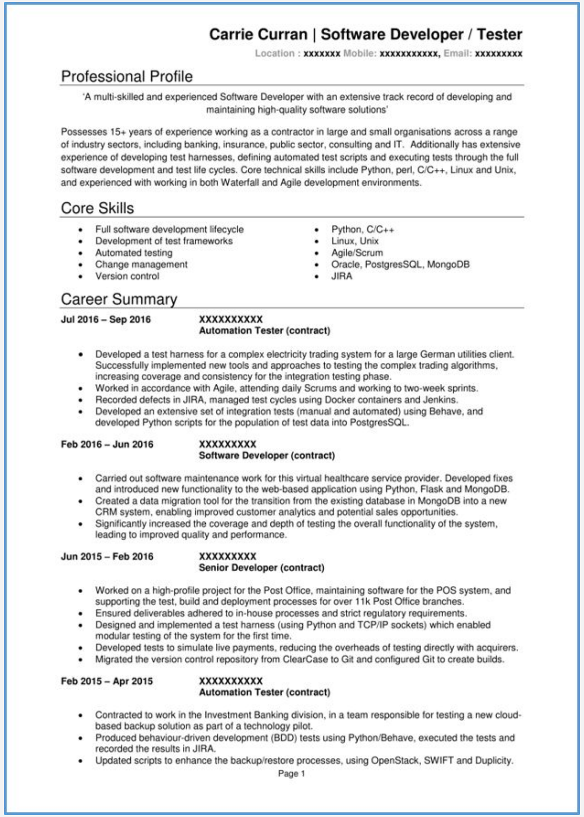
51 Present Employer/Authorized Agent Signature over Printed Name — Date Signed
CONFORME:
52 Employee Signature over Printed Name — Date Signed
CTC/Valid ID No. of Employee — Place of Issue — Date Issued — Amount paid, if CTC

To be accomplished under substituted filing

I declare, under the penalties of perjury that the information herein stated are reported under BIR Form No. 1604-C which has been filed with the Bureau of Internal Revenue.
53 Present Employer/Authorized Agent Signature over Printed Name (Head of Accounting/Human Resource or Authorized Representative)

I declare, under the penalties of perjury that I am qualified under substituted filing of Income Tax Return (BIR Form No. 1700), since I received purely compensation income from only one employer in the Philippines for the calendar year; that taxes have been correctly withheld by my employer (tax due equals tax withheld); that the BIR Form No. 1604-C filed by my employer to the BIR shall constitute as my income tax return; and that BIR Form No. 2316 shall serve the same purpose as if BIR Form No. 1700 has been filed pursuant to the provisions of Revenue Regulations (RR) No. 3-2002, as amended.
54 Employee Signature over Printed Name

*NOTE: The BIR Data Privacy is in the BIR website (www.bir.gov.ph)



Carrie Curran | Software Developer / Tester
Location : xxxxxxx Mobile: xxxxxxxxxxx, Email: xxxxxxxxx

Professional Profile

'A multi-skilled and experienced Software Developer with an extensive track record of developing and maintaining high-quality software solutions'

Possesses 15+ years of experience working as a contractor in large and small organisations across a range of industry sectors, including banking, insurance, public sector, consulting and IT. Additionally has extensive experience of developing test harnesses, defining automated test scripts and executing tests through the full software development and test life cycles. Core technical skills include Python, perl, C/C++, Linux and Unix, and experienced with working in both Waterfall and Agile development environments.

Core Skills

- Full software development lifecycle
- Development of test frameworks
- Automated testing
- Change management
- Version control
- Python, C/C++
- Linux, Unix
- Agile/Scrum
- Oracle, PostgresSQL, MongoDB
- JIRA

Career Summary

Jul 2016 – Sep 2016 XXXXXXXXXX
Automation Tester (contract)

- Developed a test harness for a complex electricity trading system for a large German utilities client. Successfully implemented new tools and approaches to testing the complex trading algorithms, increasing coverage and consistency for the integration testing phase.
- Worked in accordance with Agile, attending daily Scrums and working to two-week sprints.
- Recorded defects in JIRA, managed test cycles using Docker containers and Jenkins.
- Developed an extensive set of integration tests (manual and automated) using Behave, and developed Python scripts for the population of test data into PostgresSQL.

Feb 2016 – Jun 2016 XXXXXXXX
Software Developer (contract)

- Carried out software maintenance work for this virtual healthcare service provider. Developed fixes and introduced new functionality to the web-based application using Python, Flask and MongoDB.
- Created a data migration tool for the transition from the existing database in MongoDB into a new CRM system, enabling improved customer analytics and potential sales opportunities.
- Significantly increased the coverage and depth of testing the overall functionality of the system, leading to improved quality and performance.

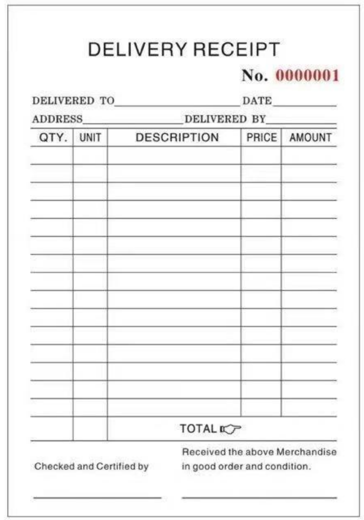
Jun 2015 – Feb 2016 XXXXXXXXX
Senior Developer (contract)

- Worked on a high-profile project for the Post Office, maintaining software for the POS system, and supporting the test, build and deployment processes for over 11k Post Office branches.
- Ensured deliverables adhered to in-house processes and strict regulatory requirements.
- Designed and implemented a test harness (using Python and TCP/IP sockets) which enabled modular testing of the system for the first time.
- Developed tests to simulate live payments, reducing the overheads of testing directly with acquirers.
- Migrated the version control repository from ClearCase to Git and configured Git to create builds.

Feb 2015 – Apr 2015 XXXXXXXXXX
Automation Tester (contract)

- Contracted to work in the Investment Banking division, in a team responsible for testing a new cloud-based backup solution as part of a technology pilot.
- Produced behaviour-driven development (BDD) tests using Python/Behave, executed the tests and recorded the results in JIRA.
- Updated scripts to enhance the backup/restore processes, using OpenStack, SWIFT and Duplicity.

Page 1



DELIVERY RECEIPT

No. 0000001

DELIVERED TO______ DATE______
ADDRESS______ DELIVERED BY______

QTY.	UNIT	DESCRIPTION	PRICE	AMOUNT
		TOTAL ☞		

Checked and Certified by

Received the above Merchandise in good order and condition.

What is a File System?

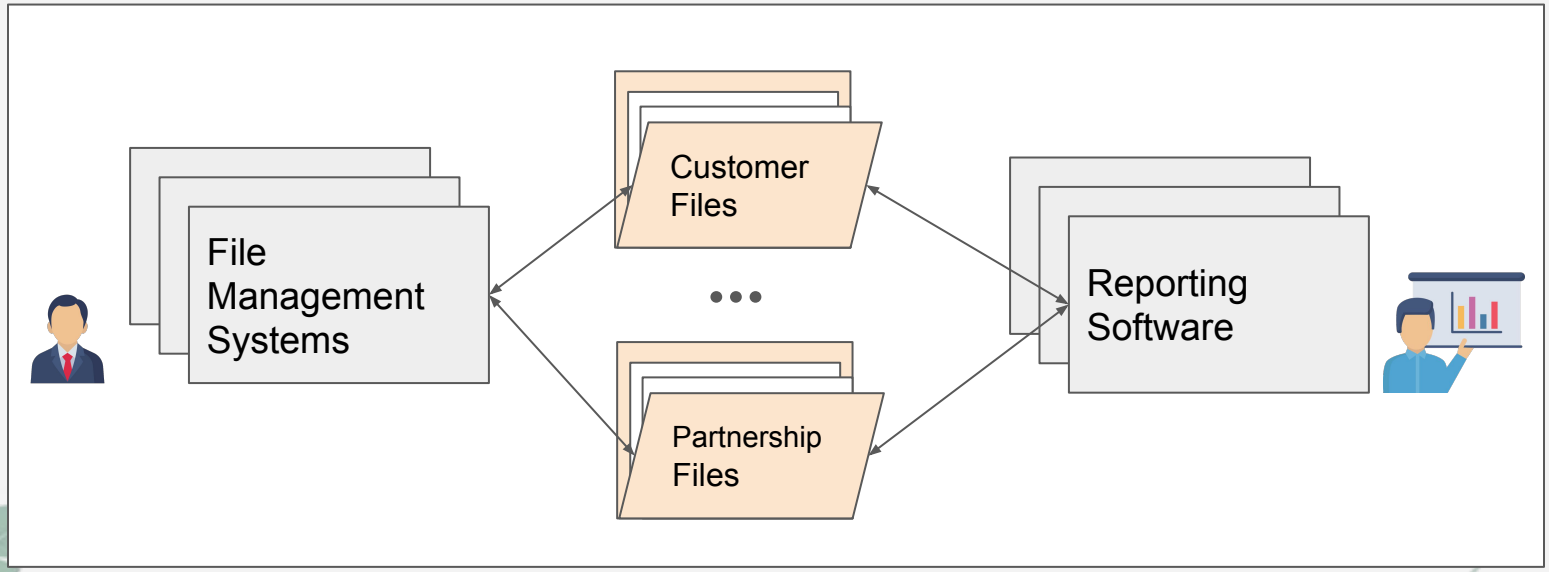
It is a systematic method of categorizing, arranging, and storing of items for easy organization and retrieval.

It is now more closely associated to computerized systems that refers to the software and structure that manages how digital data is organized, stored, and accessed.

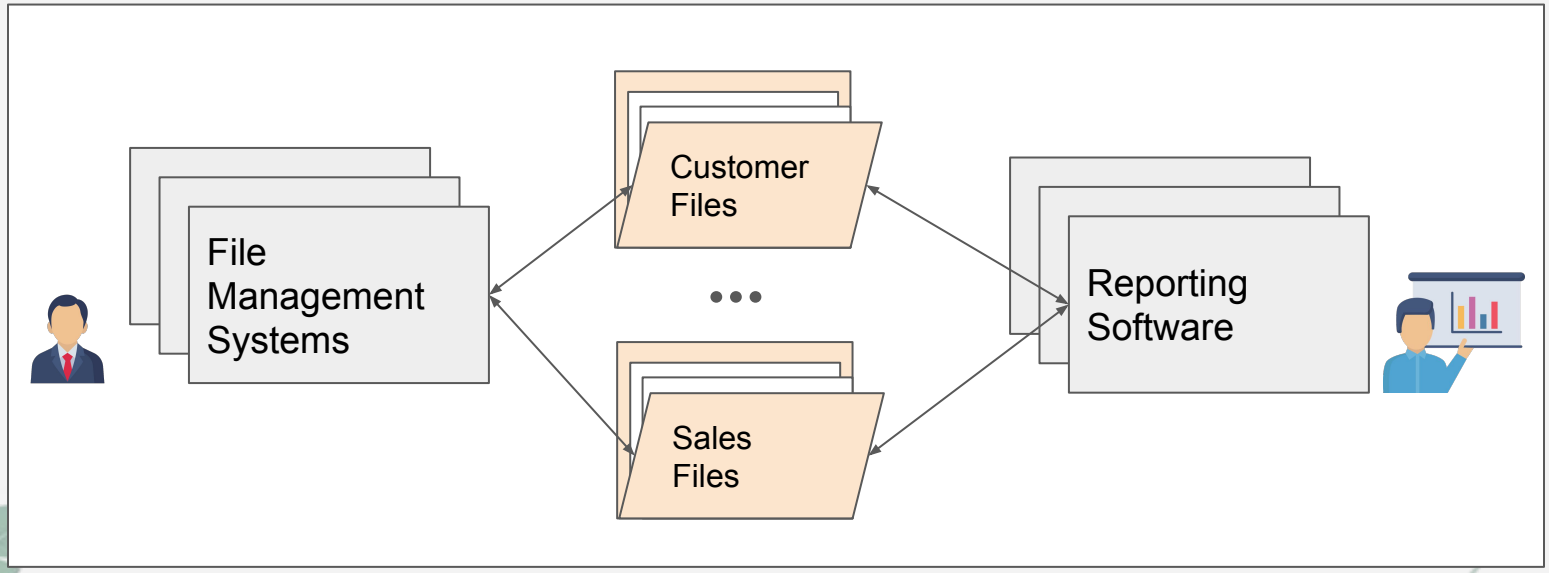
Different File Systems

- File Allocation Table 32 (FAT32)
- New Technology File System (NTFS)
- Extended File Allocation Table (exFAT)
- Universal Disk Format (UDF)
- Apple File System (APFS)
- B-tree File System (btrfs)

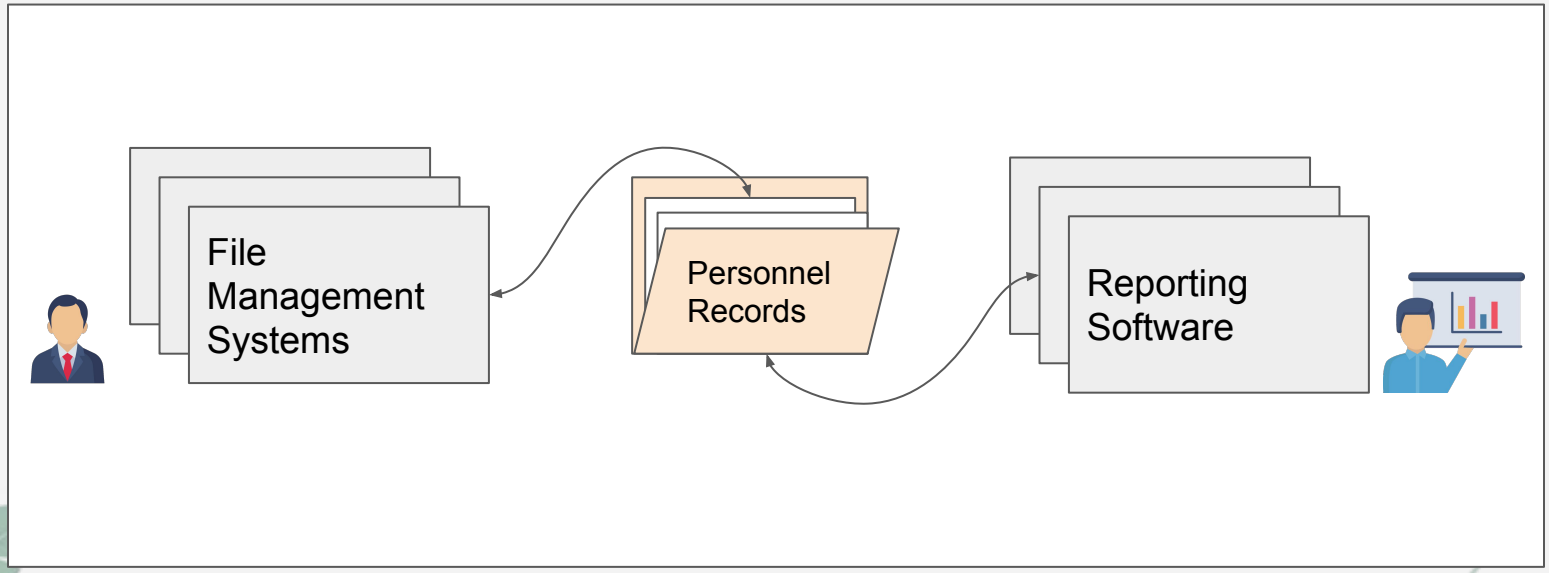
Marketing Department



Sales Department



Personnel Department



Disadvantages

- Extensive programming
- Complex development process
- Complex system administration
- Slow information retrieval
- Limited data sharing
- Lack of security

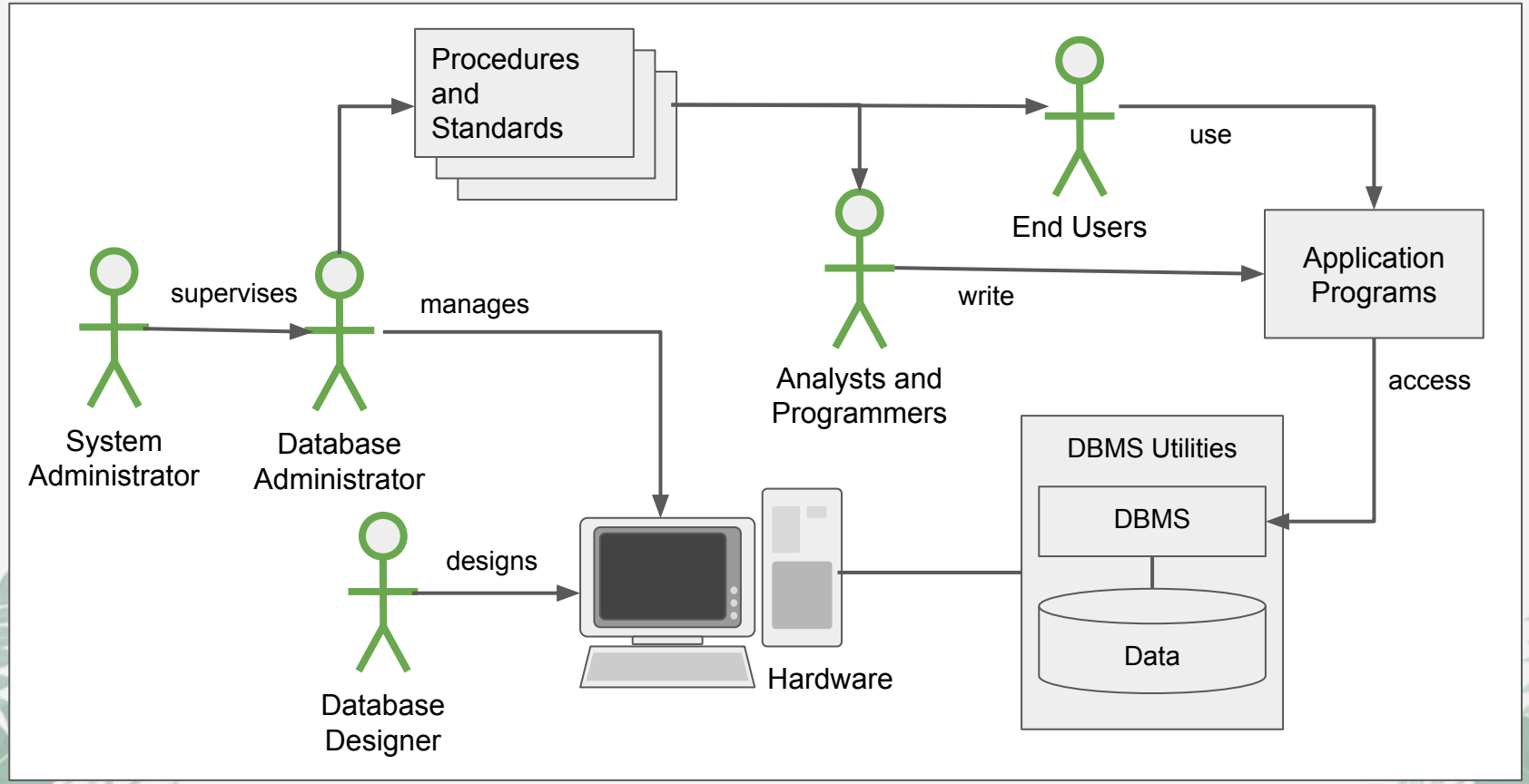
Database Systems

What is a Database System?

It is a computerized system that aims to organize and manage structured data and make that it available on demand.

Examples

- Computerized personnel records system
- Distributed version control system
- Digital data repository


Procedures and Standards
use
End Users
Application Programs
supervises
manages
write
Analysts and Programmers
access
System Administrator
Database Administrator
DBMS Utilities
DBMS
designs
Data
Hardware
Database Designer

Major Parts

1. Hardware
2. Software
   - Operating System
   - Database Management System
   - Application programs
3. People
   - System and DB Administrators
   - Database Designers
   - Systems Analysts
   - Programmers
   - End-users
4. Procedures
5. Data

Database Management Systems

What is a Database?

It is a structured and organized collection of end-user data and its metadata stored centrally, allowing management, retrieval, and manipulation.

Advantage #1: Reduced Data Redundancy

A well-designed database system aims to store data only once while ensuring that it's accessible whenever needed, preventing inefficiencies, increased storage requirements, and data inconsistency.

Advantage #2: Data Integrity

It ensures that the data stored is accurate, consistent, and reliable through various constraints, validation rules, and other mechanisms that prevent unauthorized modifications.

Advantage #3: Data Independence

It allows changes to the structure of the database or to the physical storage without directly affecting how data is accessed by an application or a system.

Advantage #4: Data Security

It protects the data from unauthorized access, tampering, or theft through authentication, authorization, and encryption.

Advantage #5: Data Consistency

It ensures that data remains accurate and valid across the database through transactional mechanisms.

Advantage #6: Ease of Use

Allows interaction with the data without the need to understand the underlying complexities of data storage and retrieval.

Advantage #7: Less Storage

Different techniques such as normalization and compression help reduce storage requirements while maintaining data integrity resulting to optimized storage compared to traditional file-based approaches.

Disadvantages

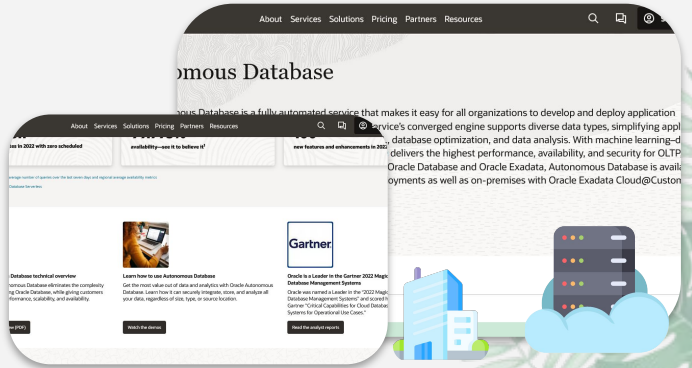
- Costs
- Complexity
- Compatibility
- Vulnerability
- Lack of lower level control

Database Management System

It is a structured and organized collection of end-user data and its metadata stored centrally, allowing management, retrieval, and manipulation.

Oracle Database

A proprietary, multi-model DMBS widely used in large enterprises for mission-critical applications.



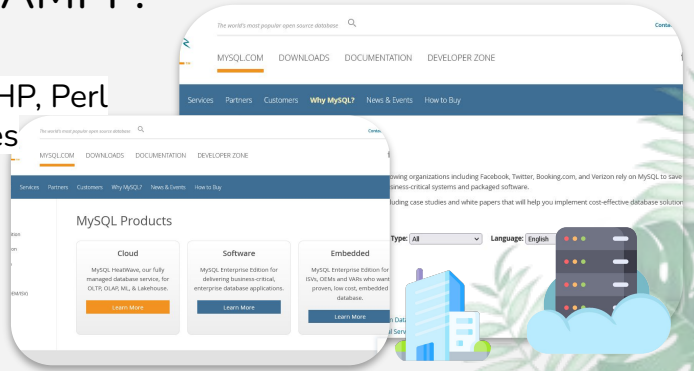
https://www.oracle.com/autonomous-database/

MySQL

An open-source rDBMS popularly used with smaller-scale web applications.

MySQL is popular with the PHP programming language, and previously supported by XAMPP.

XAMPP = Cross-platform, Apache, MySQL, PHP, Perl
Forked with MariaDB, due to proprietary issues

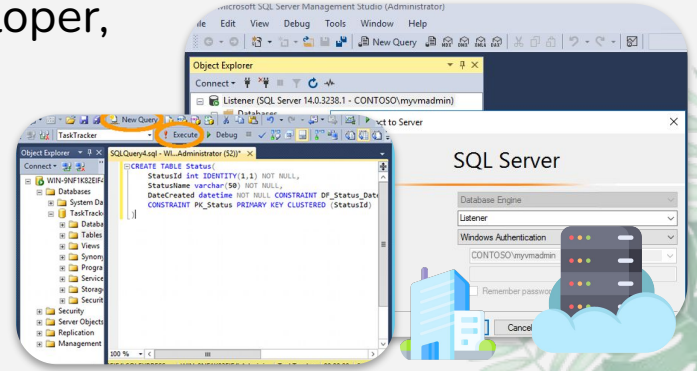


https://www.mysql.com/products/

Microsoft SQL Server

An rDBMS developed by Microsoft, common in Windows platforms and is widely used in large and mid-scale enterprises.

It is available in express, developer, standard, enterprise, and other editions.

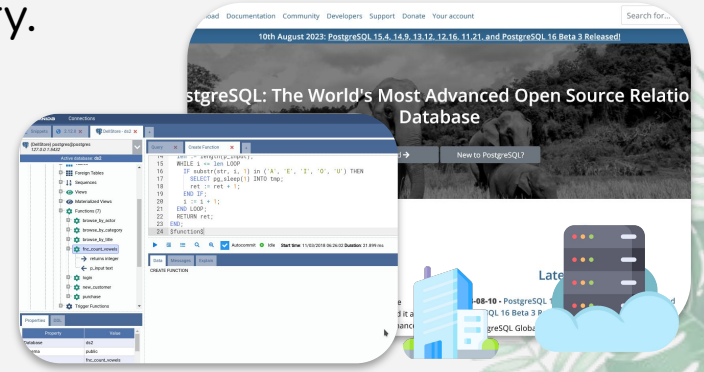


https://www.microsoft.com/en-us/sql-server/sql-server-downloads

PostgreSQL

Better known as Postgres is an open-source rDBMS advanced features, extensibility, and compliance to SQL standards.

It is popular across the industry.



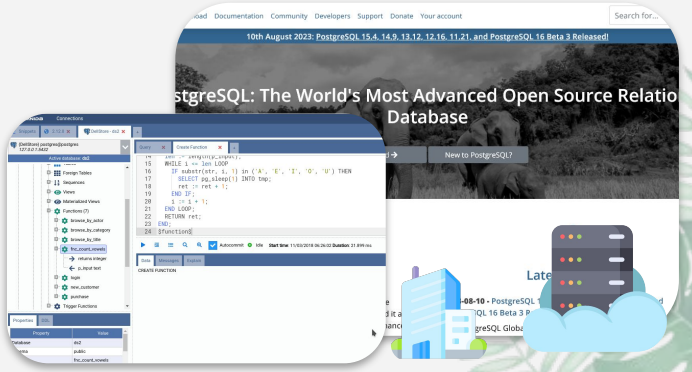
PostgreSQL

https://www.postgresql.org/

SQLite

A cross-platform, lightweight, open-source, serverless, and self-contained rDBMS that supports a subset of SQL and can be accessed directly through various programming languages.

It is popular embedded database used in mobile and desktop apps.

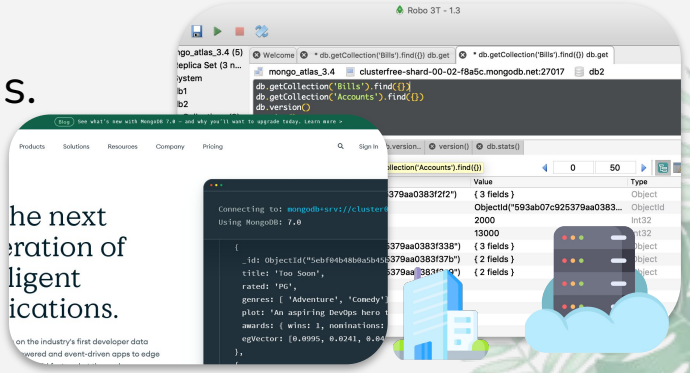


https://www.sqlite.org/fullsql.html

MongoDB

An open-source NoSQL DBMS designed to handle large volumes of unstructured and semi-structured data, stored JSON-like documents in flexible schemas.

It is popular in web applications.



https://www.mongodb.com/

Other DBMS

- IBM Db2 proprietary data management products
- Cassandra NoSQL, distributed data
- Couchbase NoSQL, distributed data
- Elasticsearch NoSQL, near real-time
- Amazon DynamoDB NoSQL, AWS-managed

Function #1: Data Dictionary Management

It can manage the metadata about the structure and organization of the database such as the information about data types, table relationships, constraints, and indices.

Function #2: Data Storage Management

It abstracts the complexities of the logical and physical storage while ensuring optimal storage and performance.

Function #3: Data Integrity Management

It enforces various constraints to ensure that the data adheres to the predefined rules to maintain accuracy and consistency.

Function #4: Data Transformation

It allows transformation and presentation of raw data into various supported visualizations that are meaningful to the users.

Function #5: Security Management

It ensures a restrictive access to stored data through various authentication, authorization, and encryption mechanisms.

NOTE: Industry-grade DBMSs offer different methods of network and storage level encryptions to secure data from malicious access and during transmission between clients and the database server.

Function #6: Multi-user Access Control

It can handle concurrent access from multiple applications and users to the the same database, through locking mechanisms when different users try to modify data simultaneously.

NOTE: Systems with high traffic loads may implement more advanced architectures and algorithms including serialization and rollback mechanisms to improve scalability, such as in airline booking and banking systems. These mechanisms are typically implemented on the system level, on top of the DBMS functionalities.

Function #7: **Backup and Recovery**

It facilitates database backup and recovery mechanisms as protection against hardware failures, software glitches, civil unrest, calamities, and other unforeseen events.

NOTE: In order to allow *Business Continuity* during unforeseen events, large organizations employ continual research and the development of protocols that may help to minimize financial loss and operational disruption.

Function #8: DB Communication Interfaces

DBMS developers or vendors provide standard drivers and APIs to allow clients to seamlessly communicate between database servers.

Examples

- Open Database Connectivity (ODBC)
- Java Database Connectivity (JDBC)
- Object Linking and Embedding Database (OLE DB)
- Representational State Transfer APIs (REST)
- Native DB Drivers

Function #9: DB Access Languages and APIs

DBMS provides an interface to programmatically communicate with and manipulate databases.

Categories of Operations

- Data Definition Language (DDL)
- Data Manipulation Language (DML)

NOTE: SQL is one of the popular language used in DBMS, but other open-source and proprietary languages and SQL variants exists.

ACID Properties

Atomicity

Each transaction is a single unit of work, when an operation fails, the transaction itself fails.

Consistency

A valid data remains valid during and after a transaction, and after one state to another.

Isolation

The operations of a transactions do not interfere with the other concurrent transactions, until the first one is completed.

Durability

Successful transactions are guaranteed to be permanent even with system failures, crashes, or power outages.

Evolution of DBMS

pre-1960s

File Systems

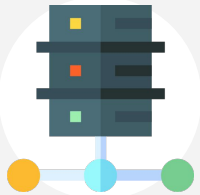
1960s

Hierarchical & Network Models

1970s 1980s

Relational & SQL DBMS

1990s



ORMs

Distributed & Client-Server Architectures

Data Warehousing & OLAP

2000s

NoSQL

2010s

Cloud-native & Serverless

Blockchains

DBMS #1 Relational Databases

RDBMS is an ACID-compliant DBMS that supports SQL and structured data, making it suitable for well-defined schemas, transactional requirements, and complex queries.

Examples: MySQL, MariaDB, PostgreSQL, Microsoft SQL Server, SQLite

DBMS #2 NoSQL

Not only SQL is a class of DBMS that supports dynamic and flexible schema and different data models, it aims for high availability and horizontal scalability, making it suitable for applications with evolving data structures, high write loads, and distributed systems.

Examples: MongoDB, Redis

DBMS #3 Graph

It supports Graph structures, excelling in modeling and powerful querying traversals and analyses of complex relationships of data, making it suitable for social network analysis, recommendation systems, and fraud detection.

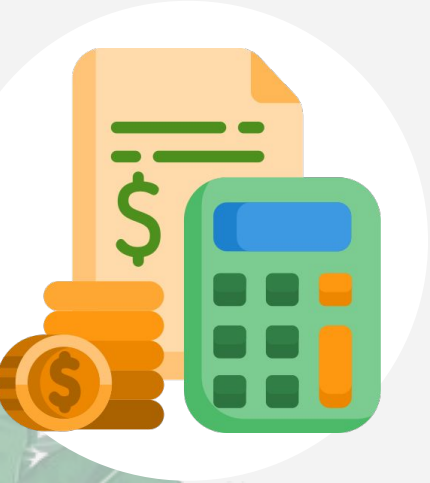
Examples: Neo4j, Amazon Neptune

DBMS Other Types

- Columnar Apache Cassandra, HBase
- Time-series InfluxDB, TimescaleDB
- In-memory Redis, Memcached
- NewSQL Google Spanner, CockroachDB
- Multi-model ArangoDB, Couchbase
- Distributed Apache Cassandra, Amazon DynamoDB
- Cloud Amazon RDS, Azure SQL Database

Major Applications of DBMS

Business and Finance



Businesses use databases in mission-critical Transaction Processing, Customer Relationship Management (CRM), and Accounting and Financial Management systems.

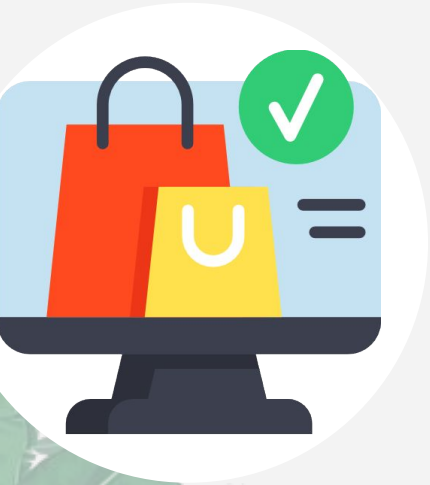
Healthcare

Databases help healthcare and medical professionals perform informed decisions and precise diagnostics through Electronic Health Records and Clinical Decision Support systems.

Education

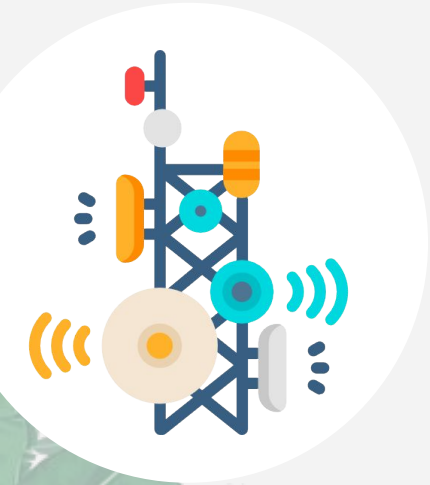
Different specialized Learning Management Systems (LMS) and Student Information Systems allow institutions to manage interactive educational experiences and academic history.

E-commerce



Databases allows on-demand and real-time interaction between businesses and customers through online Order and Inventory Management and custom-tailored product catalogs.

Telecommunications

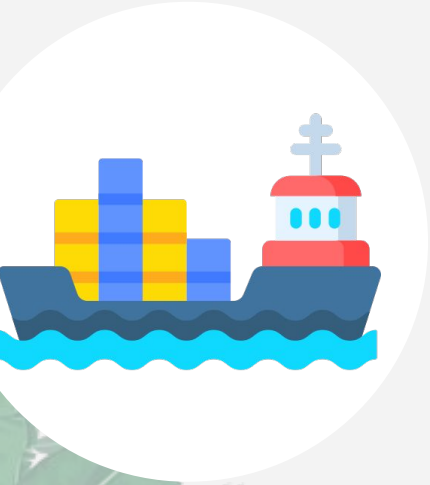


People and communities worldwide become interconnected through up-to-date Network and Subscriber Management Systems that manage both its infrastructure and its stakeholders data.

Government and Public Services

Government operations are now becoming more transparent and optimized through the digitalization of Public Records and Law Enforcement.

Logistics and Supply Chain

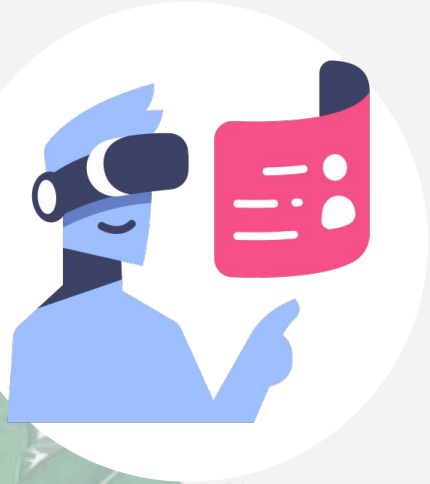


The interconnected operations of national and transnational industries are optimized through Inventory and Warehouse Management and Route Optimization systems.

Scientific Research

Databases promote scientific breakthroughs and innovations through online scientific archives, Data Analysis and Simulation, and Genomics and Bioinformatics.

Media and Entertainment



Digital media and entertainment portals proliferate through the highly-available Content Management Systems (CMS) and automated Digital Rights Management.

Energy and Utilities

Utility companies optimize operations through the implementation of Smart Grids and Asset Management systems.

Summary

- A file-based approach is prone to inefficiency redundancy.
- A database system includes HW, SW, people, procedure, and data.
- A database is a collection of a structured of data, designed to support efficient data storage, retrieval and maintenance.
- A DBMS is software that facilities the creation, organization, and management of databases, including features such as data storage, retrieval, manipulation, integrity, and security.

Questions?

Thank You!

References

- Elmasri, R., et al. (2016). Fundamentals of Database Systems, 7th ed. Pearson Higher Education, 221 River Street, Hoboken, NJ 07030.
- Aparajitha. R.S.V, et al.. (2010). Database Management Systems. International Journal of Computer Applications. 1. 10.5120/179-310.
- Date, C.J. (2004), Introduction to Database Systems (8th Edition) Addison Wesley.
- Dietrich, S. W. (2021). Understanding databases: Concepts and practice. Wiley.