Clemson *Means*
BUSINESS

Chapter 1
Introduction to Data Management

Instructor: He Li

DATA MANAGEMENT

is a discipline that focuses on the proper *generation*, *storage*, and *retrieval* of data.

Why do we need data management?

In today's world, **DATA** is

- Ubiquitous (i.e., abundant, global, and everywhere)
- Pervasive (i.e., unescapable, prevalent, and persistent)

A Day In Susan's Life

See how many databases she interacts with each day

Before leaving for work, Susan checks her Facebook and Twitter accounts

Where is the data about the friends and groups stored?

Where are the "likes" stored and what would they be used for?

On her lunch break, she picks up her prescription at the pharmacy

Where is the pharmacy inventory data stored?

What data about each product will be in the inventory data?

What data is kept about each customer and where is it stored?

After work, Susan goes to the grocery store

Where is the product data stored?

Is the product quantity in stock updated at checkout?

Does she pay with a credit card?

At night, she plans for a trip and buys airline tickets and hotel reservations online

Where does the online travel website get the airline and hotel data from?

What customer data would be kept by the website?

Where would the customer data be stored?

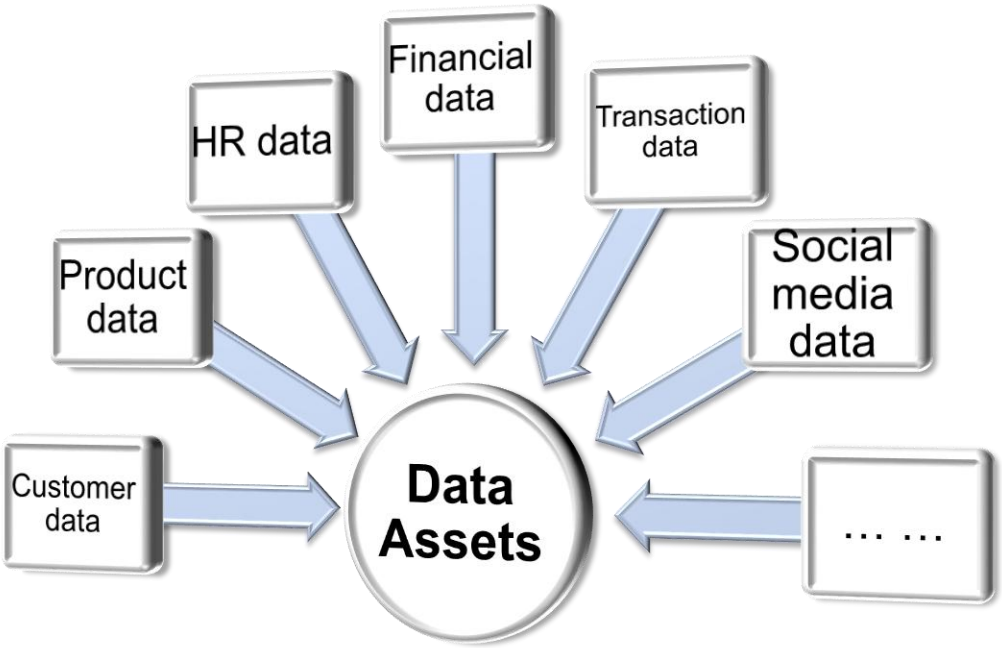
Then she makes a few online purchases

Where are the product and stock data stored?

Where does the system get the data to generate product "recommendations" to the customer?

Where would credit card information be stored?

DATA MANAGEMENT for business …

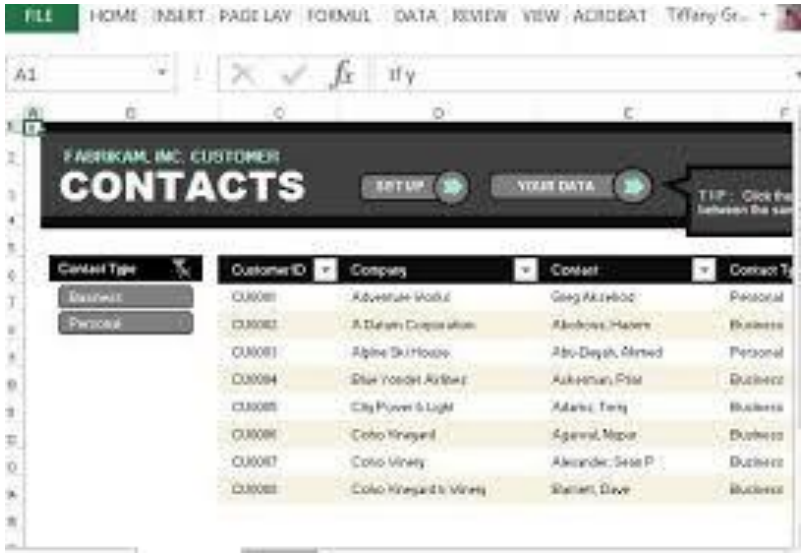


Data as Organizational Resources

- Data Collection
- Data Storage
- Data Aggregation
- Data Manipulation
- Data Dissemination
- Data Management

Size and Complexity of DATA varies …

Department of MANAGEMENT Clemson University



Customer Contact Lists

AT&T's trillions of phone calls Data

Google responds to over 91 million searches per day

Applications of Database

Department of MANAGEMENT
Clemson University

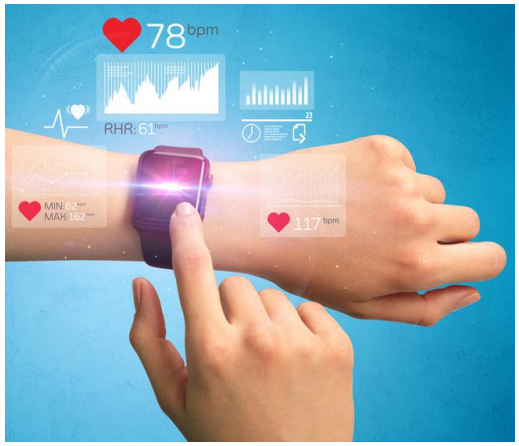
City	State	Type of breach	Type
Novato	California	PORT	BSF
Evansville	Indiana	INSD	BSF
Columbia	South Carolina	DISC	EDU
Honolulu	Hawaii	INSD	BSO
New York	New York	PORT	MED
Columbus	Georgia	PORT	BSF
New York	New York	DISC	BSF
Fairbanks	Alaska	HACK	EDU
Newport Beach	California	PORT	BSF
Seattle	Washington	PORT	BSO
Austin	Texas	HACK	EDU
New Paltz	New York	HACK	EDU
Charlottesville	Virginia	STAT	EDU
West Lafayette	Indiana	HACK	EDU
Syracuse	New York	HACK	BSO
Jersey City	New Jersey	PORT	BSF

Numeric and alphanumeric data

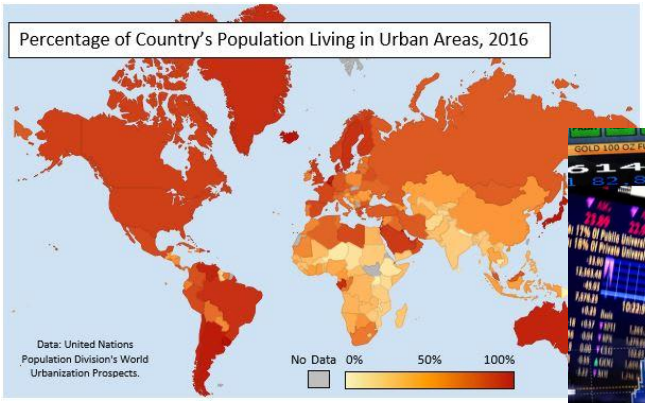
YouTube

Spotify

Multimedia data



Fitness tracking data

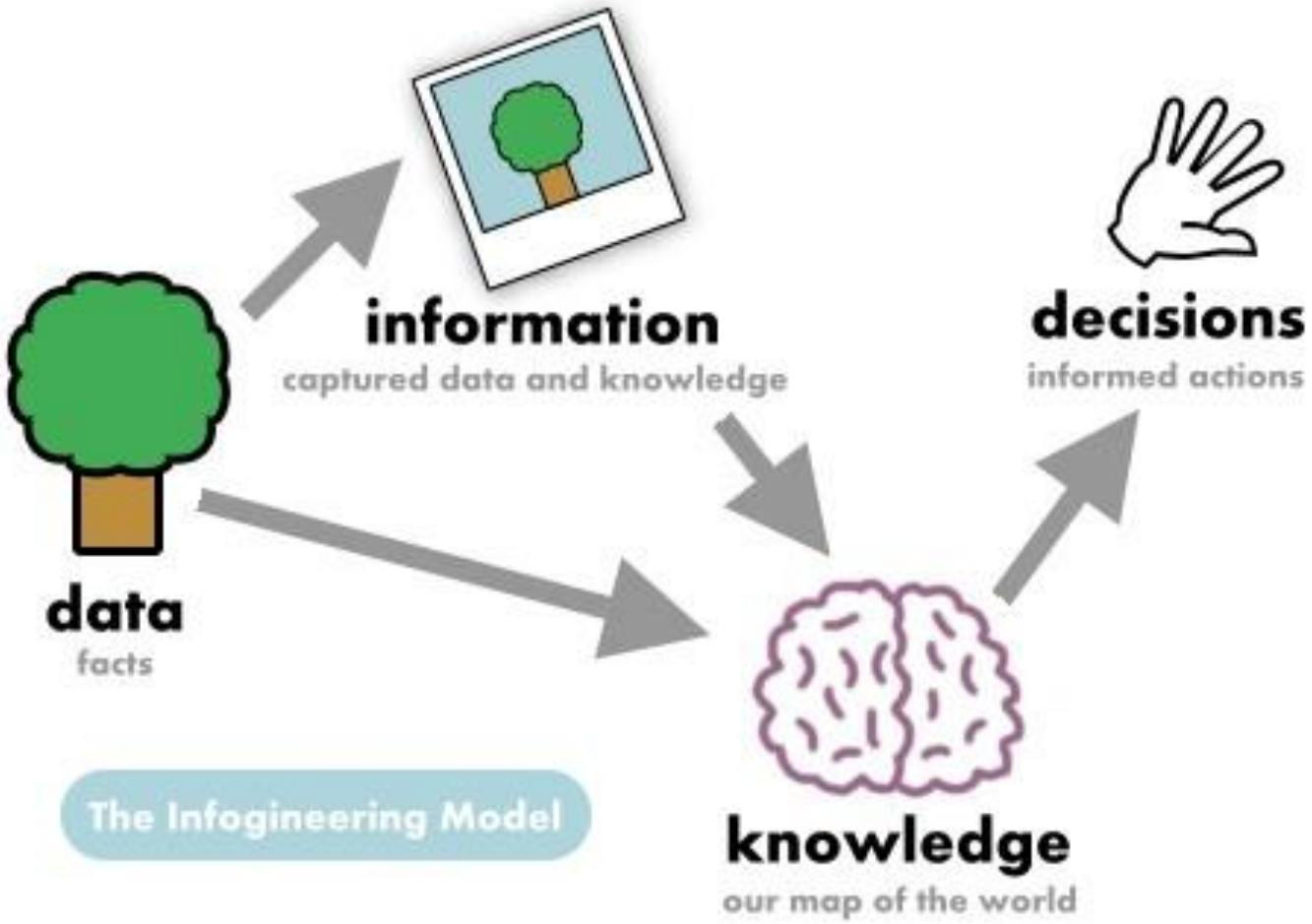


Geographical data

Biometric data

Volatile data

Big data


information
captured data and knowledge
decisions
informed actions
data
facts
knowledge
our map of the world
The Infogineering Model

Example: from Data to Information

Department of MANAGEMENT Clemson University

Raw Data

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145
Sisneros, Michael	392416582

Class Roster

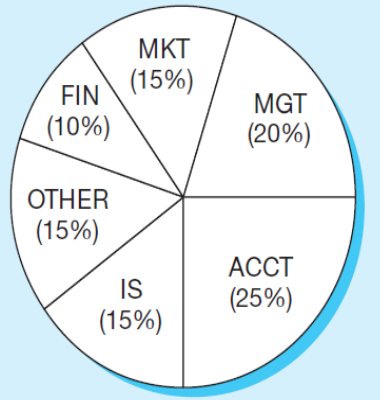
Course: MGT 500 Business Policy

Semester: Spring 2018

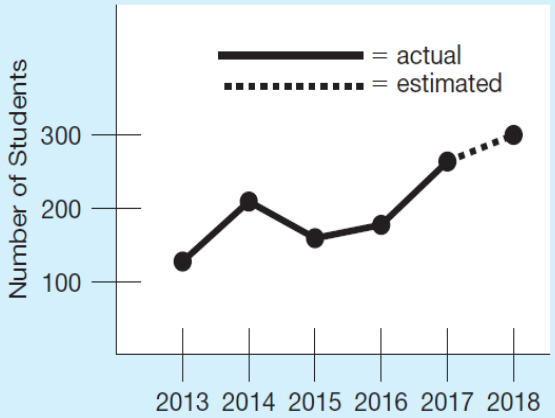
Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Summarized Data



Percent Enrollment by Major (2018)

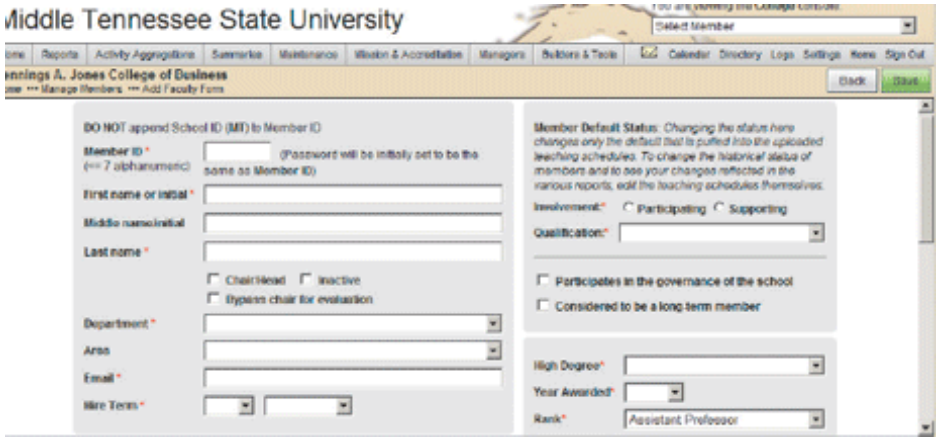


Enrollment Projections

Data in Context

Example: from Data to Information

Department of MANAGEMENT
Clemson University



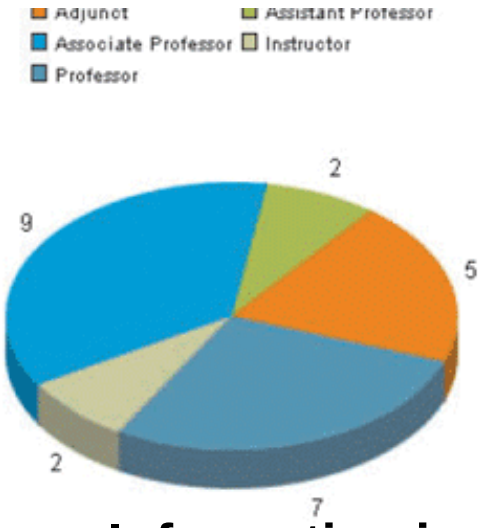
Data Entry Screen

Id	LastName	MidName	FirstName	DeptCode	Office	Email	Rank	HireYear	Degree
1	Washington	A.	George	MGMT	N135	gwashington@mtsu.edu	Professor	2001	Ph.D.
2	Adams		John	FIN	N313	jadams@mtsu.edu	Professor	1984	Ph.D.
3	Jefferson	L.	Thomas	ECON		tjefferson@mtsu.edu	Instructor	2002	M.B.A.
4	Madison	D.	James	FIN	N236	jmadison@mtsu.edu	Associate Professor	1994	Ph.D.
5	Monroe	N.	James	ACCT	N411	jmonroe@mtsu.edu	Assistant Professor	1995	Ph.D.
6	Adams	G.	John	ACCT	N418	jqadams@mtsu.edu	Associate Professor	1989	Ph.D.
7	Jackson	C.	Andrew	ECON	N303	ajackson@mtsu.edu	Associate Professor	1999	Ph.D.
8	Van Buren	T.	Martin	FIN	N306	mvanburen@mtsu.edu	Professor	1988	Ph.D.
9	Harrison	R.	William	MKTG	N418	wharrison@mtsu.edu	Professor	1994	Ph.D.
10	Tyler	M.	John	MGMT		JTyler@mtsu.edu	Assistant Professor	2000	Ed.D.
11	Polk		Cheryl	MKTG	N143	cpolk@mtsu.edu	Associate Professor	2002	Ph.D.
12	Taylor	G.	Zachary	ACCT	N415	gtaylor@mtsu.edu	Associate Professor	1996	Ph.D.
13	Fillmore		Millard	JCB	N219	mfillmore@mtsu.edu	Professor	1992	Ph.D.
14	Pierce	A.	Franklin	MKTG	N359	pfranklin@mtsu.edu	Instructor	2005	M.B.A.
15	Buchanan	T.	James	MGMT	N146	jbuchanan@mtsu.edu	Associate Professor	1996	D.B.A.
16	Lincoln	W.	Leroy	MGMT	N158	llincoln@mtsu.edu	Associate Professor	1996	Ph.D.
18	Johnson		Andrew	ISYS	N368	ajohnson@mtsu.edu	Professor	1987	Ph.D.
19	Grant		Kelly	MKTG	N128	kgrant@mtsu.edu	Assistant Professor	1989	D.B.A.
20	Rutherford		Hayes	ACCT	N408	hrutherford@mtsu.edu	Professor	1992	Ph.D.
21	Garfield	T.	Denise	ACCT		dgarfield@mtsu.edu	Assistant Professor	2010	Ph.D.
22	Arthur		Emily	ACCT	N413	earthur@mtsu.edu	Associate Professor	2003	J.D.
23	Cleveland	G.	Robert	ACCT	N401	rcleveland@mtsu.edu	Associate Professor	1997	Ph.D.
24	Harrison	X.	Patricia	BULA	N406	pharrison@mtsu.edu	Associate Professor	2001	J.D.
25	McKinley	B.	Priscilla	ISYS	N363	pmckinley@mtsu.edu	Adjunct	1994	M.S.
26	Roosevelt	F.	Hillary	MGMT	N104	hroosevelt@mtsu.edu	Associate Professor	2002	Ph.D.
27	Wilson		Laura	BCEN	N448	lwilson@mtsu.edu	Professor	1992	Ph.D.
28	Harding		Warren	MKTG	N414	wharding@mtsu.edu	Professor	1984	Ed.D.
29	Coolidge		Calvin	ECON	N316	ccoolidge@mtsu.edu	Professor	1975	Ph.D.
30	Hoover		Lisa	MGMT		lhoover@mtsu.edu	Adjunct	1978	M.B.A.
31	Truman		Betty	ACCT	N416	btruman@mtsu.edu	Professor	1971	Ed.D.
32	Johnson		Robert	BCEN	N248	rjohnson@mtsu.edu	Professor	2001	Ph.D.

Raw Data

Rank	COUNT	%/NFS	TOT/COL	%/COL. TOT.	%/COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

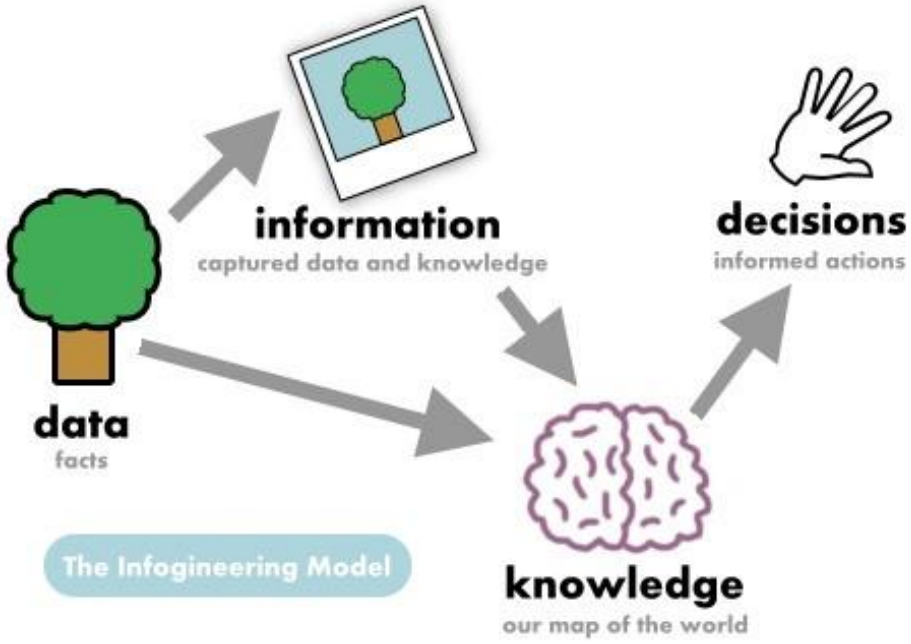
Information in Summary Format



Information in Graphical Format

Summary: Data → Information

- Data constitutes the building blocks of information.
- Information is produced by processing data.
- Information is used to reveal the meaning of data.
- Accurate, relevant, and timely information is the key to good decision making.
- Good decision making is the key to organizational survival in a global environment.



File System Data Processing

Department of MANAGEMENT Clemson University

Manual File Systems

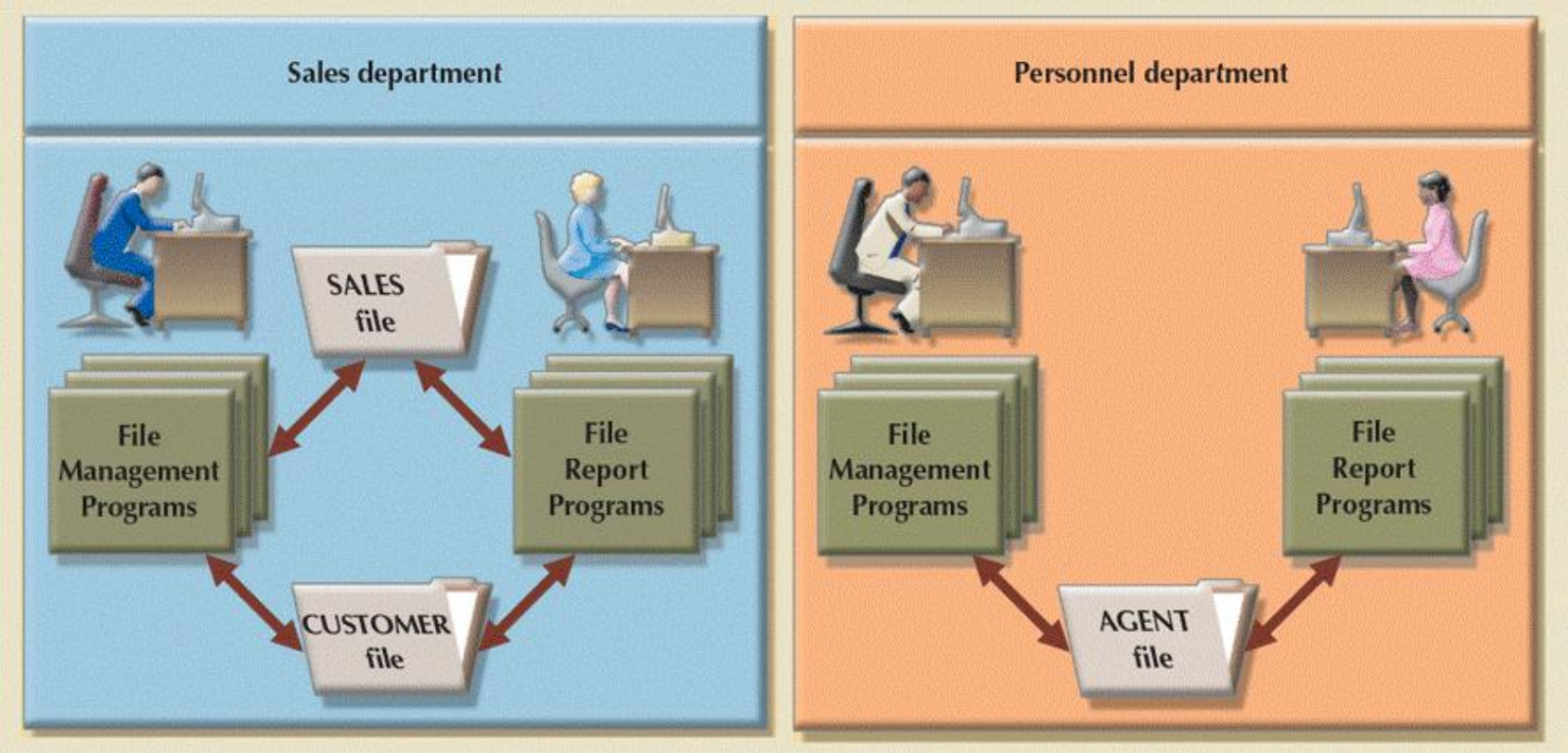
Computerized File Systems

Excel

Google Sheets

File Systems Redux

Disadvantages of File System Data Processing



Program-Data Dependence

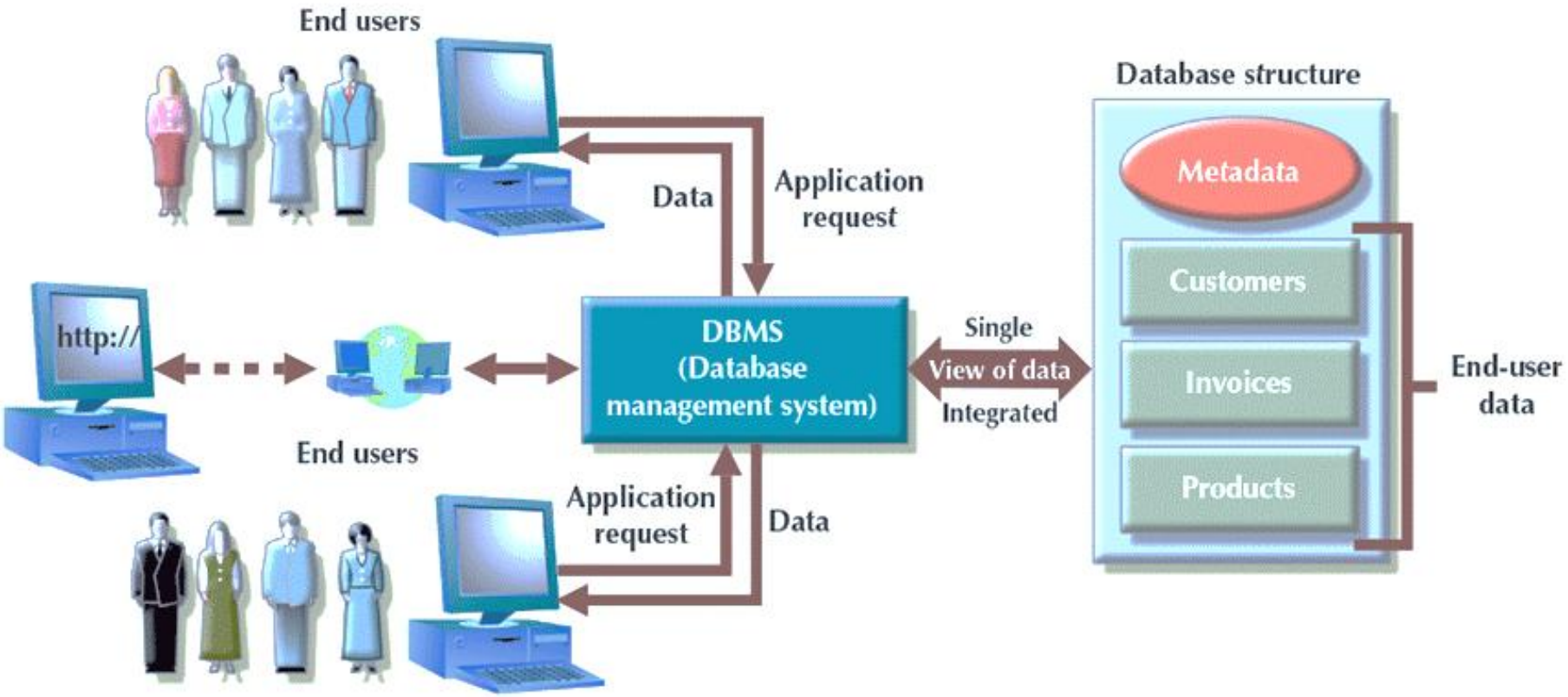
Duplication of Data

Limited Data Sharing

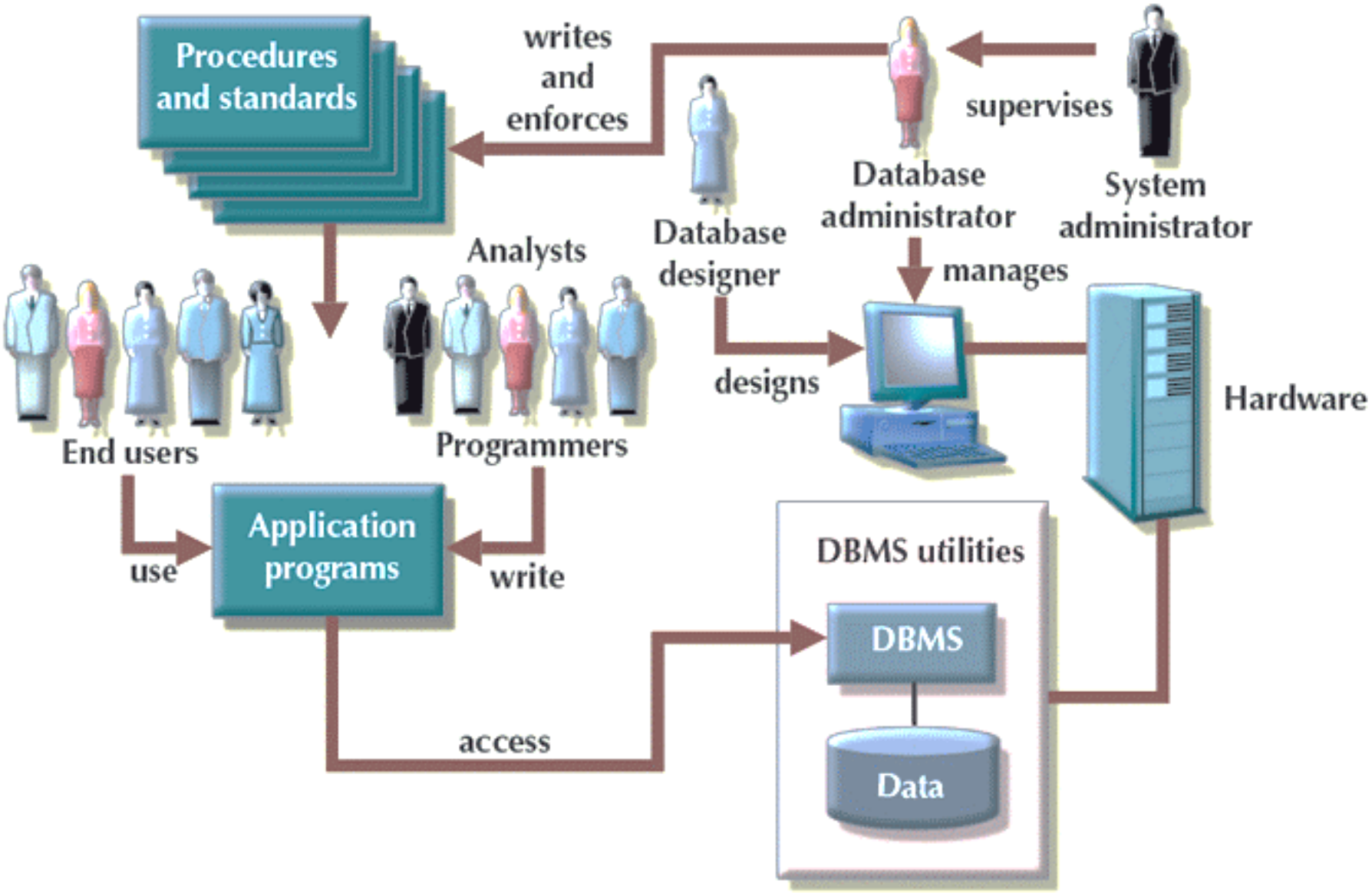
Lengthy Development Times

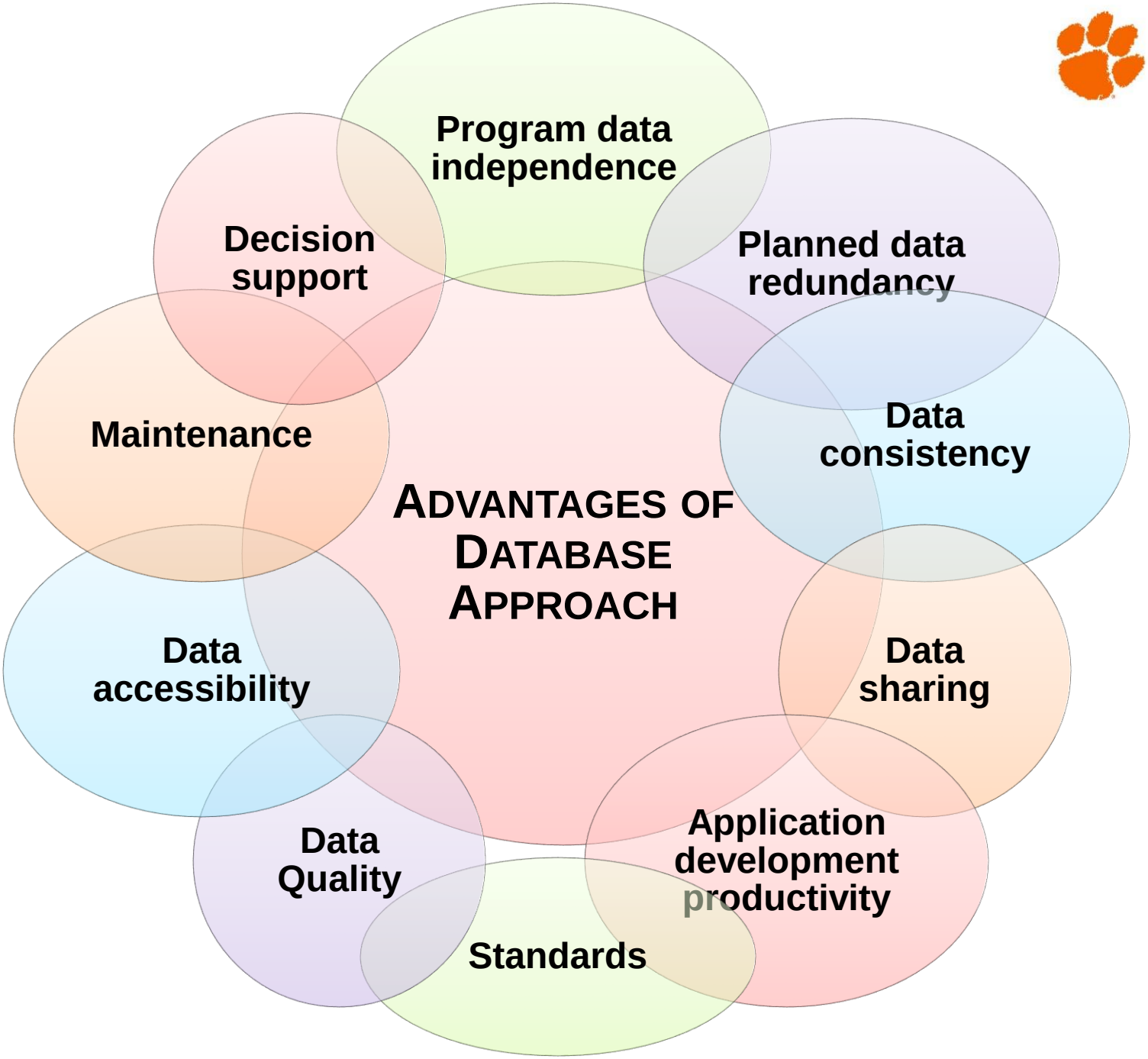
Excessive Program Maintenance

The Database Approach

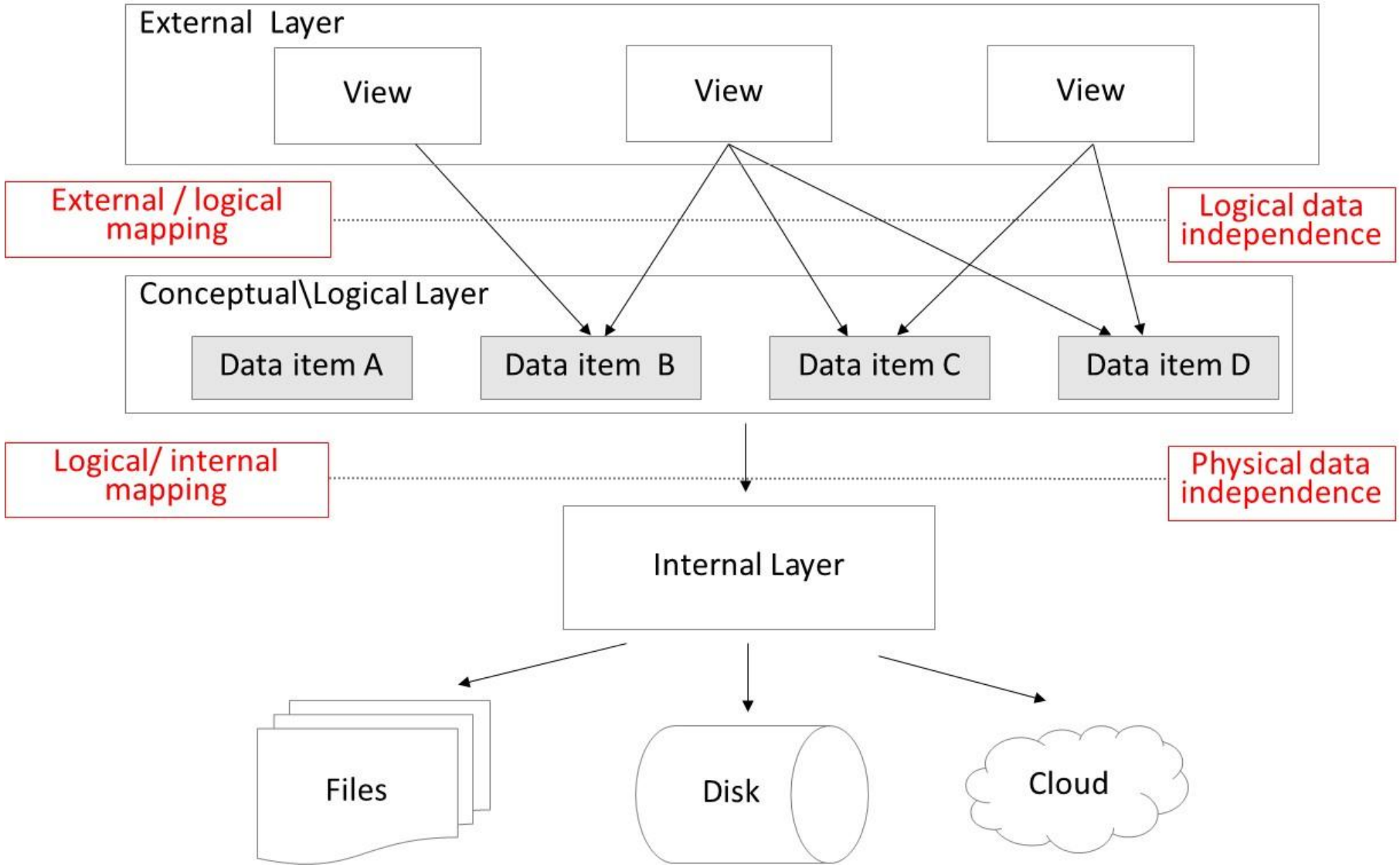


Database System Environment

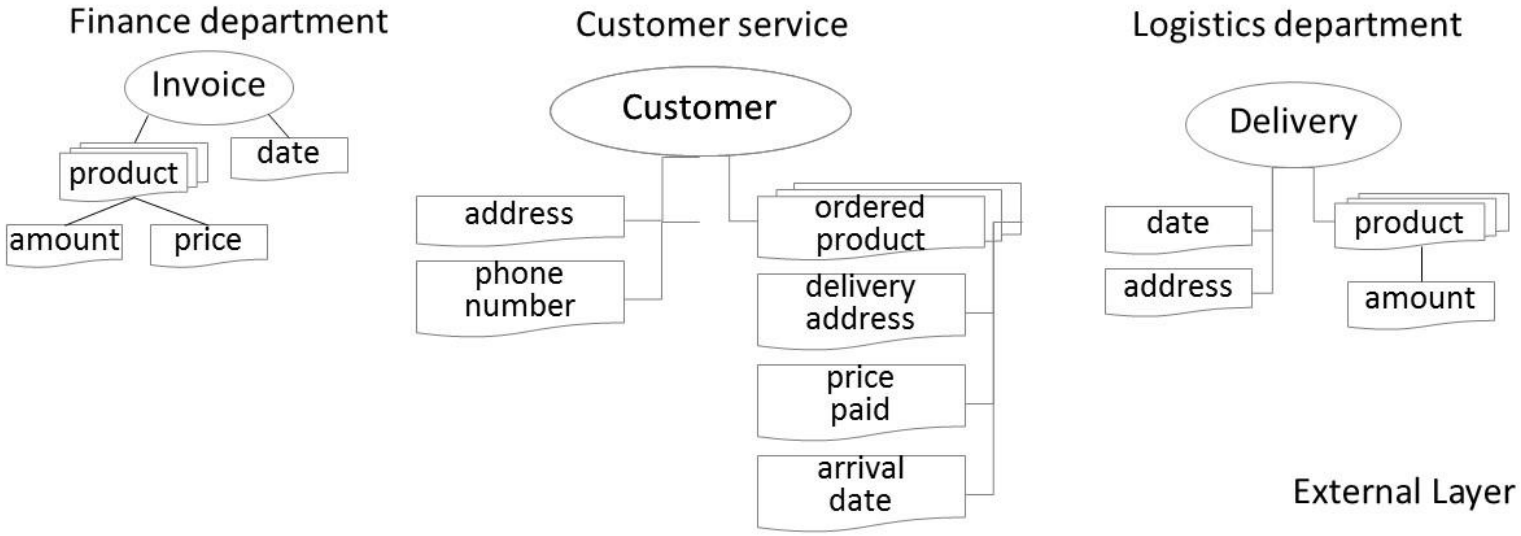



Program data independence
Decision support
Planned data redundancy
Maintenance
Data consistency
ADVANTAGES OF DATABASE APPROACH
Data accessibility
Data sharing
Data Quality
Application development productivity
Standards

Three-Layer Database Architecture

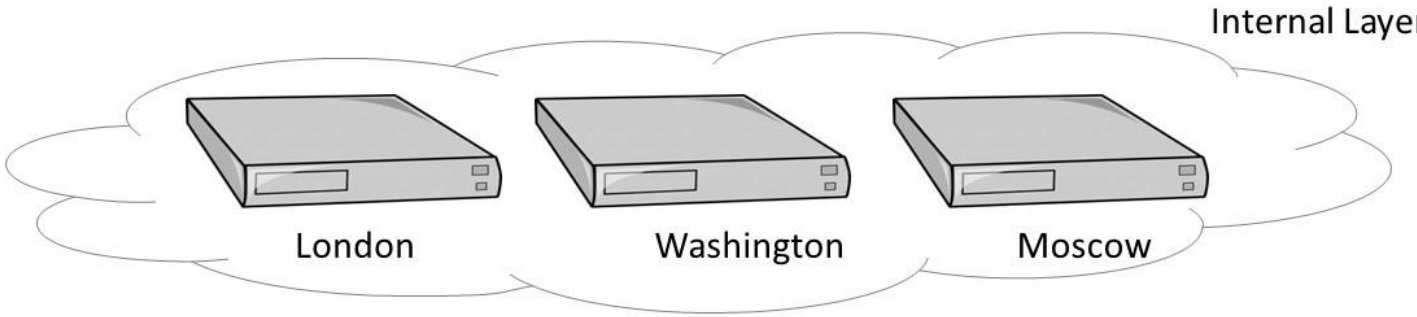


Example of the Three-Layer Database Architecture



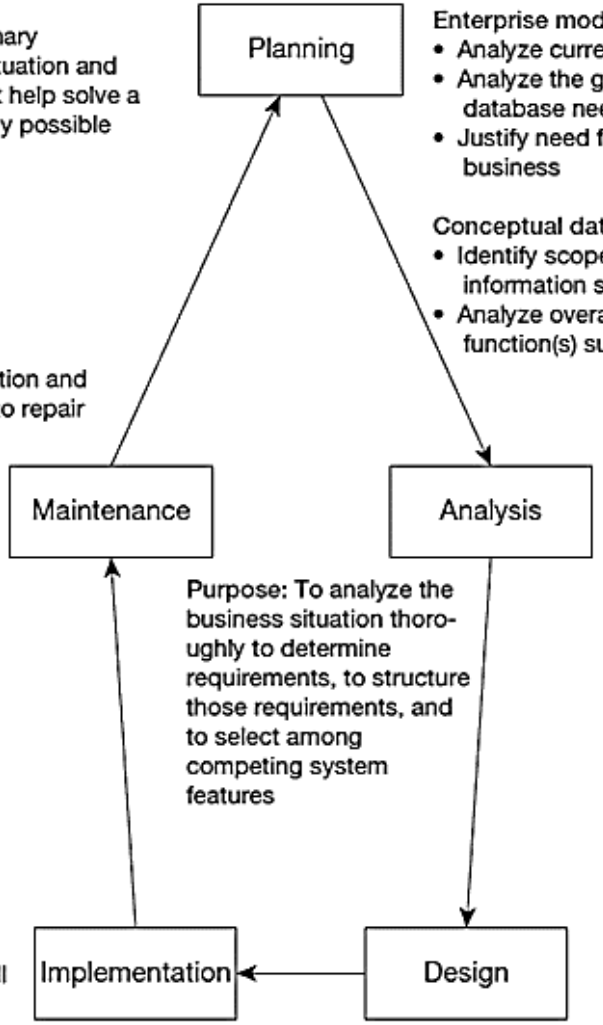
Product	name, description, cost, …
Customer	name, phone, address, …
Invoice	customer, date, products (with price and amount), …
Delivery	invoice, address, date, …

Conceptual\Logical Layer



Department of MANAGEMENT Clemson University

Developing a Database



Purpose: To develop a preliminary understanding of a business situation and how information systems might help solve a problem or make an opportunity possible

Enterprise modeling
- Analyze current data processing
- Analyze the general business functions and their database needs
- Justify need for new data and databases in support of business

Conceptual data modeling
- Identify scope of database requirements for proposed information system
- Analyze overall data requirements for business function(s) supported by database

Purpose: To monitor the operation and usefulness of the system, and to repair and enhance the system

Database maintenance
- Analyze database and database applications to ensure that evolving information requirements are met
- Tune database for improved performance
- Fix errors in database and database applications and recover database when it is contaminated

Purpose: To analyze the business situation thoroughly to determine requirements, to structure those requirements, and to select among competing system features

Conceptual data modeling, cont'd.
- Develop preliminary conceptual data model, including entities and relationships
- Compare preliminary conceptual data model with enterprise data model
- Develop detailed conceptual data model, including all entities, relationships, attributes, and business rules
- Make conceptual data model consistent with other models of information system
- Populate repository with all conceptual database specifications

Purpose: To write programs, build databases, test and install the new system, train users, and finalize documentation

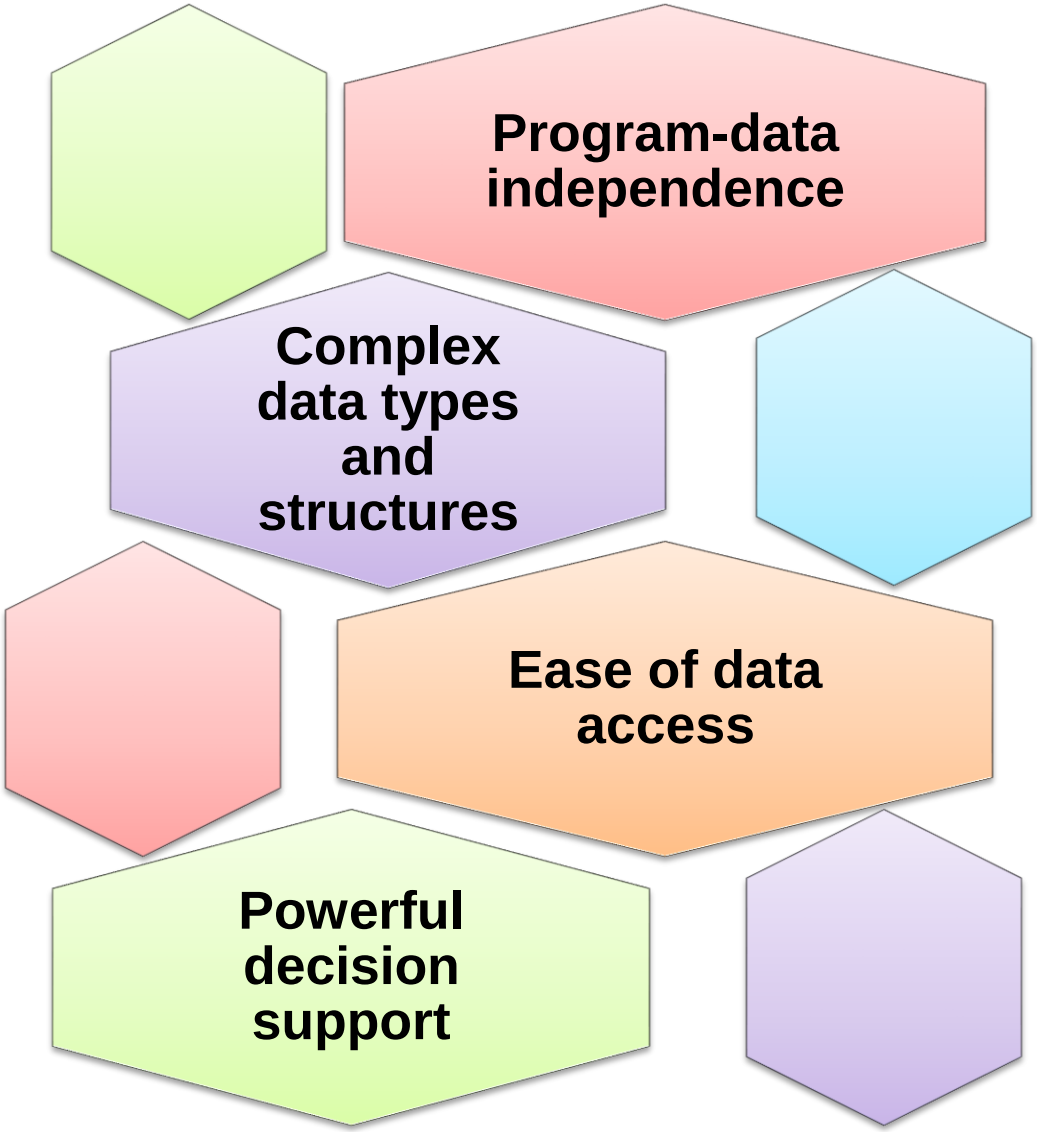
Purpose: To elicit and structure all information requirements; to develop all technology and organizational specifications

Logical database design
- Analyze in detail the transactions, forms, displays, and inquiries (database views) required by the business functions supported by the database
- Integrate database views into conceptual data model
- Identify data integrity and security requirements, and populate repository

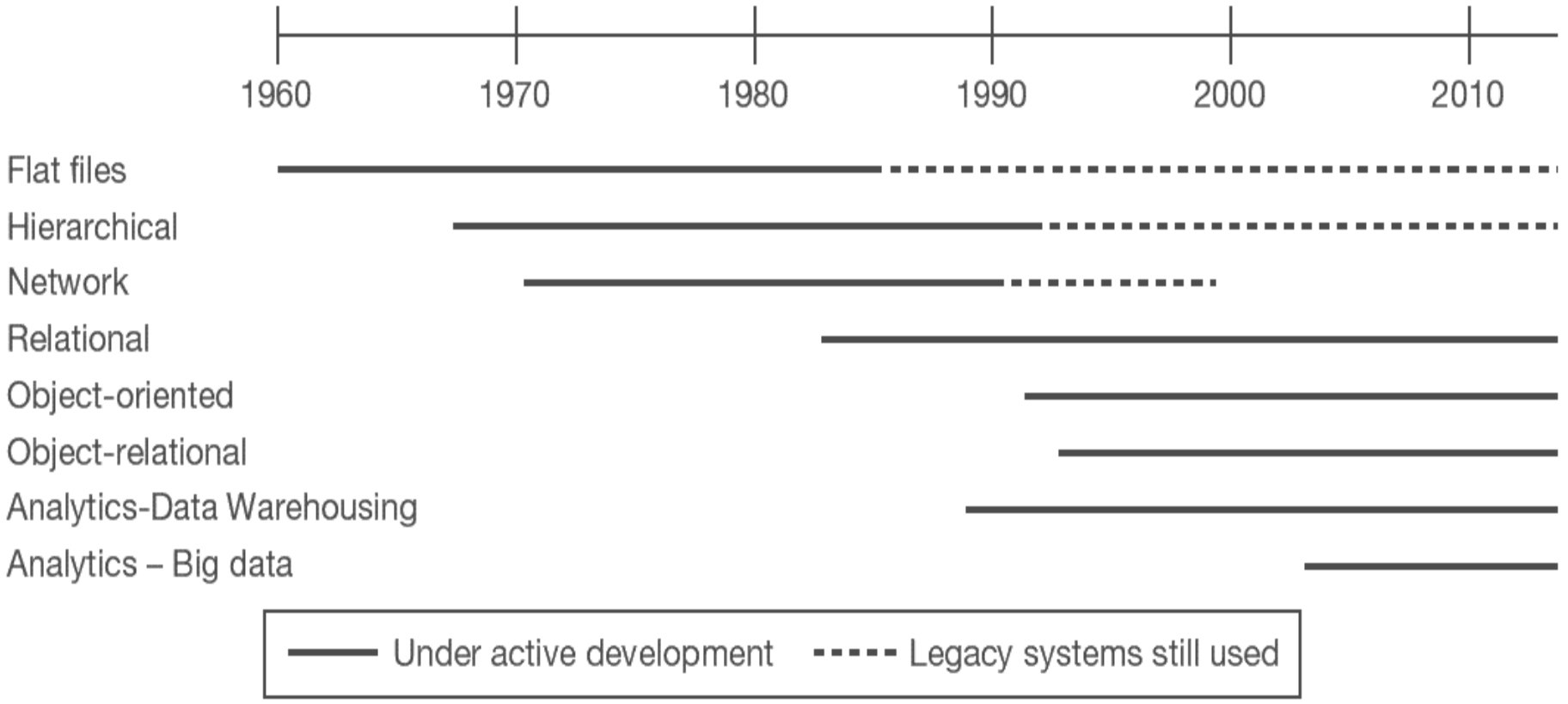
Physical database design and definition
- Define database to DBMS (often generated from repository)
- Decide on physical organization of data
- Design database processing programs

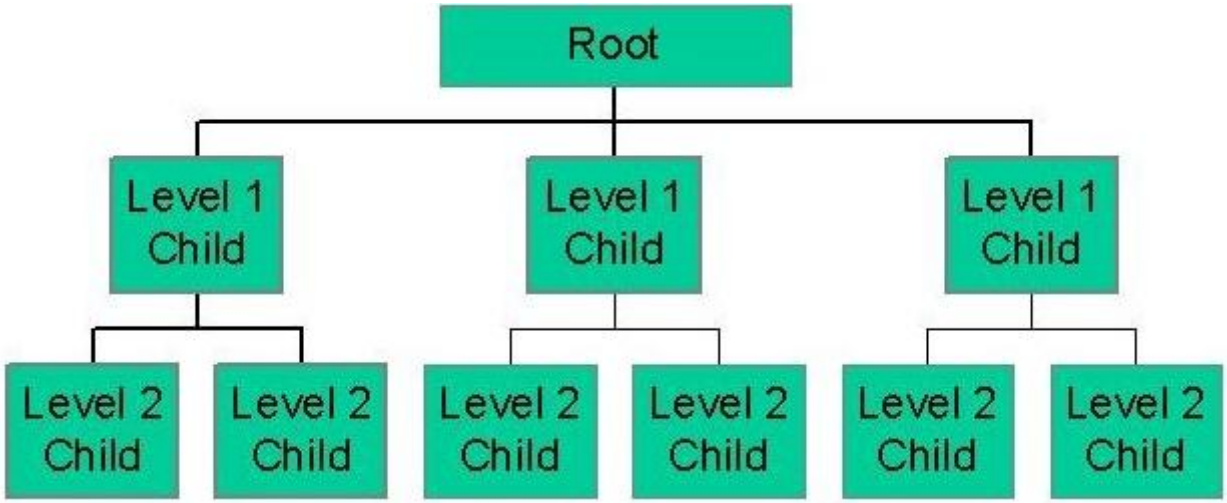
Database implementation
- Code and test database processing programs
- Complete database documentation and training materials
- Install database and convert data from prior systems

Evolving Database Systems



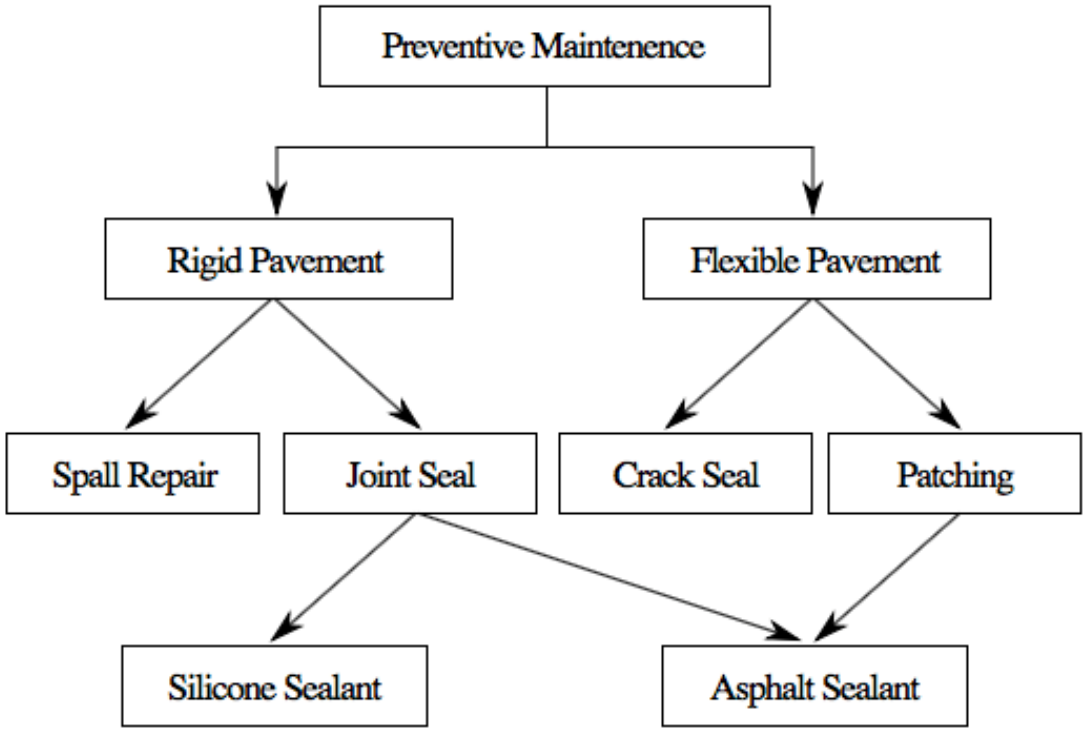
Evolution of Database Systems





Hierarchical Database

Network Database



RELATION 1 (PRIMARY KEY, ATTRIBUTES...)

RELATION 2 (PRIMARY KEY, FOREIGN KEY, ATTRIBUTES...)

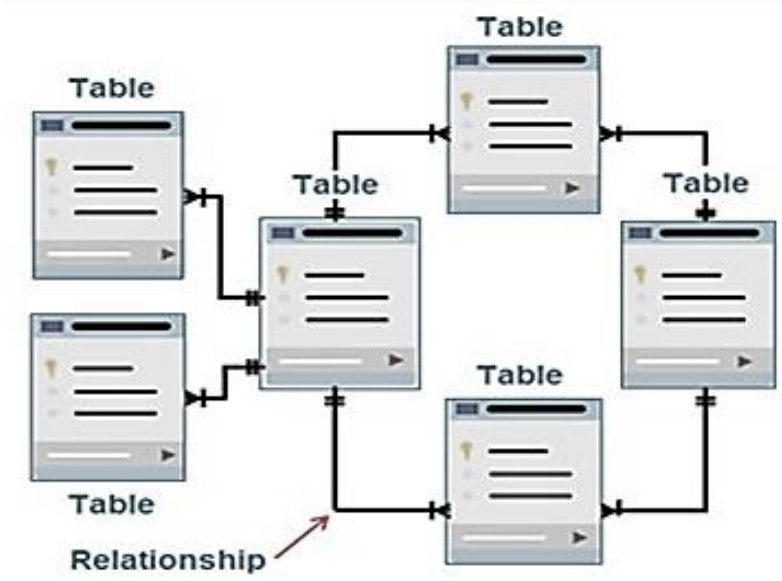
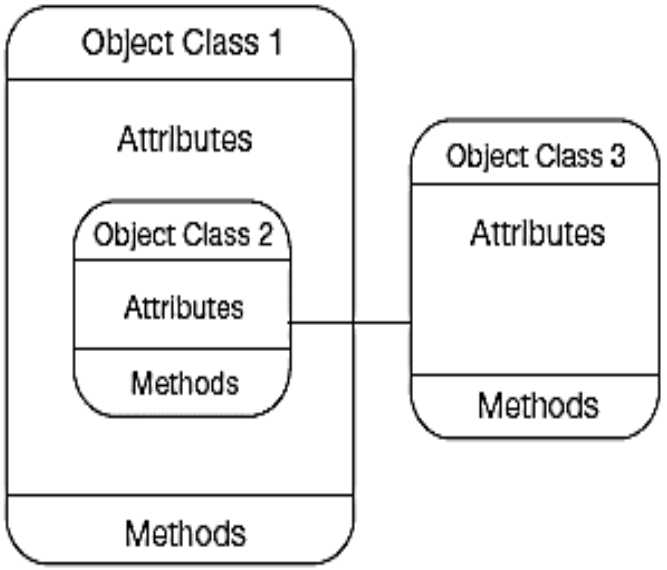


Table = Relation = Entity = Concept = Object

PK - Primary Key And FK - Foreign Key

Relational Database

Department of MANAGEMENT
Clemson University



Object 1: Maintenance Report

Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

Object 1 Instance

01-12-01
24
I-95
2.5
6.0
6.0

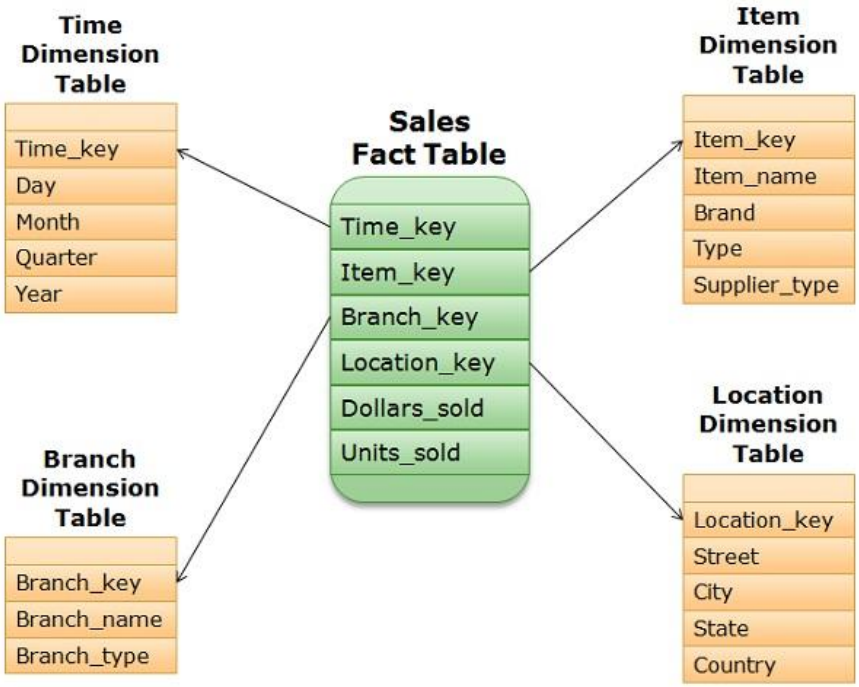
Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

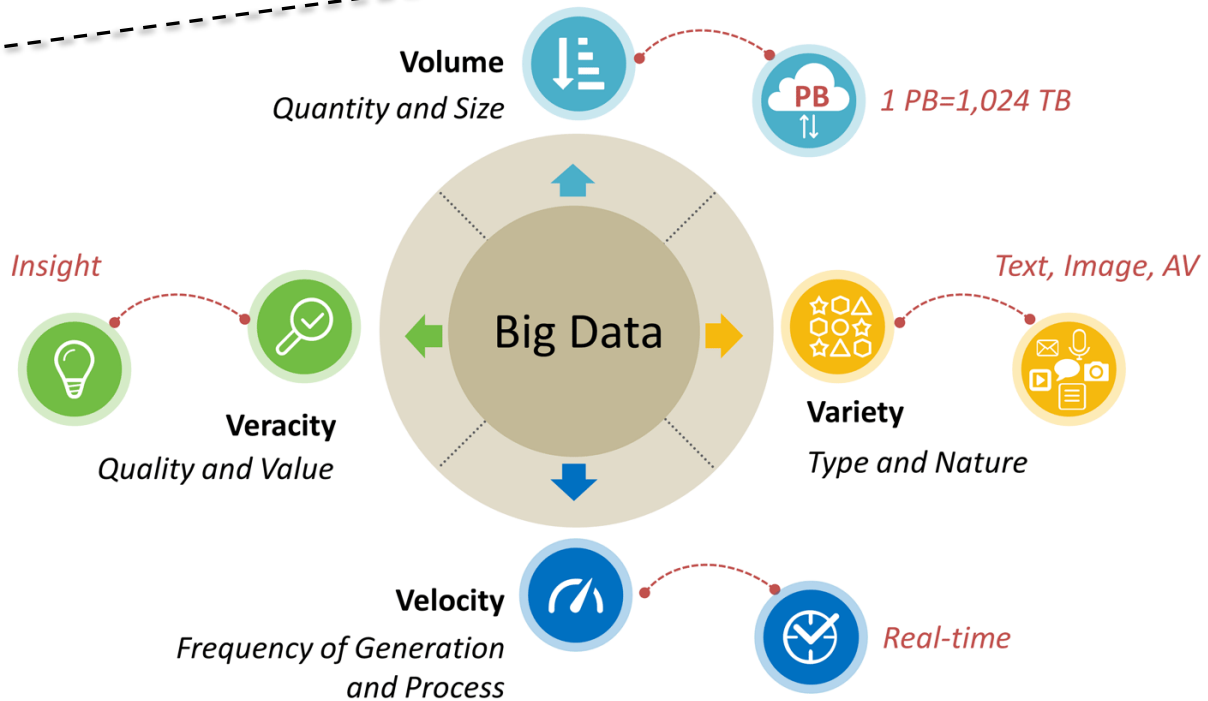
Object-Oriented Database

Department of MANAGEMENT Clemson University

Multidimensional Database – star-schema view



Key Characteristics of Big Data – no predefined data model



Integrated Data Management Framework

Department of MANAGEMENT Clemson University

	Operational	Informational	
	Transactional	Data Warehousing	Big Data
Technology	Relational	Relational	Non-Relational
Modeling	Conceptual Data Modeling (ERD and EER)	Data Warehousing and Integration	Big Data (Hadoop, NoSQL)
Design	Logical Data Modeling (relational tables and normalization)		
Access	SQL		
Data Analysis	Analytics and Its Applications		
Governance and Data Management	Lifecycle, Governance, and Data Quality		