

# AT82.02

DATA MODELING AND MANAGEMENT

---

UNIT 0: INTRODUCTION

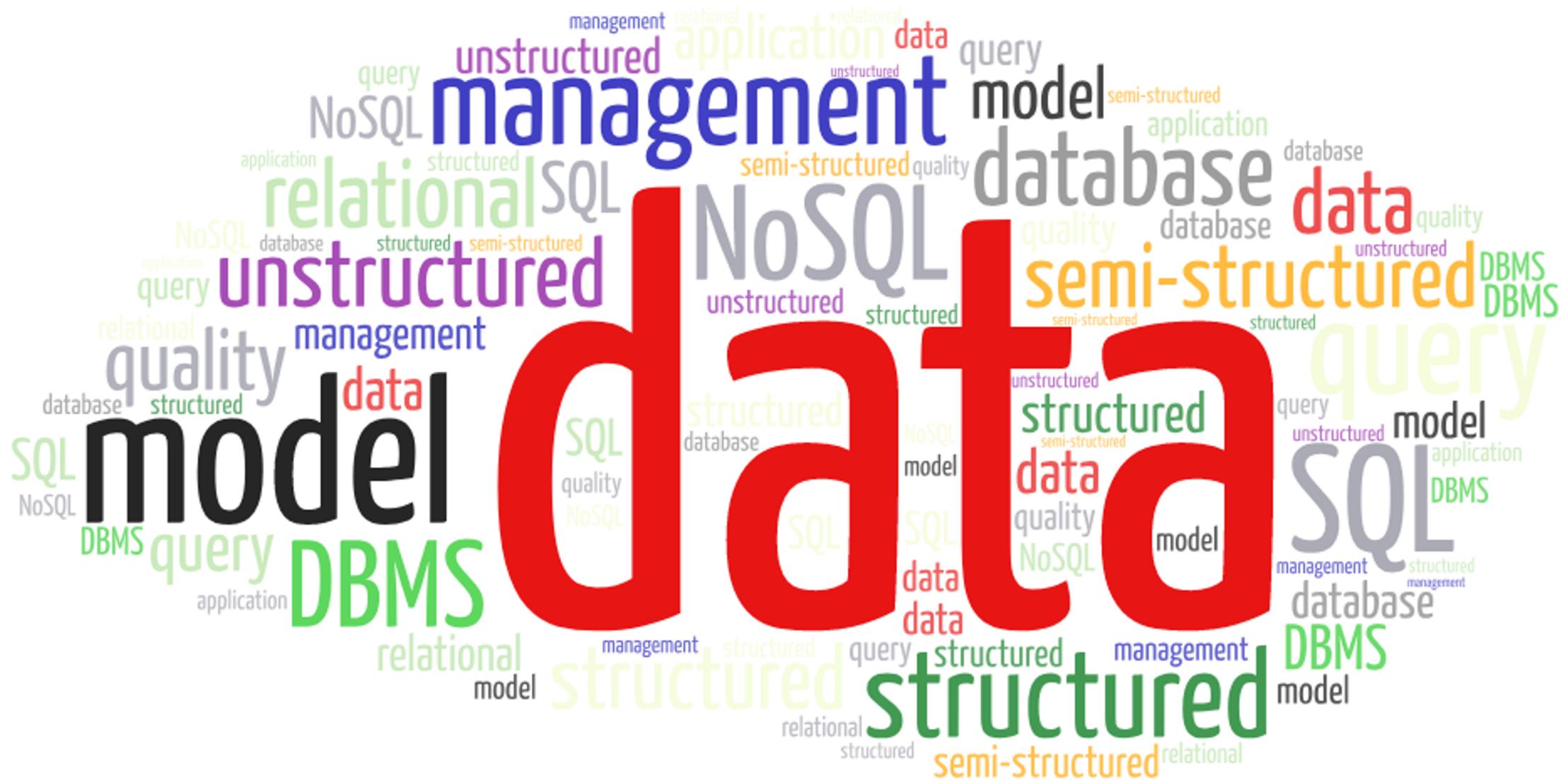
CHUTIPORN ANUTARIYA (CHUTI AT AIT DOT AC DOT TH)



## WARM UP Discussion:

*Data ...*

*What are you thinking of when we  
talk about *data*?*





Let's Discuss:

*Sources of Data ...*

*Where do data come from?*

*Who/what generates data?*

# Three major sources of data...

---

## Machines

Data generated from real time sensors, machines, vehicles, web logs, etc.



## People

Tweets, status updates, social media data, photos, videos



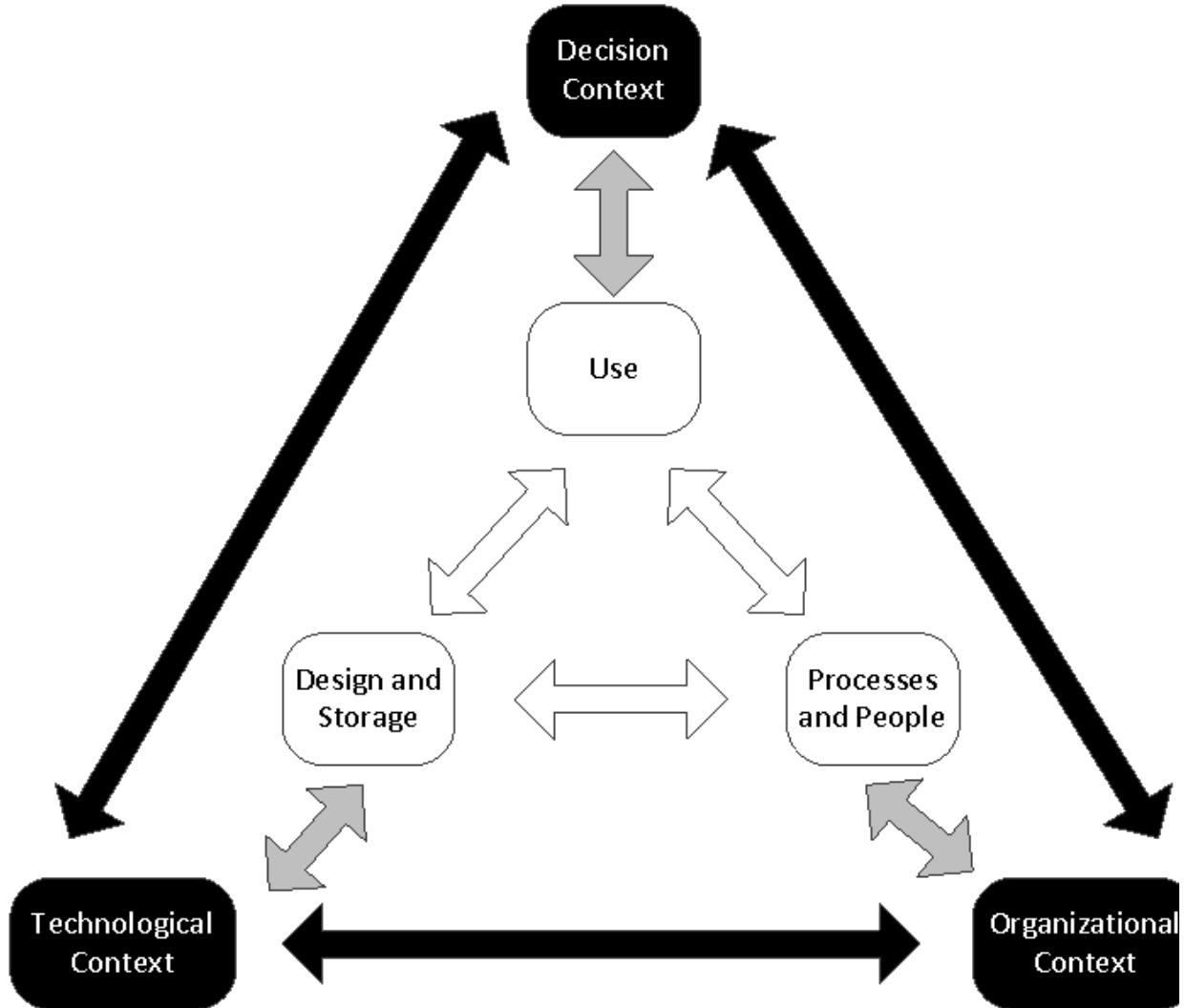
## Organizations

More traditional type of data: transaction information, databases, data warehouses





# *Data Management Perspectives & Contexts*



# Data Management: Perspectives & Contexts

---

# Business-Oriented Perspective

---

- **Decision Context (What?)**
  - How can business goals be quantified? What decisions need to be made?
  - What is the impact of decision on revenue enhancement, cost reduction, and risk management?
  - What needs to be done to enhance the effectiveness of decision making?
- **Technological Context (How? When? Where?)**
  - Does the technology for decision making need to be centralized or decentralized?
  - What technology is required that would support the requisite immediacy in decision making?
  - How do different tools support clarification of assumptions in decision making?
- **Organizational Context (Who?)**
  - Who are the key constituents in decision making?
  - What are the beliefs and culture that can influence decision making?
  - How do regulations influence decision-making?

# Data-Oriented Perspective

- Use
  - Access
    - Operational
    - Reconciled
    - Big Data
  - Analysis
    - Predictive Analysis
    - Forecasting
    - Simulation
    - Optimization
    - Text analytics
    - Web Analysis
    - Social Network Analysis
  - Delivery
    - Dashboard
- Design and Storage
- Processes and People

# Data-Oriented Perspective



Use



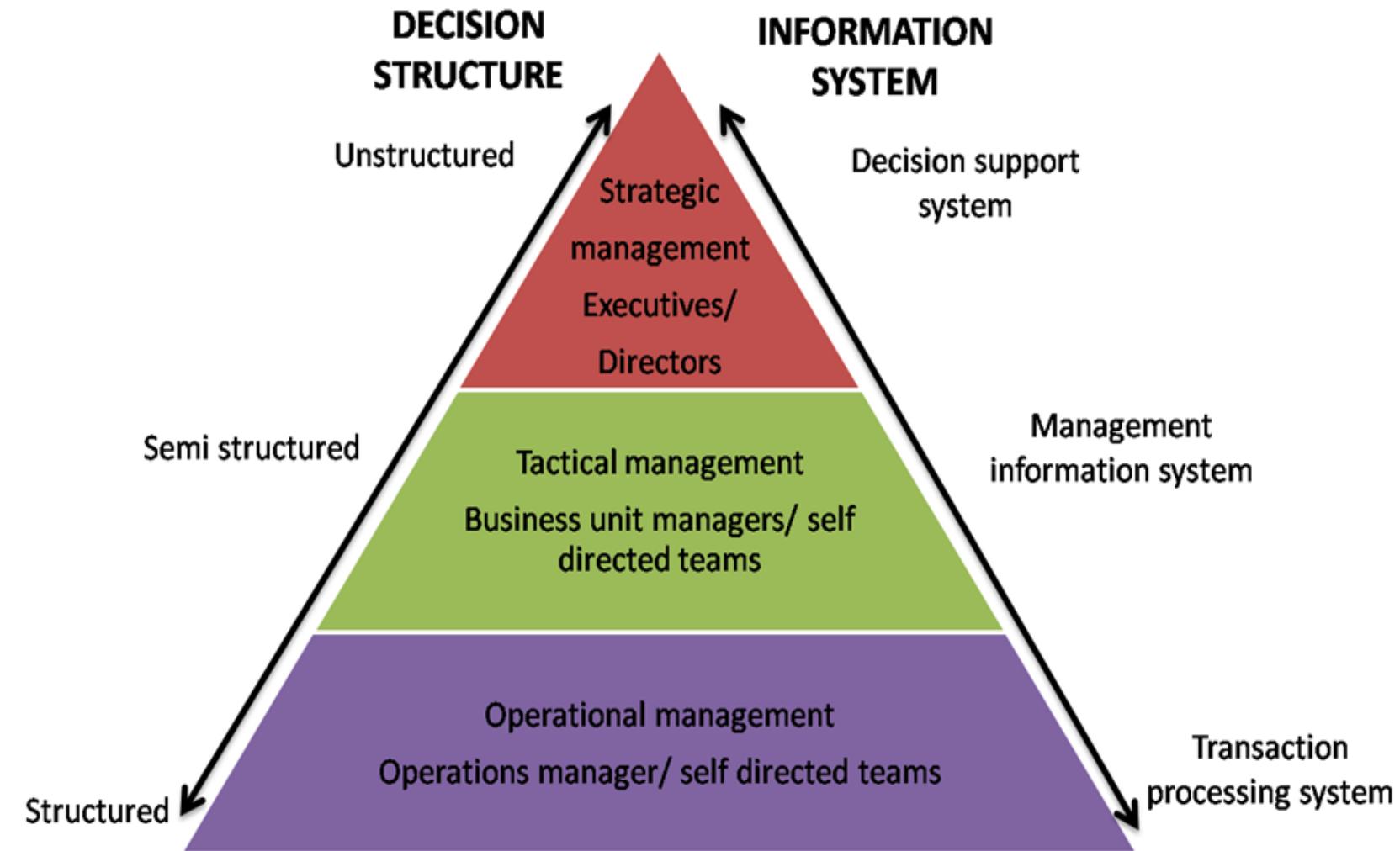
Design and Storage

- o Data Acquisition and Lifecycle Management
- o Data Storage Management
- o Metadata Management
- o Master Data Management
- o Data Quality Management
- o Data Integration



Processes and People

- o Who holds decision rights for various data decision domains? What skills are required to better use data? What skills are the training-related requirements that would enable desirable uses of data?
- o What processes need to be designed to enable desirable uses of data? Who owns the data?
- o What are the acceptable processes of data uses? How to ensure and monitor data-related regulations? How is data inventoried?

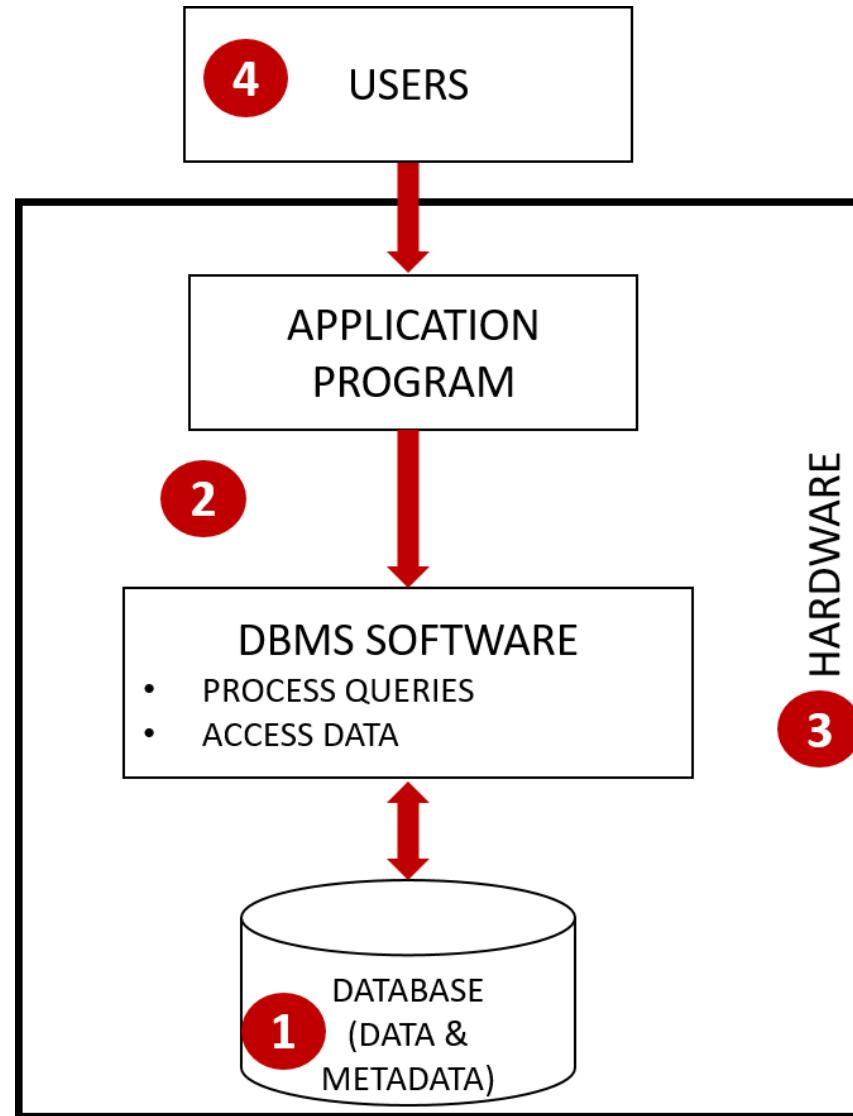


# Decision Making in Enterprises

---

# OLTP: Online Transaction Processing Database System Components

---



# Top Enterprise Relational DBMS

---



SYBASE®

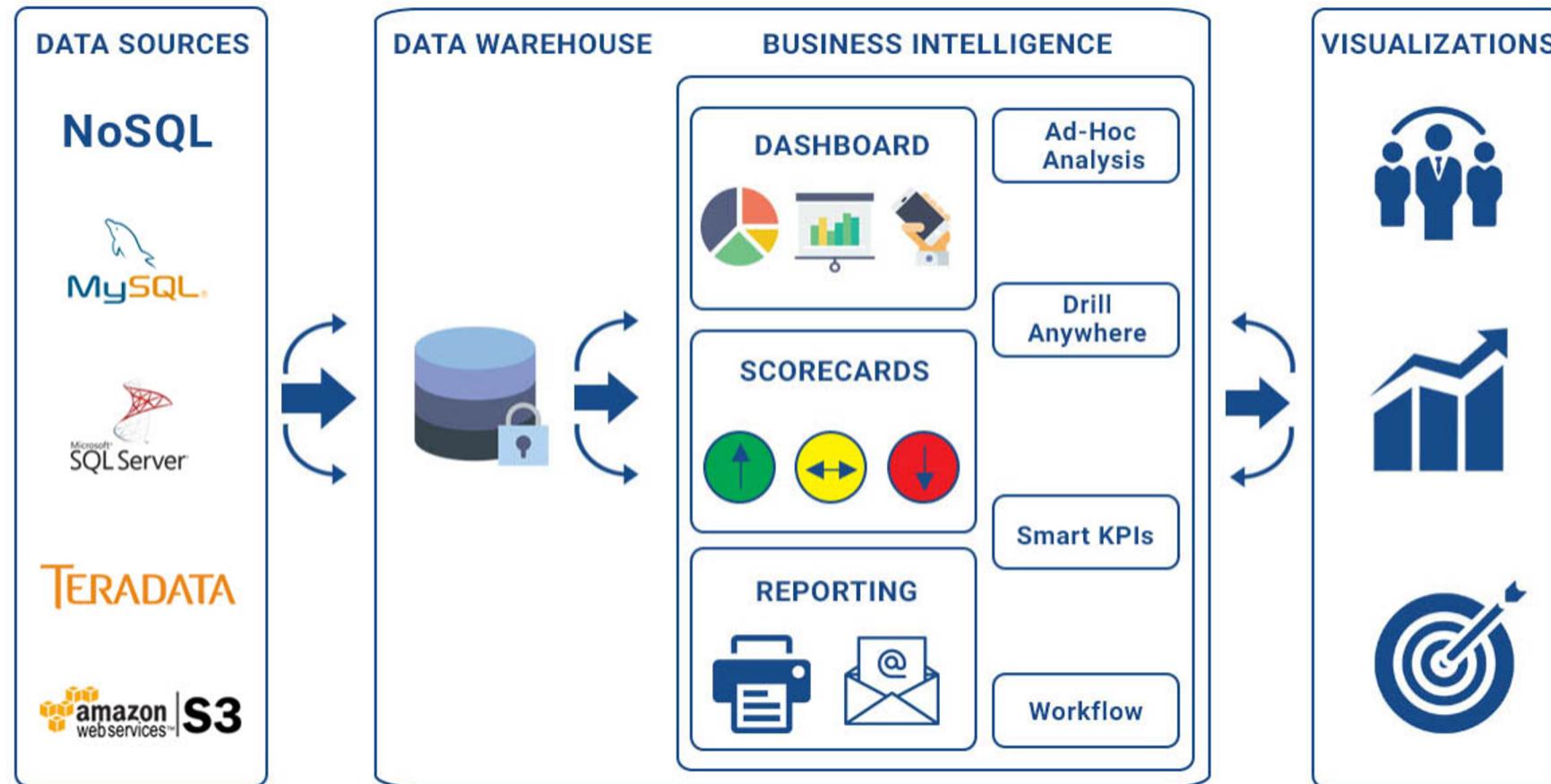


PostgreSQL



ORACLE®

# Business Intelligence Stack





*Let's Discuss:*

***What is a Data Model?***

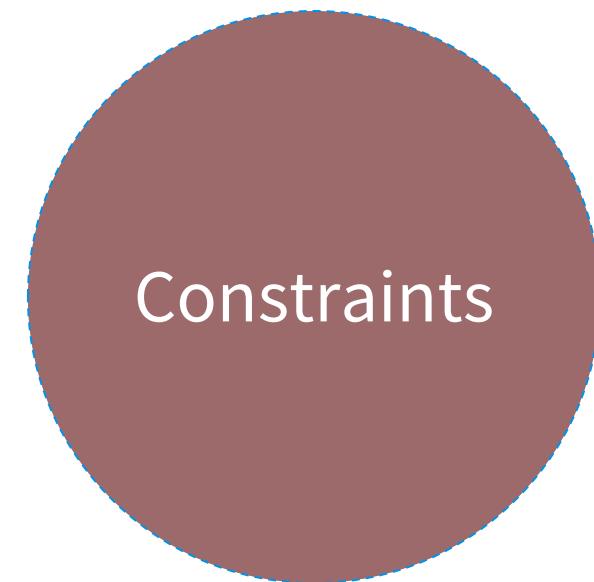
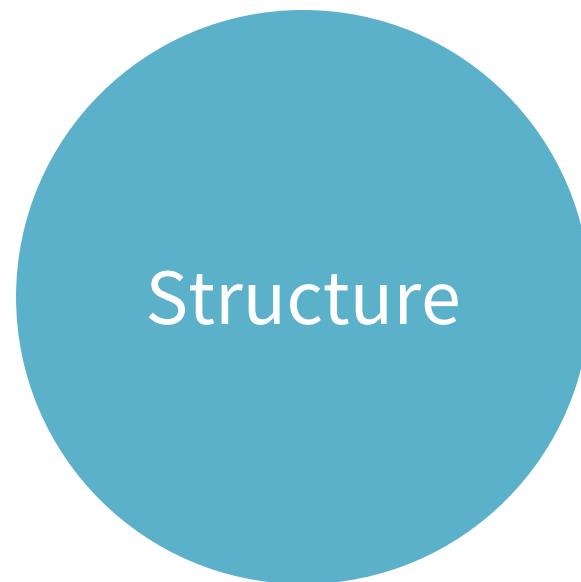
*Why do we care about data  
model?*



*Data Model*  
describes data  
characteristics

# Data Model describes data characteristics

---



# Data Model describes data characteristics

---

Structure

```
Employee {  
    firstName: string,  
    lastName: string,  
    DOB: date  
}
```

Operations

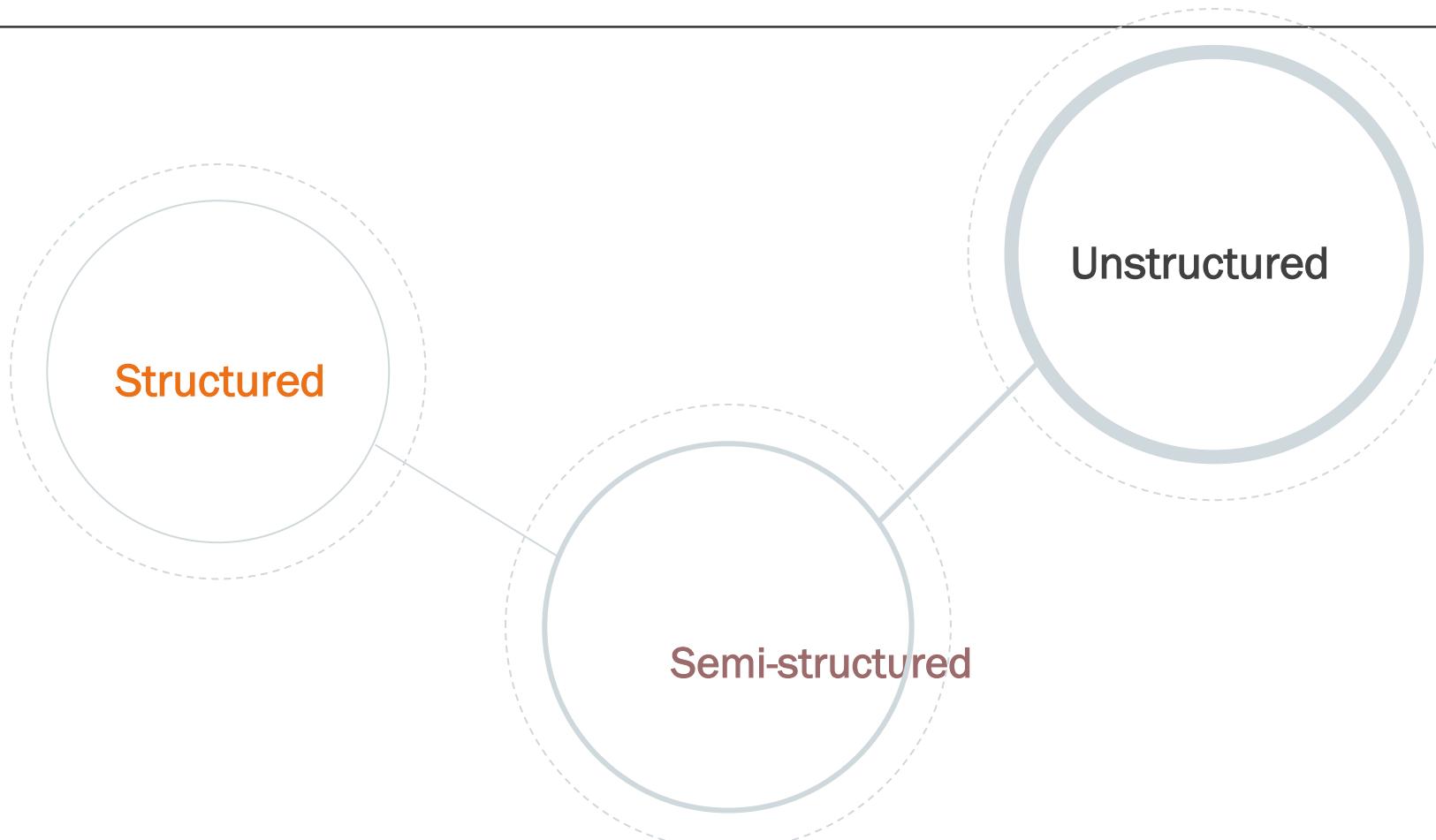
Select employee whose  
DOB is after 1990.

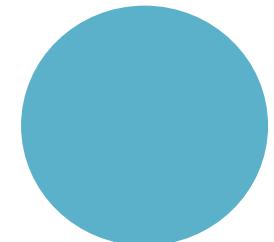
Constraints

Today's date minus DOB  
must be greater than 20.

# Data Model Structures

Structure





Structure

# Data Model Structures

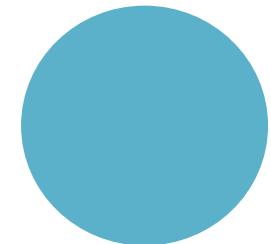
---

Structured

Unstructured

Semi-structured

- ◎ The data that has a structure, is typed and is well organized either in the form of tables or in some other way.
  
- ◎ It has a high level of organization making it predictable, easy to organize and very easily searchable using basic algorithms.



Structure

# Data Model Structures

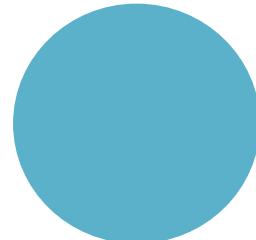
---

Structured

Unstructured

Semi-structured

- ◎ The data that is unstructured or unorganized.
  
- ◎ Operating such type of data becomes difficult and requires advance tools and softwares to access information.



Structure

# Data Model Structures

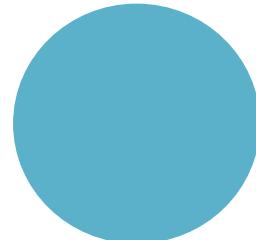
---

Structured

Unstructured

Semi-structured

- ◎ Semi-structured data lies somewhere between the two.
- ◎ It is a form of **structured data** that does not conform with the formal structure of data models associated with **relational databases** or other forms of **data tables**
- ◎ But contains **tags** or other markers to separate semantic elements and enforce hierarchies of records and fields within the data.
- ◎ It is also known as **self-describing** structure.



## Structure

# Data Model Structures

Structured?

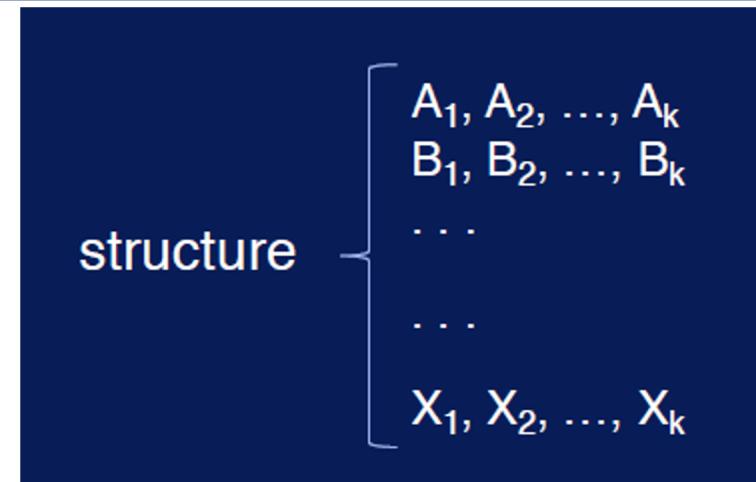
File 1	File 2
(John, Smith, 10-12-1989) (Liz, Spencer, 09-29-1980) (Marie, Bishop, 11-07-1992)	(John, Smith, 10-12-1989, Mechanical, 70000) (Liz, Spencer, 09-29-1980, Electrical, 65000) (Marie, Bishop, 11-07-1992, Driver, ) (Steve, Richards, 04-16-1958, 140000)

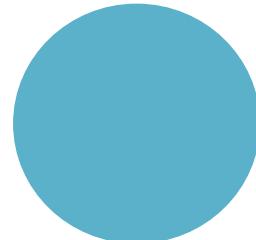
Are the files structured?

Do they have the same structure?

Unstructured?

Semi-structured?





## Structure

# Data Model Structures

Structured?

Unstructured?

Semi-structured?

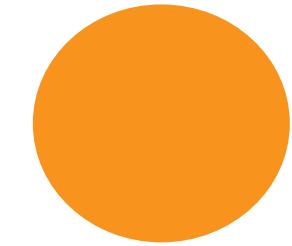
```
&#2453;&#2494;&#2480; &#2453;&#2507;&#2469;&#2527;&#2469;&#2494;&#2453;&#2494;  
&#2441;&#2458;&#2495;&#2468;&#2476;&#2507;&#2461;&#2494;&#2479;&#2494;&#2458;&#2509;&#2459;&#2503;  
&#2472;&#2494; &#2439;&#2470;&#2494;&#2472;&#2496;&#2434;! &#2456;&#2480;&#2503;  
&#2469;&#2494;&#2453;&#2476;&#2503; &#2453;&#2503;  
&#2438;&#2480;&#2476;&#2494;&#2439;&#2480;&#2503;&#2439; &#2476;&#2494; &#2453;&#2503;;  
&#2476;&#2480;&#2509;&#2471;&#2478;&#2494;&#2472;&#2503; &#2453;&#2494;&#2480;  
&#2469;&#2494;&#2453;&#2494; &#2470;&#2480;&#2453;&#2494;&#2480;; &#2453;&#2494;&#2480;  
&#2458;&#2482;&#2503; &#2479;&#2494;&#2451;&#2527;&#2494; &#2470;&#2480;&#2453;&#2494;&#2480;  
&#2478;&#2494;&#2482;&#2470;&#2489; &#2469;&#2503;&#2453;&#2503;&mdash; &#2488;&#2476;  
&#2453;&#2503;&#2478;&#2472; &#2455;&#2497;&#2482;&#2495;&#2527;&#2503;
```

Is the file structured or unstructured?

Compressed data like JPEG images, MP3  
audio files, MPEG3 video files, encrypted data,  
are usually \_\_\_\_\_.

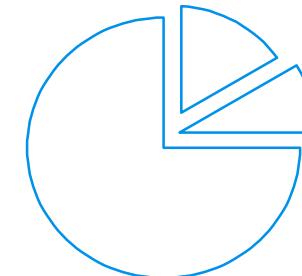
# Data Model Operations

---



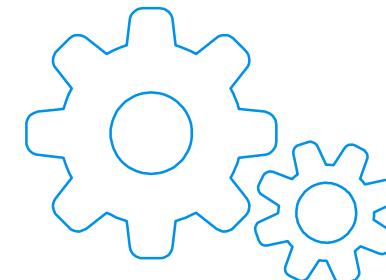
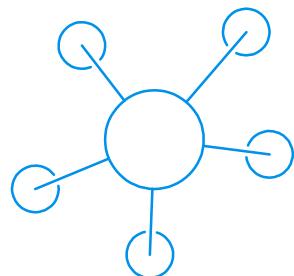
Operations

Subsetting

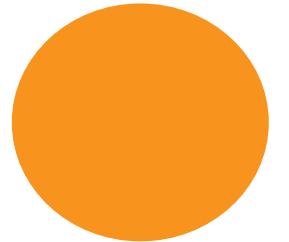


Substructure extraction

Union



Join



Operations

# Data Model Operations

---



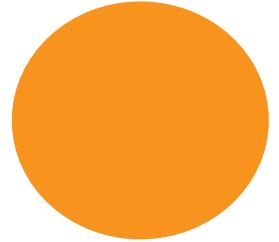
**Subsetting:** Given a collection of data, and a condition, find a subset of data from the collection so that each element in the subset satisfied

(John, Smith, 10-12-1989, Mechanical, 70000)  
(Liz, Spencer, 09-29-1980, Electrical, 65000)  
(Marie, Bishop, 11-07-1992, Driver, )  
(Steve, Richards, 04-16-1958, 140000)



Field#5 < 90000

(John, Smith, 10-12-1989, Mechanical, 70000)  
(Liz, Spencer, 09-29-1980, Electrical, 65000)



Operations

# Data Model Operations



**Substructure extraction:** Given a data collection with some structure, extract from each data item a part of the structure as specified by a condition.

(John, Smith, 10-12-1989, Mechanical, 70000)

(Liz, Spencer, 09-29-1980, Electrical, 65000)

(Marie, Bishop, 11-07-1992, Driver, )

(Steve, Richards, 04-16-1958, 140000)



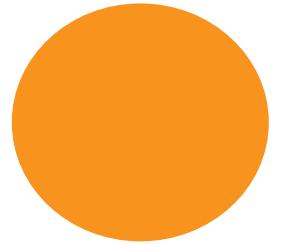
Field#1, Field#2

(John, Smith)

(Liz, Spencer)

(Marie, Bishop)

(Steve, Richards)



Operations

# Data Model Operations

---



**Union:** Given two data collections, create a new one with elements of the two input collections. Duplicate elimination.

(John, Smith, 10-12-1989)  
(Liz, Spencer, 09-29-1980)  
(Marie, Bishop, 11-07-1992)



(John, Smith, 10-12-1989)  
(Liz, Spencer, 09-29-1980)  
(Marie, Bishop, 11-07-1992)  
(Lance, Holt, 04-02-1976)

(Lance, Holt, 04-02-1976)  
(Liz, Spencer, 09-29-1980)

# Data Model Operations

---



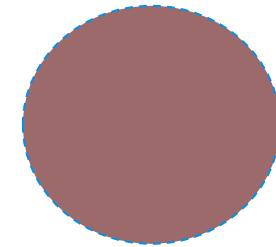
**Join:** Given two data collections, create a new one with elements of the two input collections. Duplicate elimination.

(12, John, Smith, 10-12-1989)  
(14, Liz, Spencer, 09-29-1980)  
(18, Marie, Bishop, 11-07-1992)  
(20, Sue, Daveson, 03-16-1986)



(12, John, Smith, 10-12-1989, Mechanical, 70k)  
(14, Liz, Spencer, 09-29-1980, Electrical, 65k)  
(18, Marie, Bishop, 11-07-1992, Driver, 45k)

(12, Mechanical, 70k)  
(14, Electrical, 65k)  
(18, Driver, 45k)  
(23, Student, 30k)



Constraints

## Data Model Constraints

---

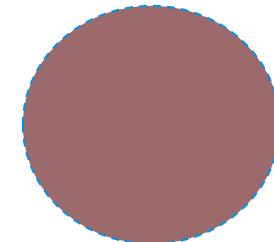
Constraints are logical statements that must hold for data.

A movie has only one title

Different data models have different ways to express constraints

# Data Model Constraints

---



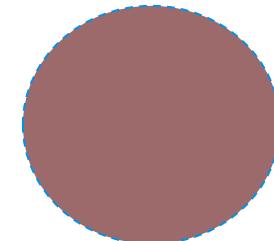
Constraints

## Types of Constraints

- ◎ Value constraint
  - Age is never negative
- ◎ Uniqueness constraint
  - A movie can have only one title
- ◎ Cardinality constraint
  - A person can take between 0 and 3 blood pressure medications at a time

# Data Model Constraints

---



Constraints

## Types of Constraints

- Type constraint
  - Lname:string
- Domain constraint
  - Month in (1 ... 12) or Month in ('Jan', 'Feb', ... 'Dec')
- A structural constraint puts restrictions on the structure of the data rather than the data values themselves.

# Different Kinds of Data Models

---

LIST SOME DATA MODELS YOU ARE AWARE OR HAVE HEARD OF?



# Semi- structured Data Model

---

DATA MODEL BEHIND THE WEB

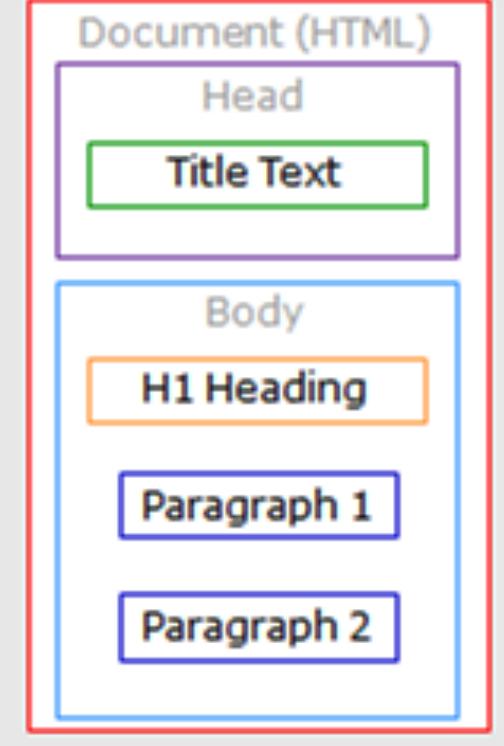
# A Simple HTML Document

---

```
<HTML>
<HEAD>
  <TITLE>Title Text</TITLE>
</HEAD>

<BODY>
  <H1>H1 Heading</H1>
  <P>Paragraph 1</P>
  <P>Paragraph 2</P>
</BODY>

</HTML>
```



# XML: A Generalization of HTML

---

XML Data  
Processing allows  
query related to  
both element  
structures and  
values.

```
<Employee>
  <Name>
    <First>Lassi</First>
    <Last>Lehto</Last>
  </Name>
  <Email>Lassi.Lehto@fgi.fi</Email>
  <Organization>
    <Name>
      Finnish Geodetic Institute
    </Name>
    <Address>
      PO Box 15,
      FIN-02431 Masala
    </Address>
    <Country CountryCode="358">Finland</Country>
  </Organization>
</Employee>
```



An open standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs.

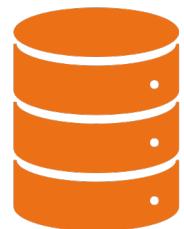


It is used primarily to transmit data between a server and web application, as an alternative to XML.



JSON has been popularized by web services developed utilizing [REST](#) principles.

# JSON: Java Script Object Notation



# JSON: Java Script Object Notation

```
{  
    "Rail Booking": {  
        "reservation": {  
            "ref_no": 1234567,  
            "time_stamp": "2016-06-24T14:26:59.125",  
            "confirmed": true  
        },  
        "train": {  
            "date": "07/04/2016",  
            "time": "09:30",  
            "from": "New York",  
            "to": "Chicago",  
            "seat": "57B"  
        },  
        "passenger": {  
            "name": "John Smith"  
        },  
        "price": 1234.25,  
        "comments": ["Lunch & dinner incl.", "\"Have a nice day!\""]  
    }  
}
```

Key-value Pair

Tuple / Object

Array

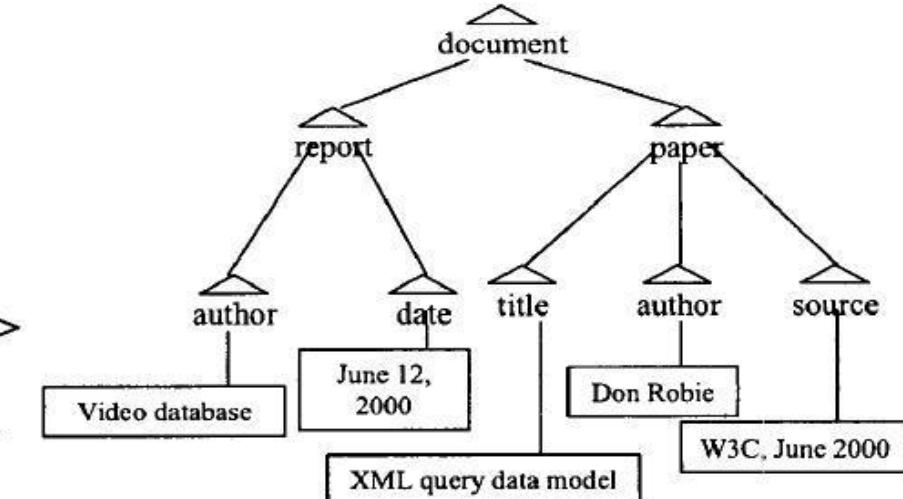


*How to generalize all these forms  
of semi-structured data?*

# Tree Data Structure

A tree has a nested and hierarchical structure representing complex relationships between entities!

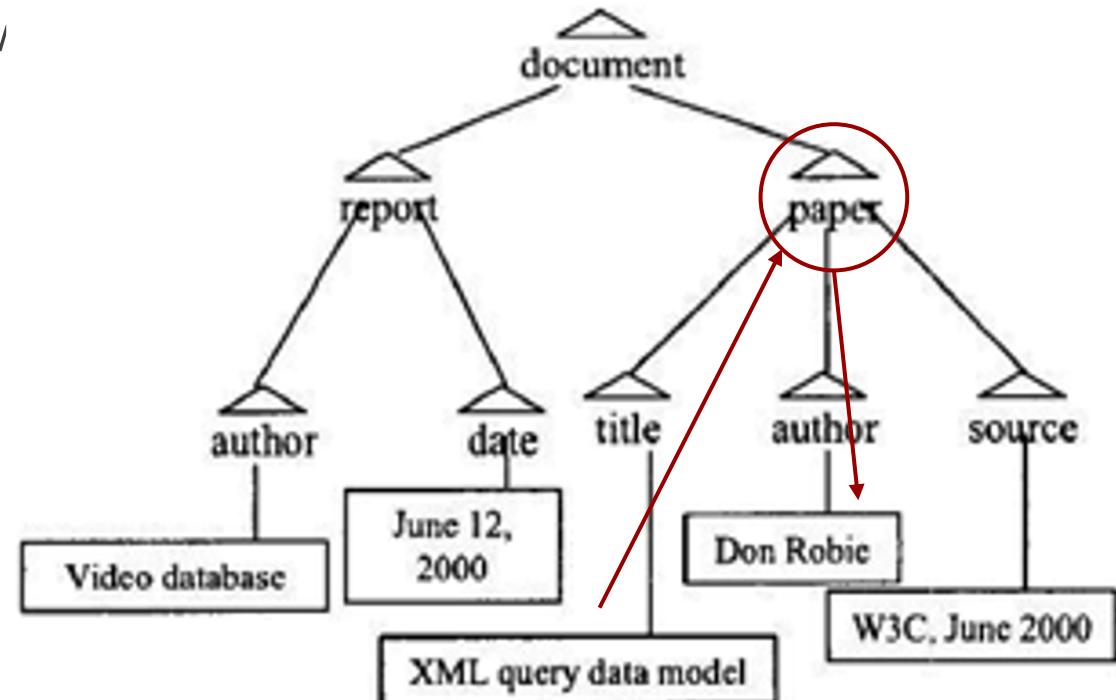
```
<document>
  <report>
    <author>Video database</author >
    <date>June 12, 2000</date>
  </report >
  <paper>
    <title>XML query data model</title>
    <author>Don Robie</author>
    <source>W3C, June 2000</source>
  </paper>
</document>
```



# Tree Data Structure

---

- Tree is well-known data structure and allow navigational access to data
- Paper
  - `getParent` → document
  - `getChildren` → title, author, source
  - `getSibling` → report
- “Video database”
  - Root-to-Node path →  
`document/report/author/“video database”`
- Queries need tree navigation
  - Author of “XML query data model”



# Let's Discuss: Utilization of XML or JSON on the Internet

---



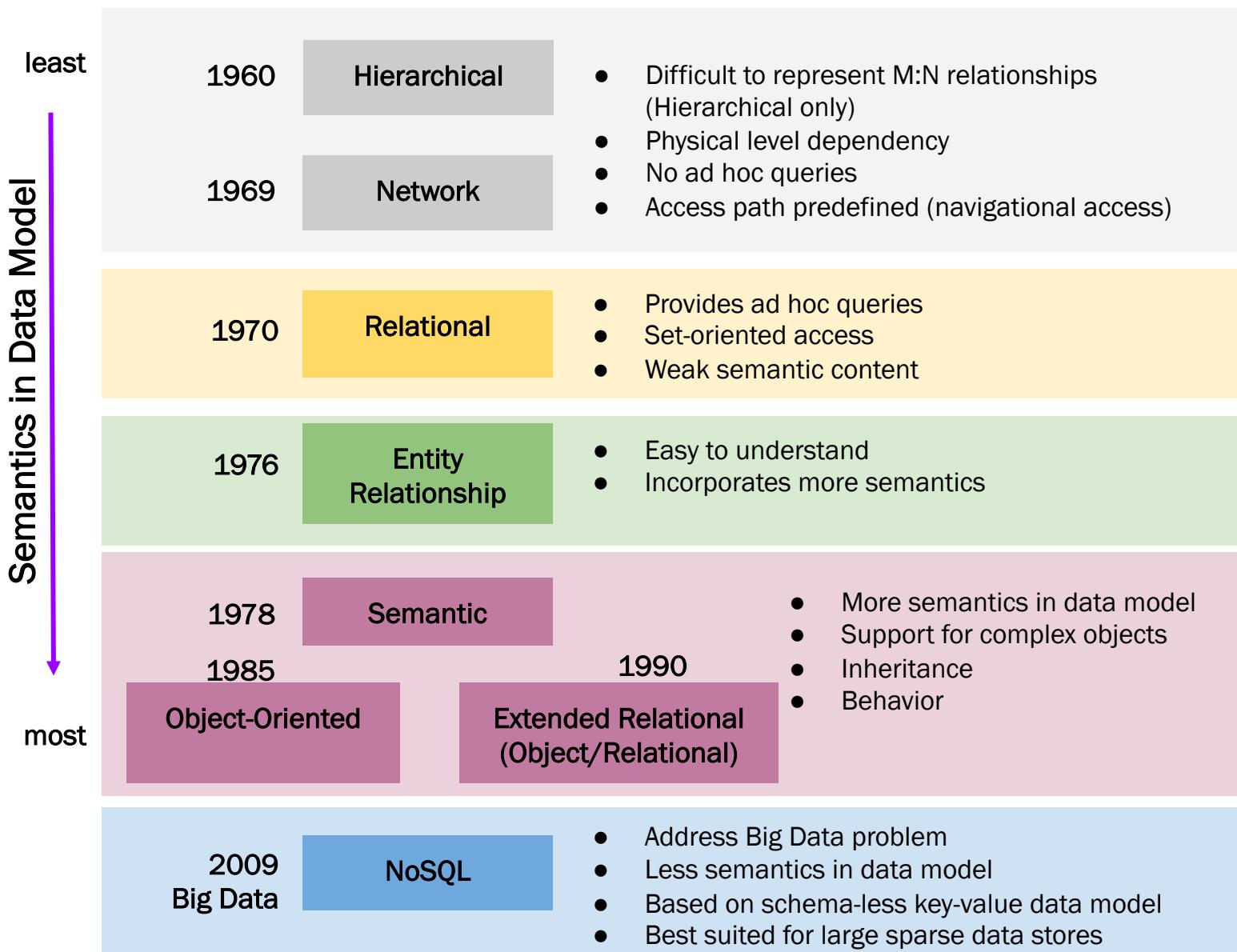
XML (and its lightweight younger sibling JSON) often gets referred to as "the language of the Internet". That means Web applications talk to each other in XML or JSON. For instance, they are used to "markup" vast amount of big data on the Internet for access and retrieval by programs running in applications. They are also used to send messages through which that programs can talk to each other.

An example application would be a program using an XML-based data standard which defines a common way of describing the news items. This application can use XML (or JSON) to gather news data from multiple web sources and aggregate them in a news portal for the readers.

---

Think of other applications that can use XML and discuss it with your fellow class mates.

# Evolution of data models



# **Data Management & Database Technology**

---



# How to Manage Data

# Database



Collection of related data



Data: known facts that can be recorded and that have implicit meaning



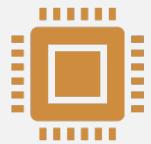
Represents some aspect of the real world



Built for a specific purpose



Crucial for daily operations and decision making in organizations



Database management technology

Major part of software industry  
Revolutionary evolution over 40 years  
Foundation for management of long-term memory of organizations



Vibrant field with employment opportunities

# Importance of Database Technology



Thank you.

**Exit Slip:**

Discuss 3 important things / concepts we have learned today.

---