Class Announcements

- Group database assignment 2 will be (was) out and due in two weeks

- Group homework #3 due on 2/10 through ecommons.

- Please read: Information Technology

  - Midterm: 2/16
Data and information

by

David G. Messerschmitt
Copyright notice

©Copyright David G. Messerschmitt, 2000. This material may be used, copied, and distributed freely for educational purposes as long as this copyright notice remains attached. It cannot be used for any commercial purpose without the written permission of the author.
Key concept

- The computer is the first machine whose functionality is not determined at the time of manufacture
  - added by software later
- Any form of information can be represented or approximated within the information technology infrastructure
- The key commodity manipulated by information technology is information
- To be manipulated in a computing/networking environment, information must be represented by data
Key concept

- The computer is the first machine whose functionality is not determined at the time of manufacture
  - added by software later
- Any form of information can be represented or approximated within the information technology infrastructure
- The key commodity manipulated by information technology is information
- To be manipulated in a computing/networking environment, information must be represented by data
The computer is the first machine whose functionality is not determined at the time of manufacture, but added by software later.

Any form of information can be represented or approximated within the information technology infrastructure.

The key commodity manipulated by information technology is information.

To be manipulated in a computing/networking environment, information must be represented by data.

Information economy: creation, capture, storage, retrieval, manipulation of information.
Key concept

- The computer is the first machine whose functionality is not determined at the time of manufacture but added by software later.
- Any form of information can be represented or approximated within the information technology infrastructure.
- The key commodity manipulated by information technology is information.
- To be manipulated in a computing/networking environment, information must be represented by data.

Information economy: creation, capture, storage, retrieval, manipulation of information.

What is information?
Information

• From a user (human) perspective...
  ....recognizable patterns that influence you in some way (perspective, understanding, behavior...), e.g., stock prices, rain, music, others?

• Includes text, numbers, other media, e.g., art, video, games, many other goods.

• In the computing infrastructure, information has a somewhat different connotation as “structure” and “interpretation” added to data.
Data: Bits as a building block

A bit is “0” or “1” — the atom of the information economy

Data is a collection of bits, like

- “0101110111010110”
- “0000011”
- “11101110101101011011011011010”
- “101111” → 47

Note: the terms “data” and “information” are not always used consistently! But, the distinction of them with “knowledge” is clear.
Information is representation of data

- Not all information in physical world is composed by bits, e.g., radio wave, others?
- Possible to represent any info (audio) by bits
- Take the place of the original, equivalent to, in the sense that the original can be reconstructed from its representation
- Often the original can only be approximately reconstructed, although it may be indistinguishable to the user
  - e.g. audio or video
### ASCII

<table>
<thead>
<tr>
<th>Alphabet</th>
<th>Hex</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7&gt;</td>
<td>/x37</td>
<td>00110111</td>
</tr>
<tr>
<td>&lt;8&gt;</td>
<td>/x38</td>
<td>00111000</td>
</tr>
<tr>
<td>&lt;9&gt;</td>
<td>/x39</td>
<td>00111001</td>
</tr>
<tr>
<td>&lt;&gt;</td>
<td>/x3A</td>
<td>00111010</td>
</tr>
<tr>
<td>&lt;=</td>
<td>/x3B</td>
<td>00111011</td>
</tr>
<tr>
<td>&lt;&lt;=</td>
<td>/x3C</td>
<td>00111100</td>
</tr>
<tr>
<td>=&gt;</td>
<td>/x3D</td>
<td>00111101</td>
</tr>
<tr>
<td>&lt;=&gt;</td>
<td>/x3E</td>
<td>00111110</td>
</tr>
<tr>
<td>?&gt;</td>
<td>/x3F</td>
<td>00111111</td>
</tr>
<tr>
<td>&lt;At&gt;</td>
<td>/x40</td>
<td>01000000</td>
</tr>
<tr>
<td>&lt;A&gt;</td>
<td>/x41</td>
<td>01000010</td>
</tr>
<tr>
<td>&lt;B&gt;</td>
<td>/x42</td>
<td>01000010</td>
</tr>
<tr>
<td>&lt;C&gt;</td>
<td>/x43</td>
<td>01000011</td>
</tr>
<tr>
<td>&lt;D&gt;</td>
<td>/x44</td>
<td>01000100</td>
</tr>
</tbody>
</table>

Note that this representation is **NOT** unique...: “structure” & “interpretation”, i.e., “rule”, e.g., 1, 2, 3, 4, ...

....this one happens to be a **standard**

(ANSI X3.110-1983)

**Structure:** *how the bits are arranged*

**Interpretation:** *significance of the bits to the users or application interpreting them as information*
A picture

This picture conveys information

This information is represented in this computer, but how?
Expanding a small portion of the picture, we see that it is represented by square pixels....

300 tall by 200 wide....

.... each pixel represented by 8 bits

....with a range of 256 intensities per pixel (why 256?)

300 • 200 • 8 bits = 480,000 bits (but it can be compressed)
The fact that all data can represent all types of information makes possible for computers, equipment, software manipulate them.

A color picture can be represented by three monochrome images...

At the expense of three times as many bits
If particular data is to represent information, the information can be recovered through a procedure called “data processing”, with presumed structure & interpretation.
Representation needs to be standardized

User 1
Information

If the representation is NOT standardized, the information is garbled!

Communicate data to another user or organization

User 2
Information

Data

The representation of information is stored in memory as data; The structure & interpretation is embodied in “computer program.”

Slide adapted from slides for Understanding Networked Applications
By David G Messerschmitt. Copyright 2000. See copyright notice
Regeneration

- Make a precise copy of the data (copy bit by bit)
- If you know the representation, this is equivalent to making a precise copy of the information
- Each such precise copy is called a generation
- process is called regeneration
Replication of information

Anything that can be regenerated can be replicated any number of times.

This is a blessing and a curse.
Analog information cannot be regenerated

Analog information can be copied, but not regenerated

We will never know exactly what the original of this Rembrandt looked like
Discrete information can be regenerated

Regeneration can preserve data (but not its original physical form, i.e., paper)

Regeneration is possible for information represented digitally (which is tolerant of physical deterioration)

\[
0 + \text{noise} \rightarrow 0 \\
1 + \text{noise} \rightarrow 1
\]
Replication of information requires knowledge of representation.

1. Replication preserves the integrity of the data, but that is not sufficient.

2. Replication of information also presumes knowledge of its representation.

Every .xxx DOS file is a representation.
Implications

Digitally represented information can be preserved over time or distance in its precise original form by occasional regeneration

- digital library
- digital telephony

Replication of data is easy and cheap
Implications (con’t)

- Replication of information requires knowledge of the structure and interpretation
  - Standardization or some other means
- Extreme supply economies of scale
- You can give away or sell and still retain
- Unauthorized replication or piracy relatively easy
Building Blocks of Network Computing

- **User**, person who interacts with and derives benefit from a networked computing application.
- **Desktop computer or host**, a computer directly accessed by users, acting as interface between users and application.
- **Server**, not directly associated with users, usually missing display, keyboard.
- **Application software**, programs running on the clients for specific function.
- **Data**, collection of bits representation information manipulated by application.
- **Infrastructure software**, program running on computers serving many common needs.
- **Network**, communicating infrastructure connecting computer running applications, e.g., LAN, WAN.

None of those is particular useful just by itself.
Architecture

by

David G. Messerschmitt
©Copyright David G. Messerschmitt, 2000. This material may be used, copied, and distributed freely for educational purposes as long as this copyright notice remains attached. It cannot be used for any commercial purpose without the written permission of the author.
Why Architecture?

Computing system is complex and requires an internal organization or system to work properly.

A **system** is something that puts together building blocks (e.g., users, hosts, servers, etc.) to interact to accomplish higher-level purpose that each block cannot accomplish individually.

A **subsystem** is an element within the system that performs some narrow-defined function on behalf of the system and cannot subdivided.

Purpose: subsystems can be dealt with as individual units, independently from others.
Why Architecture?

Computing system is complex and requires an internal organization or system to work properly.

A **system** is something that puts together building blocks (e.g., users, hosts, servers, etc.) to interact to accomplish higher-level purpose that each block cannot accomplish individually.

A **subsystem** is an element within the system that performs some narrow-defined function on behalf of the system and cannot subdivided.

Purpose: subsystems can be dealt with as individual units, independently from others.

Any examples of “system” and “subsystem”? 
Properties of System Architecture?

**Decomposition**: participation of the system into small individual systems that interact to realize the higher purpose of the system, e.g., government: executive, legislative and judicial branches;

**Functionality**: specialized capabilities assigned to each subsystem supporting overall system purposes, e.g., legislature makes laws, etc.

**Interaction**: How the subsystems communicate and cooperate to support the system purposes, e.g., executive informs the legislature of the need of new laws
Clients: Make request, server satisfies those requests
Some building blocks

Which of these can be subsystems?
What is Architecture?

How do you architect a solution?
A system is decomposed into interacting subsystems.

Each subsystem may have a similar internal decomposition.
Three elements of architecture

Decomposition
Organization
Functionality
Responsibility
Interaction
Cooperation

Slide adapted from slides for Understanding Networked Applications
By David G Messerschmitt. Copyright 2000. See copyright notice
Three elements of architecture

Any examples of “architecture”?

Slide adapted from slides for *Understanding Networked Applications*
By David G Messerschmitt. Copyright 2000. See copyright notice
System examples

Let’s quickly look at some system decomposition examples
- Quick tour of information technology systems
Evolution of client/server computing: Time sharing

Discussions:
1) Laundry (punched cards) vs Laundromat (time-sharing)
2) Wired vs LAN (Local Area Network)
Two-tier client/server

Local-area network

Server/Mainframe
Three-tier client/server

1) Distributed applications to access to databases
2) Enterprise database server (mainframe)
3) Application server: could be in database server, but different administrative domains, limited access
Inter-organizational computing
Consumer access

Grant all the consumers have access to application Servers within enterprises; but serious security implications.
System integration

Architecture

-> subsystem implementation

-> system integration

Bring together subsystems and make them cooperate properly to achieve desired system functionality

- Always requires testing
- May require modifications to architecture and/or subsystem implementation
**Emergence**

Subsystems are more specialized and simpler functionality

Higher-level system functionality arises from the interaction of subsystems

Higher-order behavior that emerges because the subsystems are composed is called “emergence.”

**Emergence** includes capabilities that arise purely from that interaction (desired or not)

- e.g. airplane flies, but subsystems can’t
Why system decomposition?

- Divide and conquer approach to containing complexity
- Reuse
- Consonant with industry structure (unless system is to be supplied by one company)
- Others?
Networked computing infrastructure

by

David G. Messerschmitt
Major subsystems

Presentation

Application software

Logic Data

Infrastructure software

Infrastructure equipment

Client host

Network

Server host

Infrastructure is decomposed into subsystem with common functionality to minimize cost
A network computing infrastructure supports:
1) Communication across distance (through network)
2) Communication over time (through storage)
3) Computation and logic
4) Human-computer interface (called presentation), e.g., keyboard, mouse, microphone, etc.
Layering builds capability incrementally by adding to what exists.
Example of Layering: networking

- Application
  - Messages
- Transport
  - Packets
- Network
  - Frames
- Link
  - Bits
- Physical
  - Signals
Messages and packets

- Simplest network communication service is the message
  - Smallest unit of communicated data meaningful to application
  - Size, but unknown structure and interpretation
  - Analogous to file in storage
- Internally, the network may fragment a message into packets, and reassemble those packets back into a message
Software Layering

Application

Middleware

Operating System
## Simplified infrastructure layering

<table>
<thead>
<tr>
<th>Application</th>
<th>Middleware</th>
<th>Operating system</th>
<th>Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributed object management</td>
<td>Database management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network software</td>
<td>File system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network equipment</td>
<td>Storage peripherals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communications</td>
<td>Storage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Operating system functions

- Graphical user interface (client only)
- Hide details of equipment from the application
- Multitasking
- Resource management
  - Processing, memory, storage, etc
- etc
Middleware Functions

- Between application and the application.
- Capabilities that can be shared by many applications, but that is not part of OS
  - Example: Database Management System (DBMS)
- Hide details of OS from application
  - Java Virtual Machine
- More purposes we’ll talk about later.
Network equipment (a subsystem), including links, switches, backbone links

- Hosts
- Switches (linked to backbone not to hosts)
- Backbone network: backbone links + switch
  - Switches (linked to backbone not directly link to hosts)
  - Backbone Links (connected switch, not directly link to hosts)
  - Access switch
  - Access links, e.g., drive way
- At least one feasible path from one host to any other hosts!
File system

- Hides details of storage equipment from applications
- File is:
  - Unit of data managed for the benefit of the application
  - Name
  - Location in naming hierarchy
What’s a database?

**Database**

- File with specified structure
- Example: relational table
<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Accommodation</th>
<th>Tourists</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Oakley</td>
<td>Bed&amp;Breakfast</td>
<td>14</td>
</tr>
<tr>
<td>2002</td>
<td>Oakley</td>
<td>Resort</td>
<td>190</td>
</tr>
<tr>
<td>2002</td>
<td>Oakland</td>
<td>Bed&amp;Breakfast</td>
<td>340</td>
</tr>
<tr>
<td>2002</td>
<td>Oakland</td>
<td>Resort</td>
<td>230</td>
</tr>
<tr>
<td>2002</td>
<td>Berkeley</td>
<td>Camping</td>
<td>120000</td>
</tr>
<tr>
<td>2002</td>
<td>Berkeley</td>
<td>Bed&amp;Breakfast</td>
<td>3450</td>
</tr>
<tr>
<td>2002</td>
<td>Berkeley</td>
<td>Resort</td>
<td>390800</td>
</tr>
<tr>
<td>2002</td>
<td>Albany</td>
<td>Camping</td>
<td>8790</td>
</tr>
<tr>
<td>2002</td>
<td>Albany</td>
<td>Bed&amp;Breakfast</td>
<td>3240</td>
</tr>
<tr>
<td>2003</td>
<td>Oakley</td>
<td>Bed&amp;Breakfast</td>
<td>55</td>
</tr>
<tr>
<td>2003</td>
<td>Oakley</td>
<td>Resort</td>
<td>320</td>
</tr>
<tr>
<td>2003</td>
<td>Oakland</td>
<td>Bed&amp;Breakfast</td>
<td>280</td>
</tr>
<tr>
<td>2003</td>
<td>Oakland</td>
<td>Resort</td>
<td>210</td>
</tr>
<tr>
<td>2003</td>
<td>Berkeley</td>
<td>Camping</td>
<td>115800</td>
</tr>
<tr>
<td>2003</td>
<td>Berkeley</td>
<td>Bed&amp;Breakfast</td>
<td>4560</td>
</tr>
<tr>
<td>2003</td>
<td>Berkeley</td>
<td>Resort</td>
<td>419000</td>
</tr>
<tr>
<td>2003</td>
<td>Albany</td>
<td>Camping</td>
<td>7650</td>
</tr>
<tr>
<td>2003</td>
<td>Albany</td>
<td>Bed&amp;Breakfast</td>
<td>6750</td>
</tr>
</tbody>
</table>
Storage Middleware example: DBMS

- **Database Management System (DBMS)**
  - Manage Multiple databases
  - Allow multiple applications to access common databases
  - Implement standard data “lookup” (query) functions.
What is the Internet

- Internet = the major global internet
- An internet is a “network of networks”
  - Interconnect standard for LAN’s and WAN’s
- A private internet is called an intranet
- An extranet is an interconnection of intranets through the Internet
Intranet

- Private internet
- May be connected to Internet
  - Firewall creates a protected enclave
Extranet

- Intranets connected through an unprotected domain (typically the Internet)
- Encryption and other security technologies used to
  - protect proprietary information
  - prevent imposters, vandals, etc
Extranet

Consumers, field workers, etc.

Internet

Intranet