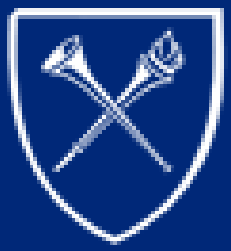


EMORY UNIVERSITY

Clinical Informatics Academy

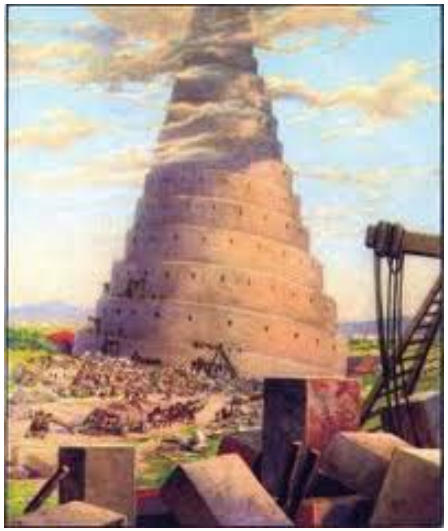
Overview of Computer Hardware: Introduction to Basic Terms & Concepts

Fusheng Wang PhD
David A Gutman MD PHD

Goals of Lecture

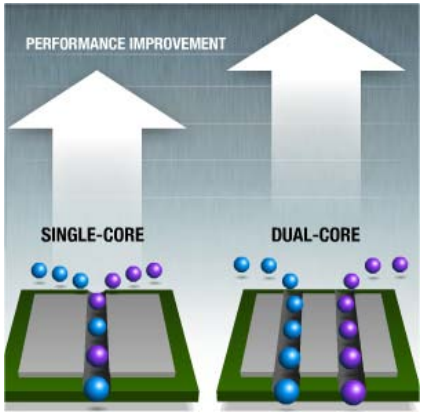
- **Provide very brief overview of certain terms and concepts that may be used throughout the day**

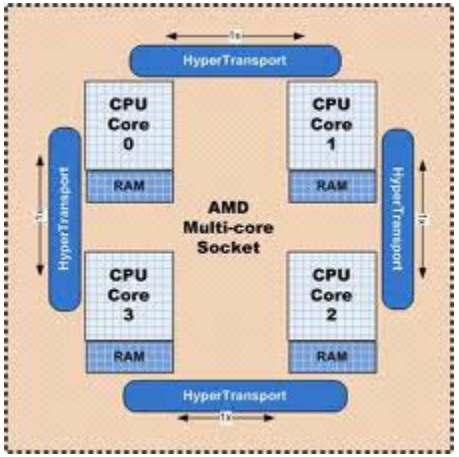
- **Begin our ascent up the tower of Babel of Informatics**



CPUs

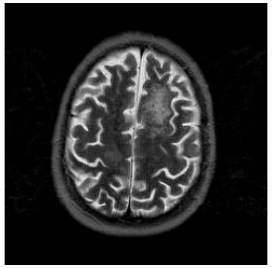
- Serves as the basic computation engine (brain) for the machine
- Intel and AMD make processors used in most healthcare applications
- Many machines now have “multi-core”, meaning a single chip contains several individual CPUs
- More cores != faster performance unless underlying programs can work in parallel





Serial vs Parallel Operations

- Have a 1 million line document and want to count # of times “diabetes” appear
- Can start from the top…. and have the computer keep going (serial operation)
- Parallel version: Split document into equal parts, and have each “core” or “node” count up their part and assemble the results
- Some problems are easier to parallelize than others e.g. run same analysis on different patients



GPU

- Graphics Processing Unit
- Specialized piece of hardware that is very good at processing certain types (generally 3D image) data
- Examples in clinical practice would be GPU-accelerated DICOM workstation for rendering/viewing 3-D image reconstructions

Networking

- Computers are now networked together to allow sharing of resources (printers/disks/etc)
- Wired vs wireless
- Networks can span different scopes: home → office → department → University (intranet) → the world (Internet)
- Amount of information that can be transmitted (bandwidth) can vary dramatically
- Slower/congested networks = longer times to load files → particularly noticeable when transferring imaging data (e.g. radiology)

RAM Memory vs Hard Drive

- Random Access Memory(RAM)- Serves as short term storage for calculations & programs
- Most RAM is volatile– turning off computer deletes this data
- More RAM generally means better performance, especially with large data sets/many programs open
- Hard Disk– generally spinning platters that store information at high density– non-volatile
- Access time for RAM is generally 10x to 100x's of times faster than corresponding disk access… allowing processors to not "wait" for information

vs

vs

Storage Types & Networking

- Local vs network storage
- Local storage is physically part of the computer– generally faster than network storage
- Network/shared storage: Use existing network infrastructure and files “live” in a different physical location but can be directly accessed
- Has many advantages as backups/redundancy can be engineered as part of the system
- Disadvantage is that slow networks and/or network outages = no access to files

Operating Systems

- Software framework that manages/controls basic operations on a given CPU allowing communication between users and underlying hardware
- Examples: Windows 7, Mac OSX, Unix/Linux and derivatives
- Programs written for a given OS generally can only be run on that platform
- Linux/Unix is sometimes used for specific high-end uses (like Radiology workstation)

DISK Storage

- Serve as main storage for files/programs/pictures/images/etc
- Much slower than memory
- New “FLASH” drives use non-volatile RAM similar to what’s used in a camera and can be used to store frequently used data and allow quicker access than conventional spinning disks

Network

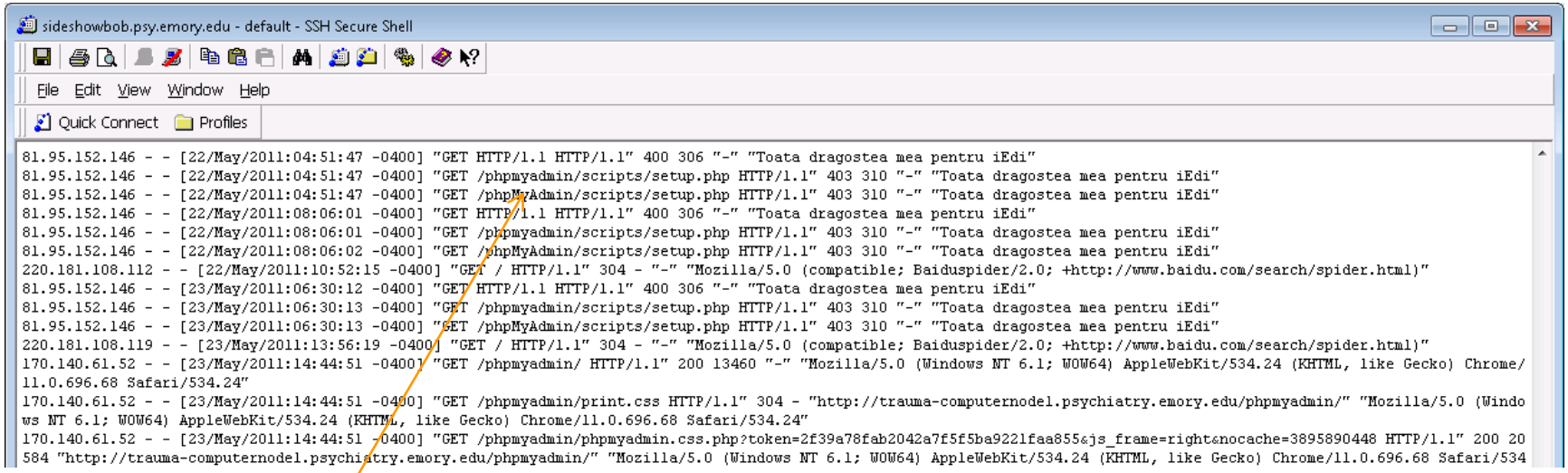
- General infrastructure that allows computers to send information to and from each other
- Can have different scopes/purposes
- Bluetooth network: Short range communication
- Intranet:

Firewalls

"A **firewall** is a device or set of devices designed to permit or deny network transmissions based upon a set of rules and is frequently used to protect networks from unauthorized access while permitting legitimate communications to pass"

- Can be lax to very restrictive (block access to the entire web)
- Mandatory in clinical settings to protect patient data
- Can allow access to certain resources only at specific locations

Why firewalls can be important!



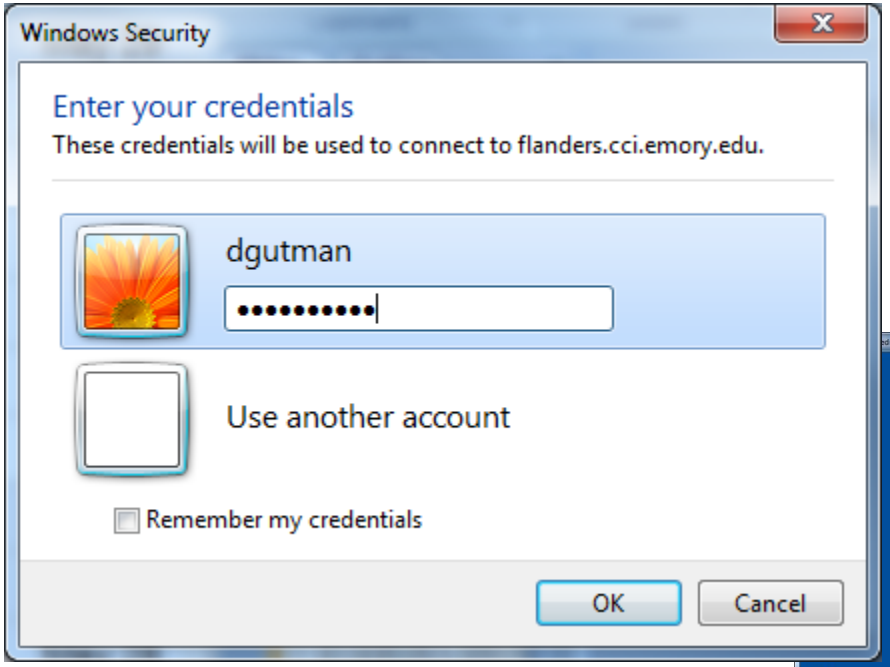
Looks like someone from Europe is trying to access one of my servers by doing a "scan"

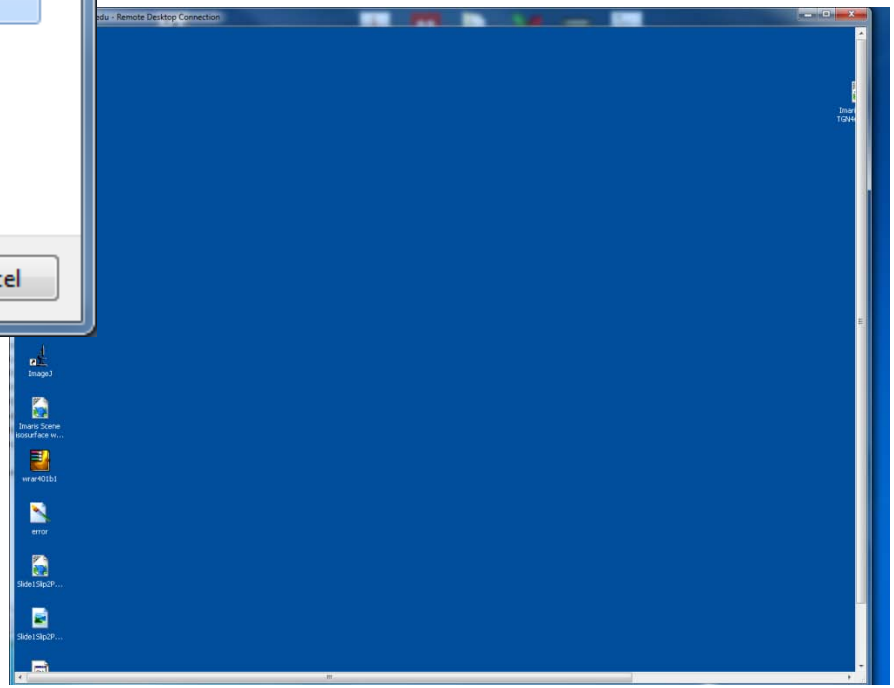
Virtualization / Virtual Machines

- Separates the applications/operating system from the underlying hardware and creates a “virtual” copy
- This virtual machine can be transferred to any computer/hardware that can host the image in case of hardware failure
- Copies/snapshots can be made of the image to facilitate backup/rollbacks/testing
- Allows pooling of resources– a single machine can host several virtual hosts
- Performance of a “virtual machine” is no longer significantly slower than a “real” machine for many applications
- Relatively inexpensive machines can be used as a “thin” client to access a VM

Examples:

Remote Desktop to my virtual machine





Application at Emory

- Emory uses a CITRIX based “virtual” desktop for many/most clinical programs
- This common framework greatly simplifies administration/backup/security
- Installation of programs is also simplified for IT as a VM can be copied and deployed
- Can access same files/applications from home (if in the firewall), at Emory Clinic, at the Hospital, etc
- Has many advantages in certain scenarios, although can prevent challenges in research environments
- Severely limits ability for clinicians/staff to install/modify programs

The Cloud

- With the rapid advancement of virtualization technology and fast networks, no need to run virtual machines locally
- Many “commodity” calculations / services can be outsourced to online service providers (Amazon S3, Godaddy, Gmail, Dropbox, etc)
- Specific machine/hardware an app runs on is controlled dynamically and can be migrated automatically in case of hardware failure at the hosting provider**
- Commodity model of software + hardware/pay as you go

Scripts/Programming

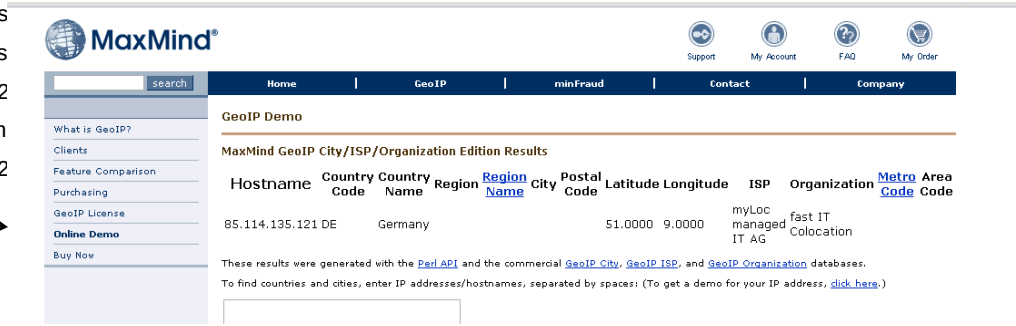
Script/Macro – A program (set of commands) that performs relatively simple action automatically

For example… can open a list of files/documents, looking for the word "diabetes" and "hypertension" and generate co-occurrence statistics vs making a medical student or resident scan through documents manually

Or…

Scripting in the wild / why we need Firewalls

85.114.135.121 - - [16/May/2011:08:44:09 -0400] "GET //scripts/setup.php HTTP/1.1" 404 302 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:09 -0400] "GET //admin/scripts/setup.php HTTP/1.1" 404 308 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:09 -0400] "GET //admin/pma/scripts/setup.php HTTP/1.1" 404 312 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:10 -0400] "GET //admin/phpmyadmin/scripts/setup.php HTTP/1.1" 404 319 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows
85.114.135.121 - - [16/May/2011:08:44:10 -0400] "GET //db/scripts/setup.php HTTP/1.1" 404 305 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:10 -0400] "GET //dbadmin/scripts/setup.php HTTP/1.1" 404 310 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:10 -0400] "GET //myadmin/scripts/setup.php HTTP/1.1" 404 310 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:10 -0400] "GET //mysql/scripts/setup.php HTTP/1.1" 404 308 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:11 -0400] "GET //mysqladmin/scripts/setup.php HTTP/1.1" 403 317 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:11 -0400] "GET //typo3/phpmyadmin/scripts/setup.php HTTP/1.1" 404 319 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows
85.114.135.121 - - [16/May/2011:08:44:11 -0400] "GET //phpadmin/scripts/setup.php HTTP/1.1" 404 311 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:11 -0400] "GET //phpMyAdmin/scripts/setup.php HTTP/1.1" 403 317 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:12 -0400] "GET //phpmyadmin/scripts/setup.php HTTP/1.1" 403 317 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:12 -0400] "GET //phpmyadmin1/scripts/setup.php HTTP/1.1" 404 314 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:12 -0400] "GET //phpmyadmin2/scripts/setup.php HTTP/1.1" 404 314 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:12 -0400] "GET //pma/scripts/setup.php HTTP/1.1" 404 306 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:12 -0400] "GET //web/phpMyAdmin/scripts/setup.php HTTP/1.1" 404 317 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98
85.114.135.121 - - [16/May/2011:08:44:13 -0400] "GET //xampp/phpmyadmin/scripts/setup.php HTTP/1.1" 404 319 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows
85.114.135.121 - - [16/May/2011:08:44:13 -0400] "GET //web/scripts/setup.php HTTP/1.1" 404 306 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:13 -0400] "GET //php-my-admin/scripts/setup.php HTTP/1.1" 404 315 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:13 -0400] "GET //websql/scripts/setup.php HTTP/1.1" 404 309 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
85.114.135.121 - - [16/May/2011:08:44:13 -0400] "GET //phpmyadmin/s
85.114.135.121 - - [16/May/2011:08:44:14 -0400] "GET //phpMyAdmin/s
85.114.135.121 - - [16/May/2011:08:44:14 -0400] "GET //phpMyAdmin-2
85.114.135.121 - - [16/May/2011:08:44:14 -0400] "GET //php-my-admin
85.114.135.121 - - [16/May/2011:08:44:14 -0400] "GET //phpMyAdmin-2

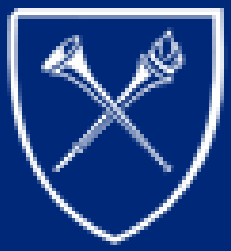


Databases and Spreadsheets

≠

Learning curve of spreadsheets is low, however search and query capability is extremely limited

Clinical Informatics Academy



EMORY UNIVERSITY

Introduction to Data Management

Fusheng Wang
Center for Comprehensive Informatics
Emory University

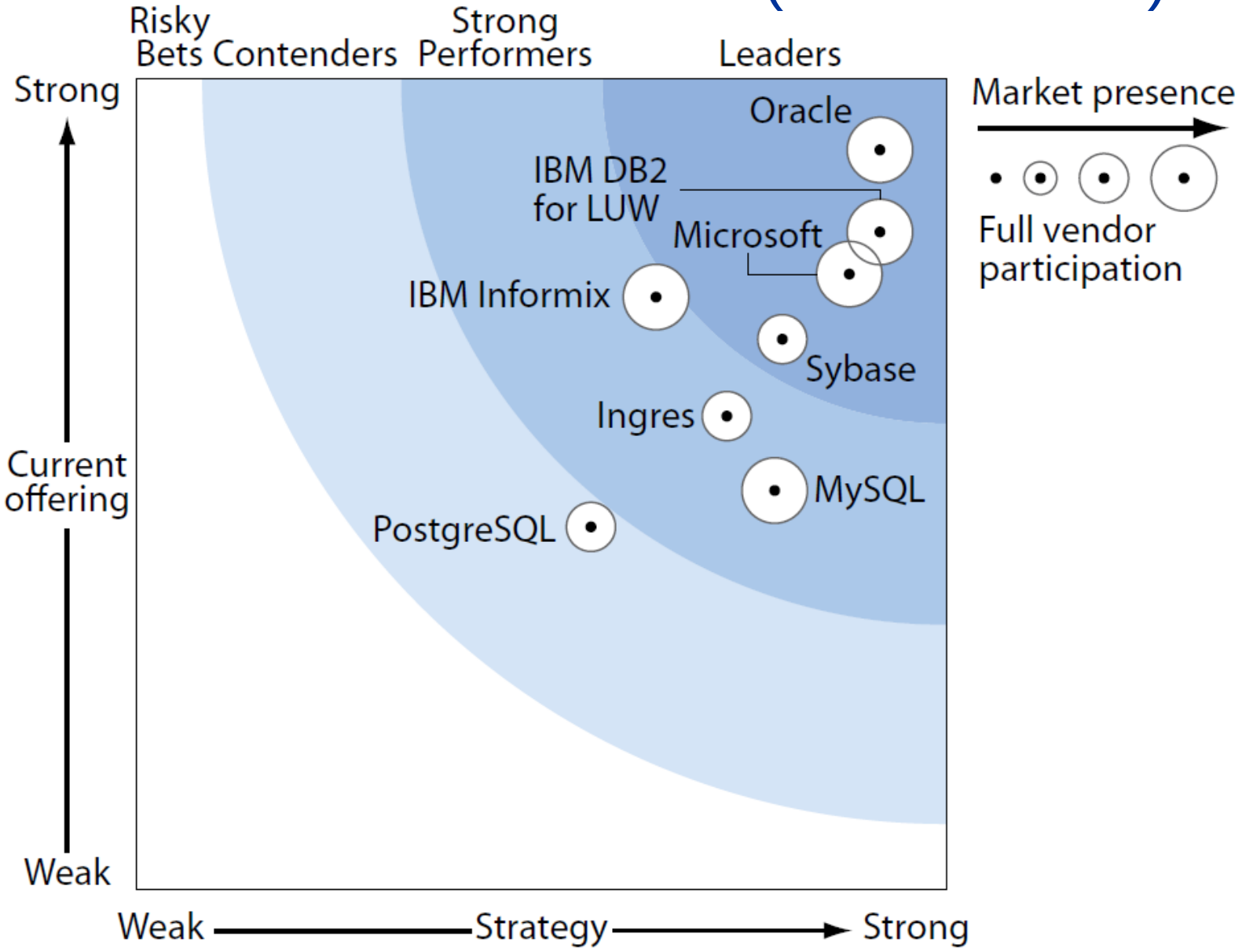
Why Databases?

- Data can be stored using multiple methods such as text files, comma delimited data files, spreadsheets, databases
- Benefits of using a database:
  - A standard interface for accessing data
  - Multiple users could simultaneously insert, update and delete data
  - Data could be changed without risk of losing data and its consistency
  - Efficiently handle huge volumes of data
  - Tools for data backup, restore and recovery
  - Security
  - Reduce redundancy
  - Data independence

Database Management Systems (DBMS)

- A Database Management System (DBMS) is a software system designed to store, manage, and facilitate access to databases
- A relational DBMS (RDBMS) is a DBMS that is based on the relational model
- Objectives of DBMS
  - Representing information: data modeling
  - declarative language for querying data: SQL, XQuery…
  - Efficient support of queries with access methods
  - Controlling concurrent access
  - Reliable data storage

The Current RDBMS Market (Forrester'09)

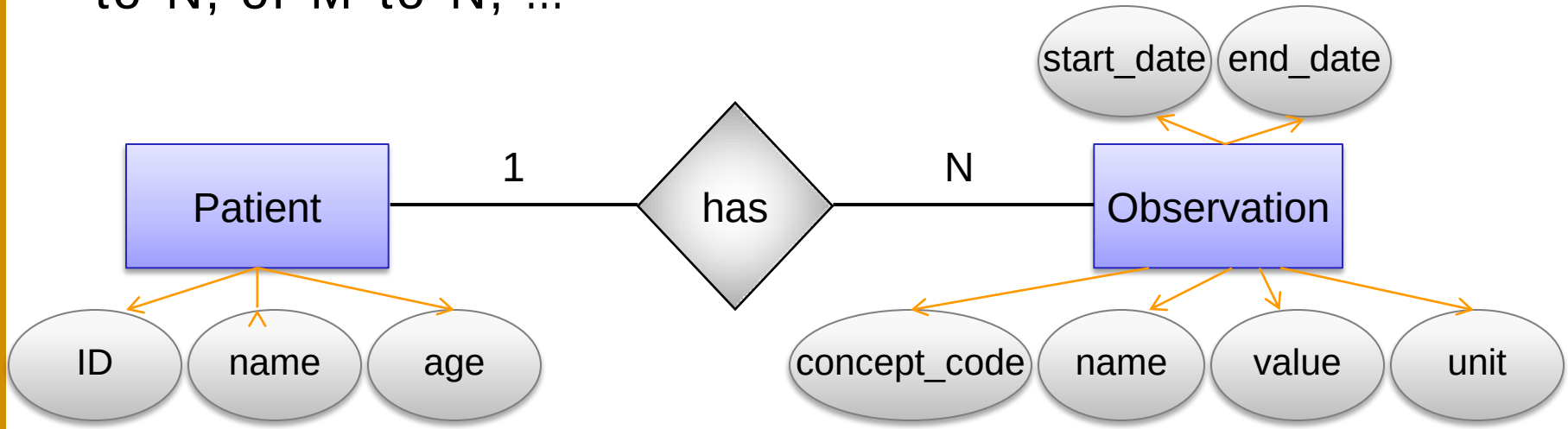


Data Model

- Define how data is to be represented, structured linked, and constrained
  - Independent of specific implementations and protocols
- Types of data models
  - Hierarchical model, network model
  - Relational model – invented by E. F. Codd from IBM
  - Entity Relational (ER) model
  - Object-oriented model
  - Semi-structured (XML) model
- Relational model
  - describes data as a collection of relations
  - Focuses on providing better data independence
  - Implemented by most DBMS in the market

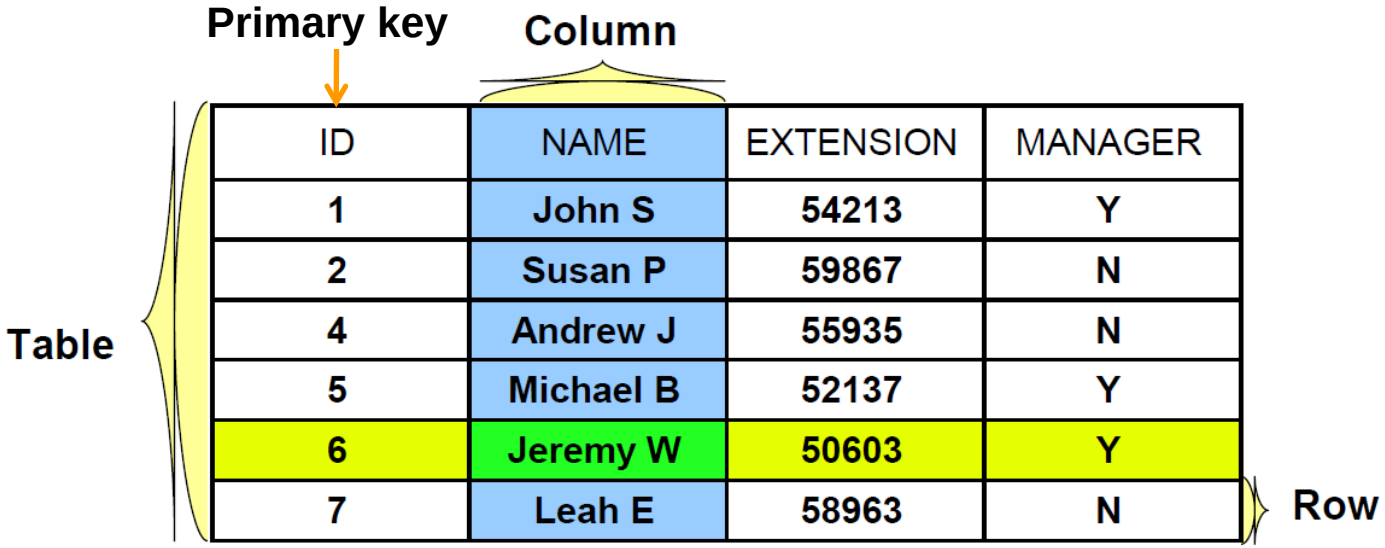
ER Model

- A popular conceptual model for database design
- A database be thought of a collection of instances of entities, independent of any other entities in the database
- Entities have attributes to characterize the entity
- There could be relationships between entities: 1-to-1, 1-to-N, or M-to-N, …



Relational Database with Relational Model

- A relational database is implemented based on relational data model
- Data stored in tables, consists of columns and rows
- Each column has a specific data type
- Constrained can be specified: primary key, uniqueness,...
- Relationships can be defined with foreign keys



ID	NAME	EXTENSION	MANAGER
1	John S	54213	Y
2	Susan P	59867	N
4	Andrew J	55935	N
5	Michael B	52137	Y
6	Jeremy W	50603	Y
7	Leah E	58963	N

SQL: Standard Language for RDBMS

- SQL (Structured Query Language) is the standard language of relational database access
- Multiple standard revisions and multiple flavors (implementations) exist
- Procedure SQL (PL/SQL, SQL PL, …) adds programming capabilities into SQL
- A SQL query is compiled and executed by the DBMS engine and the result is sent to the client
- Many approaches to optimize SQL query performance: indexes, parallel disk readings, normalization, etc.

SQL (2)

- Data Definition Language (DDL), defines properties of data objects. e.g. creation of a table:

```
CREATE TABLE OBSERVATION_FACT (
   EONCOUNTER_NO INTEGER NOT NULL,
   PATIENT_ID    INTEGER,
   AGE           INTEGER,
   CONCEPT_CODE  VARCHAR(50),
   NAME          VARCHAR(50),
   VALUE         VARCHAR(50),
PRIMARY KEY(ENCOUNTER_NO, CONCEPT_CODE) );
```

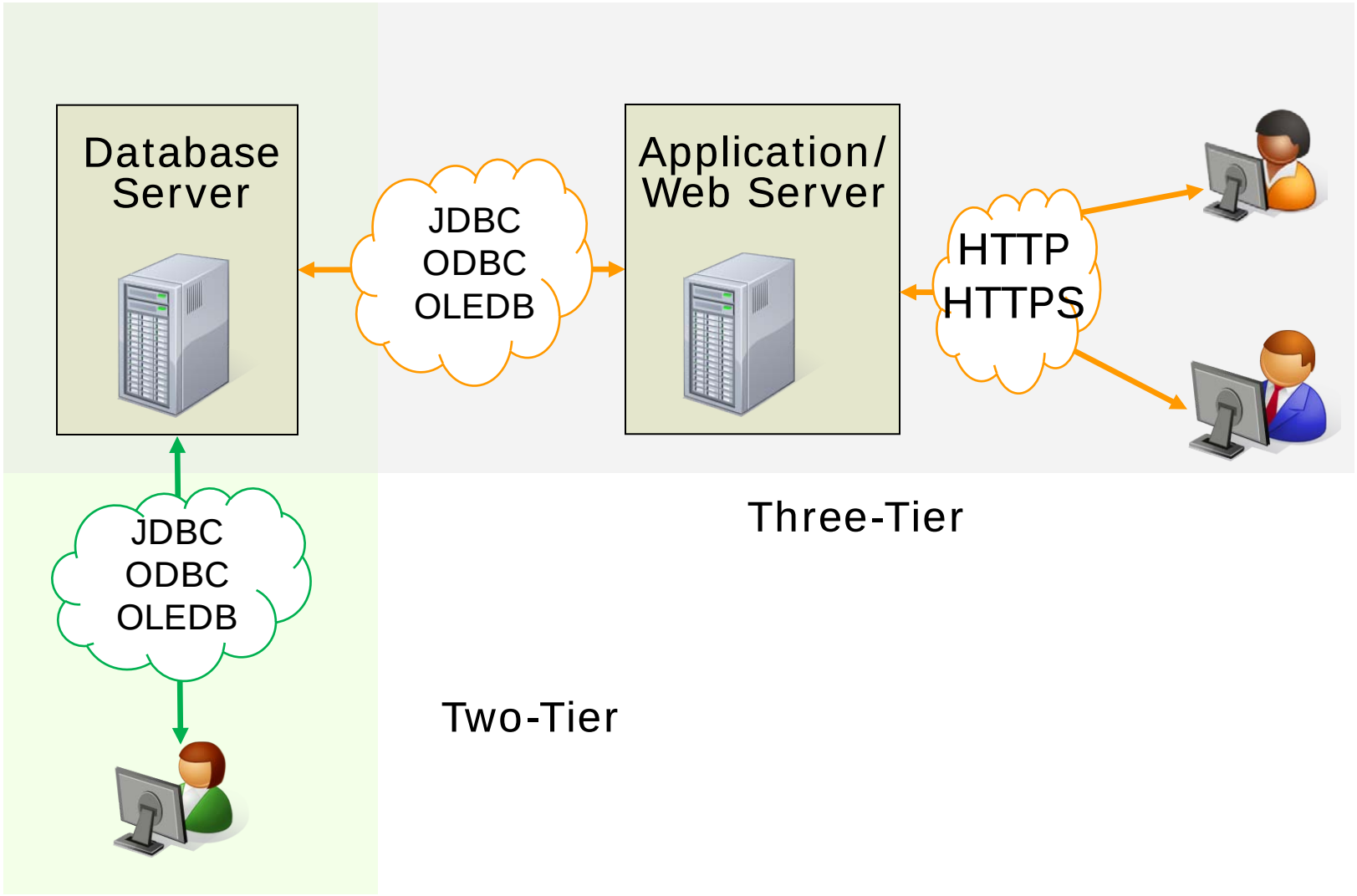
- Data manipulation language (DML): retrieve, insert, update and delete data

```
SELECT AGE, COUNT(*) FROM OBSERVATION_FACT
WHERE NAME ='CHEST PAIN'
GROUP BY AGE
ORDER BY AGE
```

Reliable Storage and High Availability

- Problems could happen: system outage, transaction failure, disk failure, disaster
- Backup/restore: DBMS tracks changes as logs, thus recovery is possible when failure happens
  - Full snapshot backup
  - Incremental backup: only changes since last successful full backup
  - Restore: rebuild database from backups + logs
- High availability: eliminate or minimize downtime
  - Creating and maintaining replica versions of database
  - Failover takes place when disaster happens

Access of Databases



Data Exchange: XML

- The e**X**tensive **M**arkup **L**anguage (XML) defines a generic syntax used to mark up data with simple, human-readable tags
- Standard language for data exchange over the Web
- HL7 CDA messaging in XML
- Data can be published directly in XML from DBMS

```
<component>
   <section>
      <templateId root="1.3.6.1.4.1.19376.1.5.3.1.3.4"/>
      <code code="10164-2" displayName="History of present illness"
            codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC"/>
      <title> History of present illness </title>
      <text>Carcinoma of breast. Post operative diagnosis: same. left UOQ
breast mass. </text>
   </section>
</component>
```

Specialized DBMSs

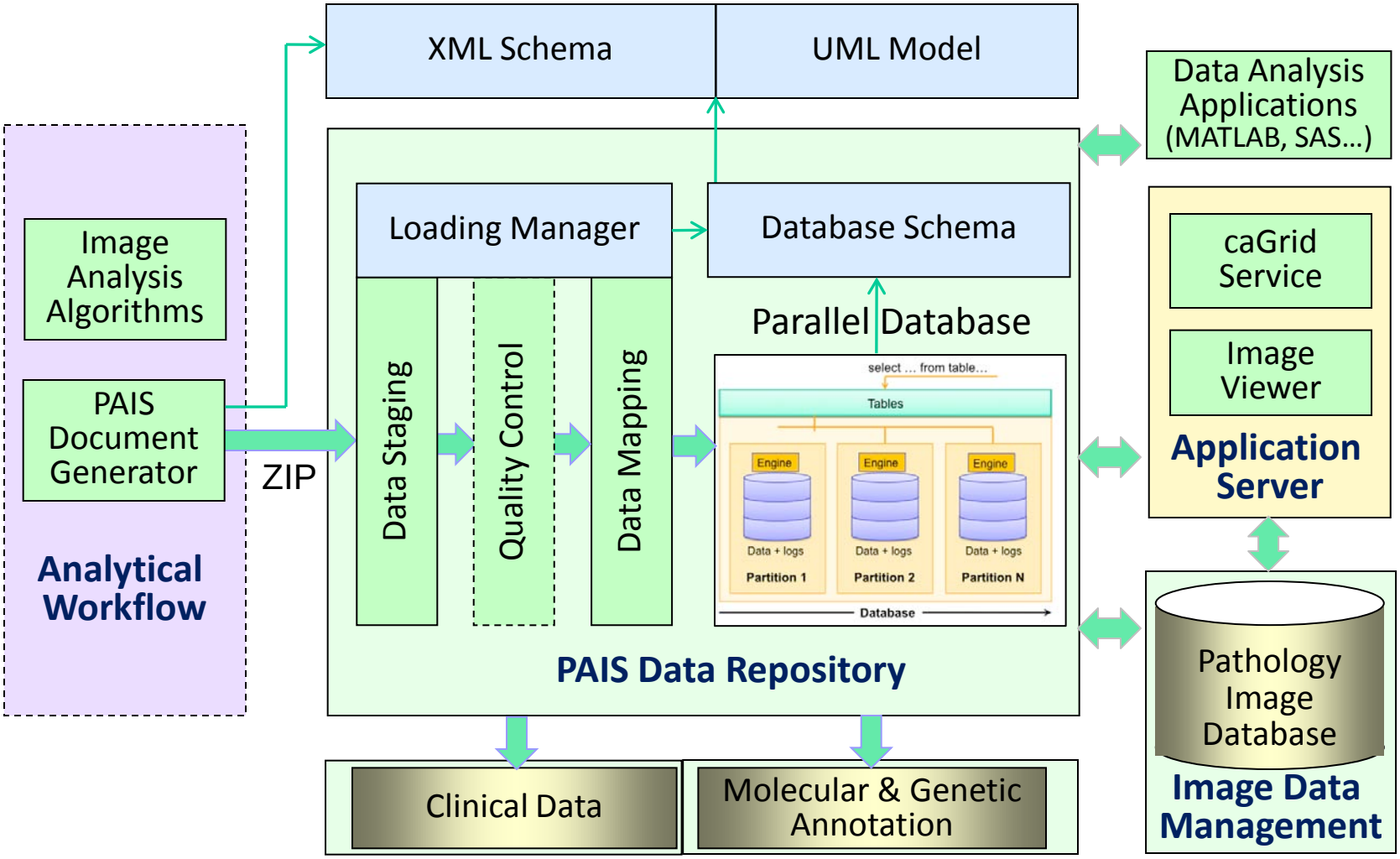
Special data models/data types/queries:

- XML DBMS: manage XML data directly in XML model and query language
  - e.g.: DB2 pureXML, Oracle XML DB
- Spatial DBMS: manage location related information. E.g., find patients located within 10 miles radius
  - e.g.: Oracle Spatial, ArchGIS, DB2 Spatial
- Temporal DBMS: manage temporal oriented information. e.g.: what drugs have been prescribed with Proventil?
  - TeraData, DB2 TemporalDB

Parallel DBMS: use data partitioning and parallel data access to increase I/O bandwidth for scalability

– Oracle RAC, DB2 DPF, DB2 pureScale, TeraData

Example Database: PAIS



Questions?