Topics

- Finish up some XML
- Class Survey
- GIS
  - GRASS Demo
  - Raster vs. Vector
- Spatial Databases
  - Multidimensional Indexes
  - Study Questions
- Hand back Homework 1

XML for Science

- Recall features of Science Data:
  - Read-oriented access
  - Provenance
    - who, what, when, where, why
  - Interesting Data Types
    - timeseries
    - spatial
    - arrays
    - images
  - Scale

XML for Science

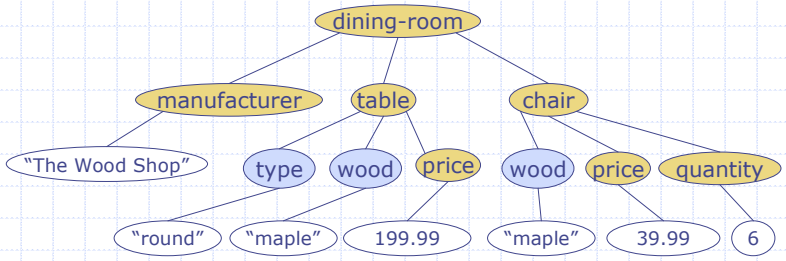
- Read-oriented access?
  - perfect!
- Provenance
  - requires some flexibility; no problem
- Interesting Data Types
  - ...and special file formats
- Scale
  - could get ugly

Interesting Data Types

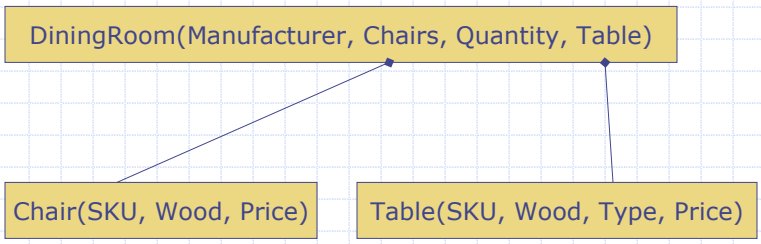
- Data locked in binary file formats
  - **Binary Format Description Language**
    - [Myers, Chappell 2000]
  - **Data Format Description Language**
    - [OpenGrid Project]
  - **Retrofitting Data Models**
    - [Howe, Maier SSDBM 2005]
  - **PADX**
    - [Fernandez et al, PLANX 2006]
  - **XDTM**
    - [Foster, Voeckler et al. Global Grid Forum 2005]

XML Storage

- Use RDBMS as your storage engine
- Two approaches:
  - Schema-aware
  - Schema-oblivious



XML Storage: Schema-aware

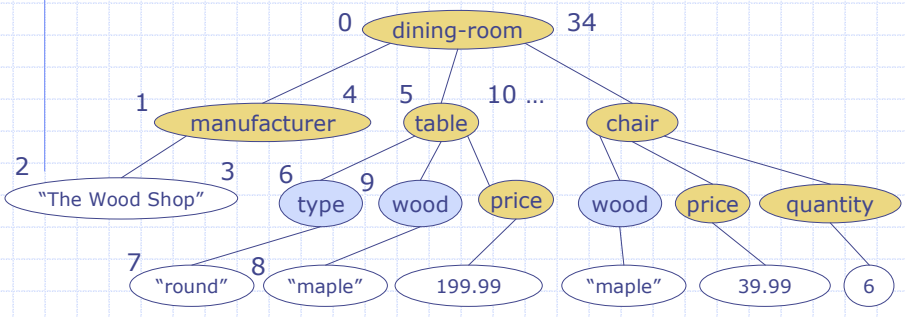


XML Storage: Schema-oblivious

Edge(NodeId, Tag, Value, ParentNodeId)

- Remember fancy node-labeling schemes…

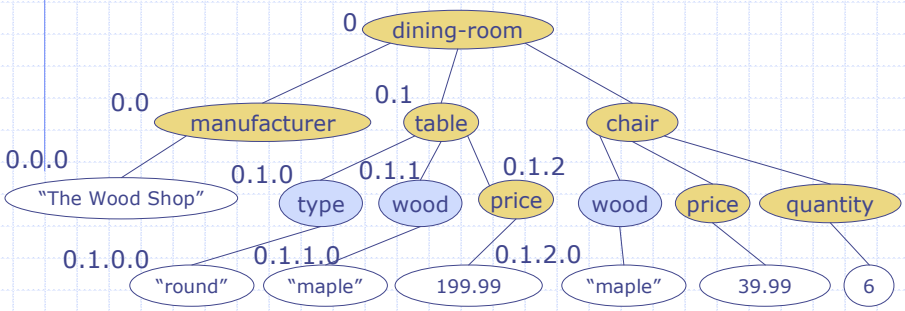
Left/Right Labeling



Which queries are easy and fast?

What did we say the problems were?

Path Labeling

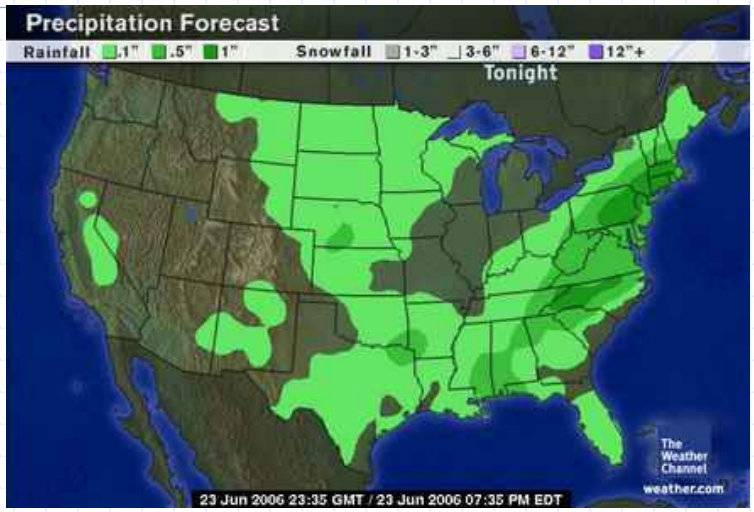


What queries are fast and/or easy?

What did we say the problems were?

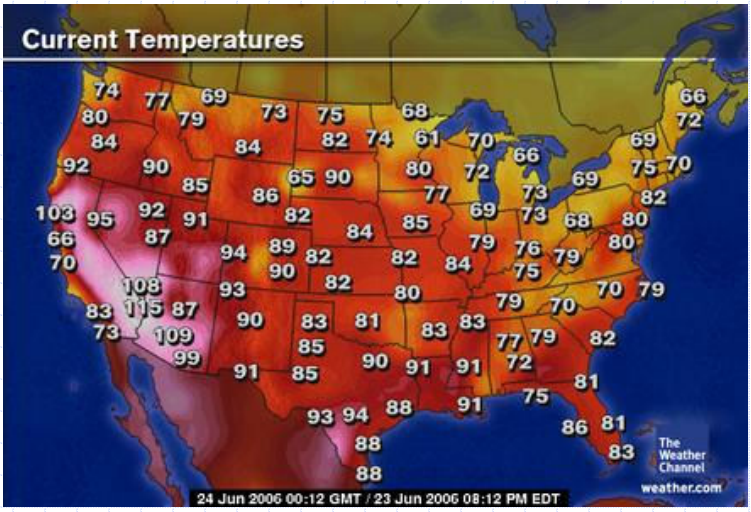
Geographic Information Systems and Spatial Databases

Represent This



source: USA Today Weather

Represent This

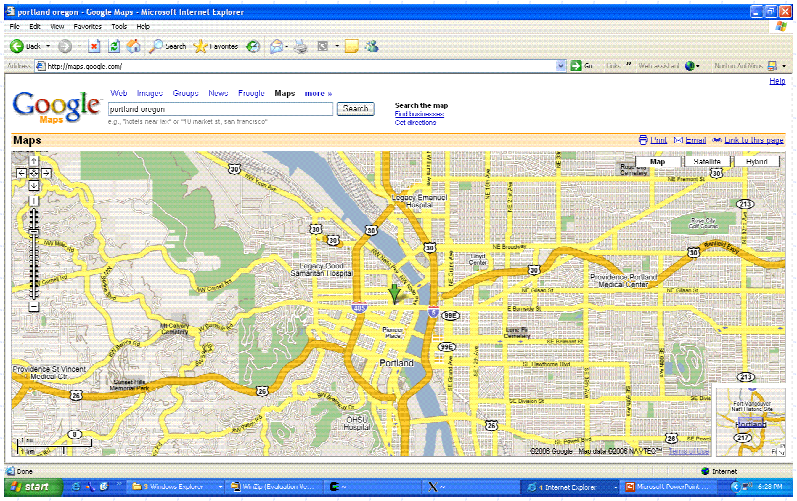


source: USA Today Weather

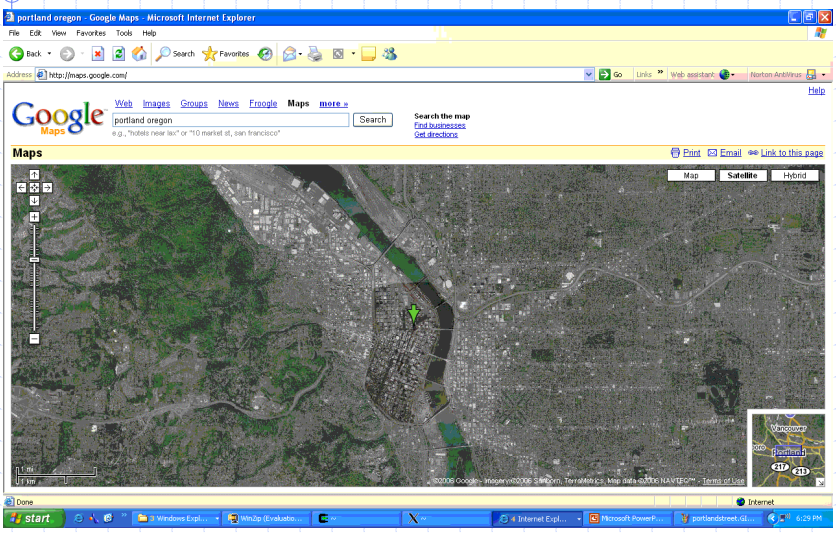
Represent This

Source: www.mapsofworld.com/

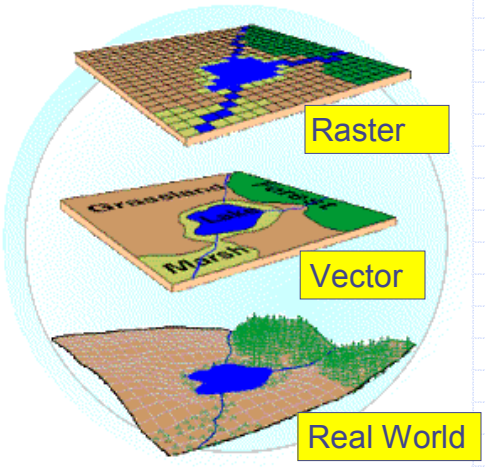
Represent This



Represent This

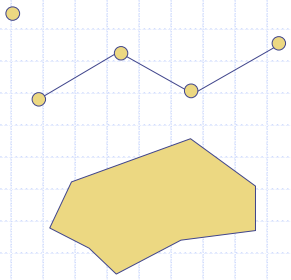


Raster-Vector Data Model



Vector Data

- The world is a collection of objects
  - Points
  - Lines
  - Polygons

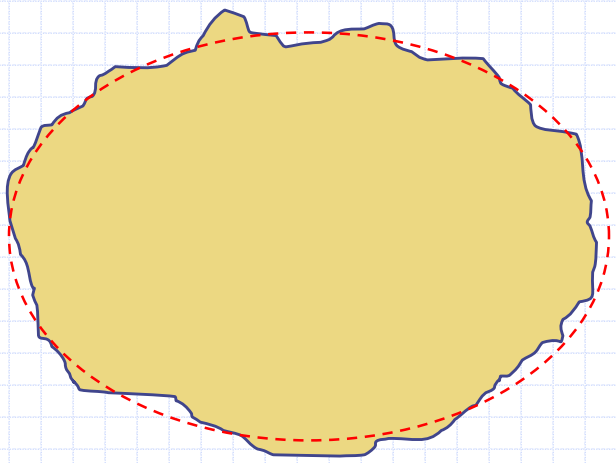


Intermission: Class Survey

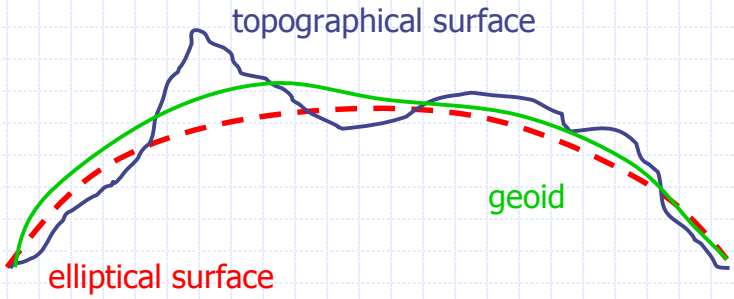
GIS: Geographic Information Systems

- GIS is a collection of:
  - file format converters and readers
  - algorithms (visibility, buffer, etc.)
  - visualization tools
  - *georeferencing libraries*…

Geodesy: Elevation

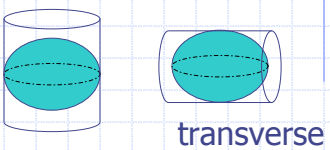


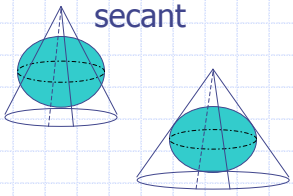
Geodesy: Elevation

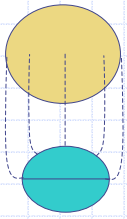


Geodesy: Map Projections

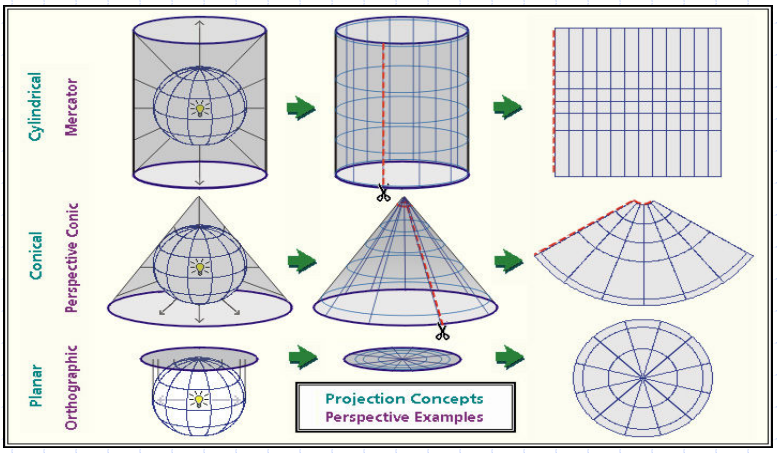
- Cylindrical (Mercator)
- Conical
- Planar (Azimuthal)







Map projection



http://www.geog.ucsb.edu/~dylan/mtpe/geosphere/topics/map/map1.html

Spatial Databases

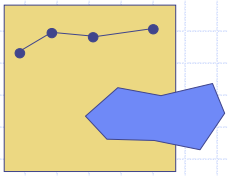
- Spatial Database is an RDBMS equipped with:
  - spatial types (points, lines, regions)
  - spatial functions (georeferencing, other algorithms)
  - *multi-dimensional indices*

Spatial Databases

- Non-spatial queries:
  - List all gas stations owned by Bob
  - List all unpaved streets
  - List galaxies with luminosity > x
- Spatial queries:
  - List all gas stations with one mile of the river
  - List all gas stations touching an unpaved street
  - List galaxies within 2 arcmin of a quasar

Spatial Predicates

- Each value has type *Geometry*
  - plus a subtype:
    - point, line, polyline, polygon
- Spatial Predicates
  - Geo Overlap(Geo, Geo)
  - bool Contains(Geo, Geo)
  - Geo Boundary(Geo)
  - bool Touches(Geo, Geo)
  - Geo Nearest(Geo, [Geo])
  - Geo Nearest(Geo, [Geo])

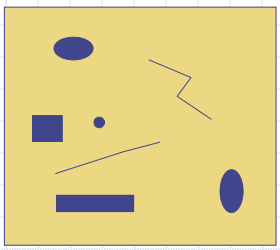


Multidimensional Indices

- How do we answer topological questions efficiently?
- B-Tree: one dimension only

Multidimensional Indices

- One way: Impose a total order on the objects, then use an ordinary one-dimensional index
- Ideas?



Space Filling Curves

- How do we organize objects on disk so proximity is preserved?