Introduction to Persistency and Berkeley DB

Philip Johnson

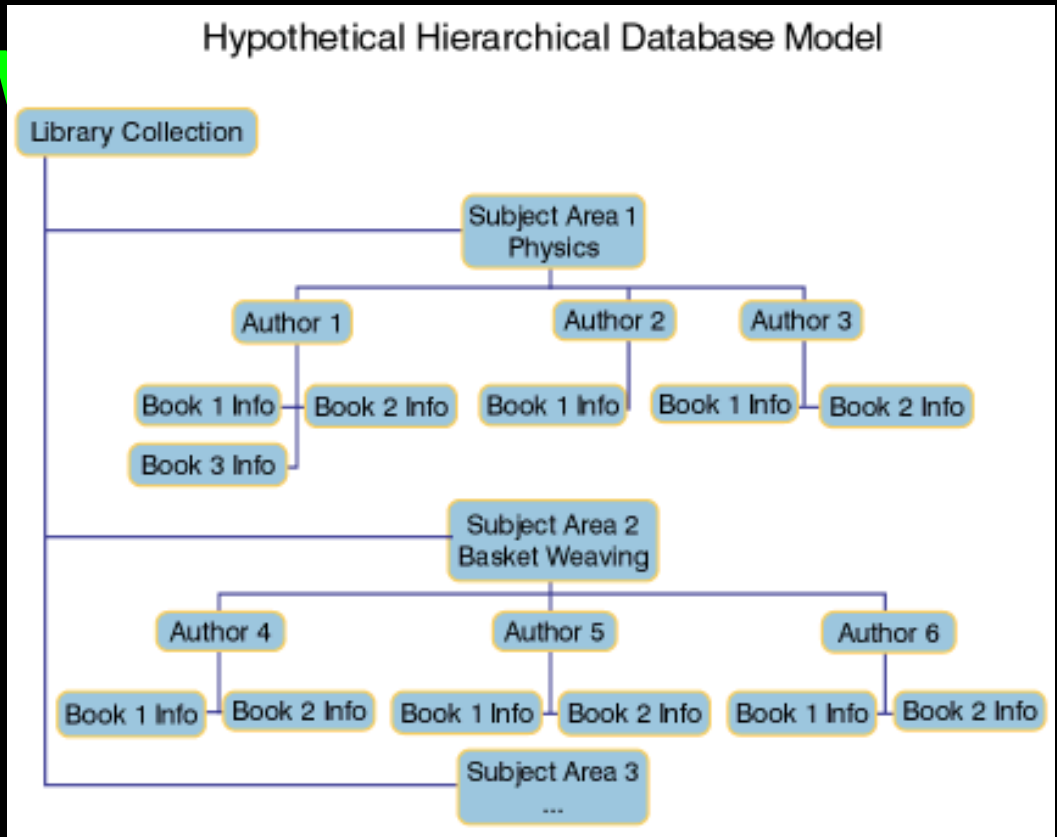
Collaborative Software Development Laboratory

Information and Computer Sciences

University of Hawaii

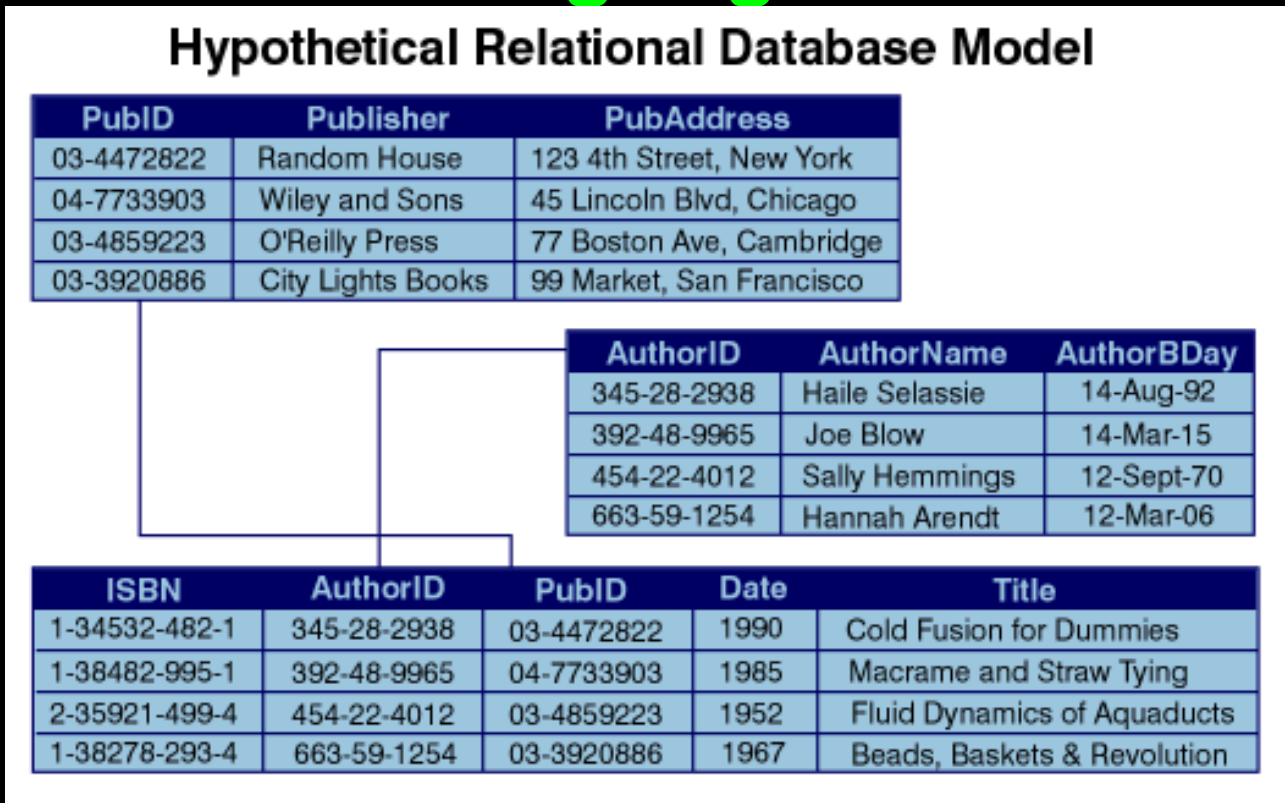
Data Model: Hierarchical

- Tree of nodes and subnodes
- "Cousins" have same structure
- Limited data access language (traversal)
- Data model for XM



Data Model: Relational

Set of tables containing fields
Limited columns, unlimited rows
Primary, secondary keys for access
Powerful data access language



Hypothetical Relational Database Model

PubID	Publisher	PubAddress
03-4472822	Random House	123 4th Street, New York
04-7733903	Wiley and Sons	45 Lincoln Blvd, Chicago
03-4859223	O'Reilly Press	77 Boston Ave, Cambridge
03-3920886	City Lights Books	99 Market, San Francisco

AuthorID	AuthorName	AuthorBDay
345-28-2938	Haile Selassie	14-Aug-92
392-48-9965	Joe Blow	14-Mar-15
454-22-4012	Sally Hemmings	12-Sept-70
663-59-1254	Hannah Arendt	12-Mar-06

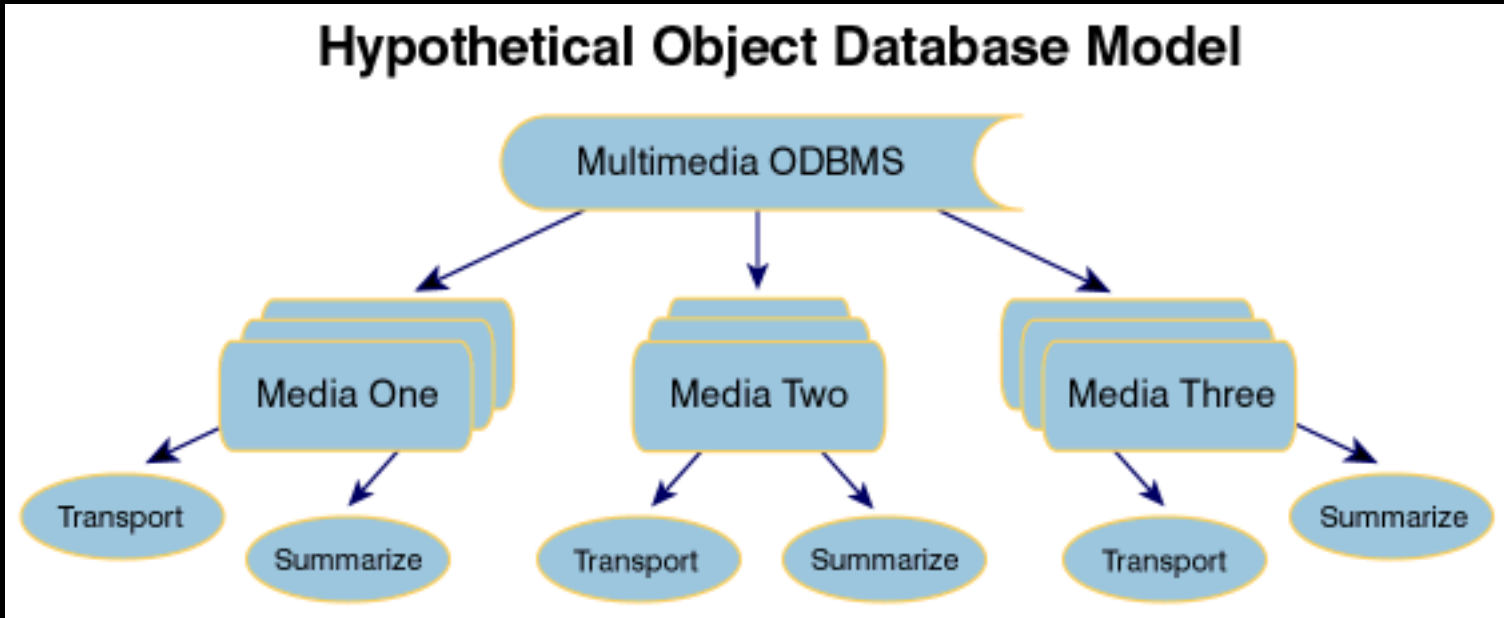
ISBN	AuthorID	PubID	Date	Title
1-34532-482-1	345-28-2938	03-4472822	1990	Cold Fusion for Dummies
1-38482-995-1	392-48-9965	04-7733903	1985	Macrame and Straw Tying
2-35921-499-4	454-22-4012	03-4859223	1952	Fluid Dynamics of Aquaducts
1-38278-293-4	663-59-1254	03-3920886	1967	Beads, Baskets & Revolution

Data model: Object

Similar to hierarchical

- no limitation on node structure.

Nodes can contain methods to guide retrieval.



Data Model: NoSQL

Generic term for all non-relational DBs

- key-value stores, document DBs, etc.

A rejection of the RDBMS "orthodoxy"

Examples:

- Google's BigTable, CouchDB

Object-relational "impedance mismatch"

A primary difficulty with relational DBs

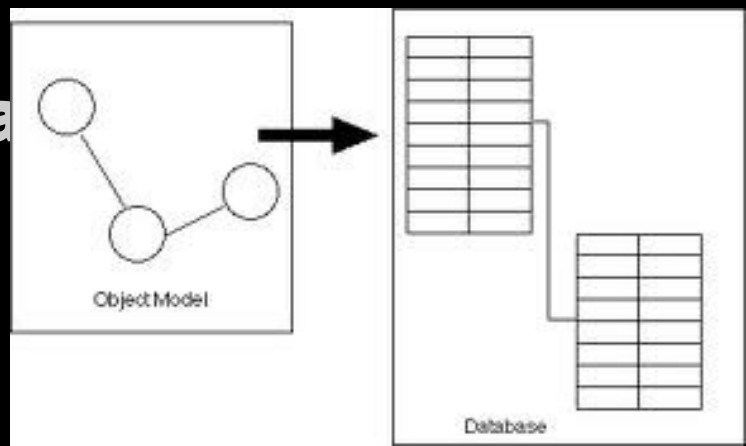
- How to map table-based data to objects?

Issues:

- Objects nested, tables flat
- Objects traversed, tables searched
- Conversion back and forth

Approaches:

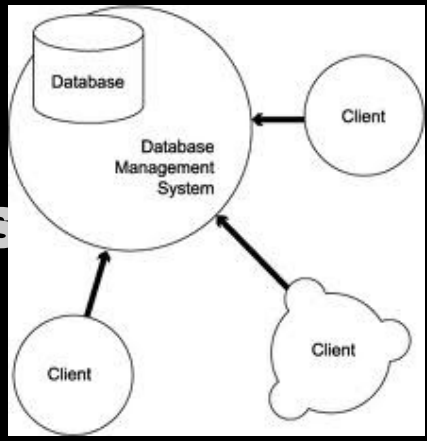
- ORM (Object-relationa frameworks
- Object databases



Client-server vs. embedded

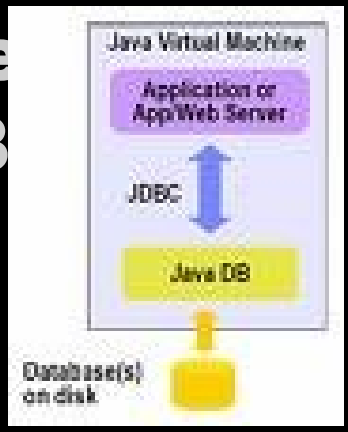
Client-server:

- Database runs in one process
- Supports multiple connections
- Relatively complicated to set administer.



Embedded:

- Database runs "within" the clie
- Only one client connects to DB
- Relatively simple to set up, administer.



How to choose?

Beyond the scope of this lecture!

- Take ICS 321
- Talk to ICS Professor Lipyeow Lim

BerkeleyDB

Open source, embedded, object-oriented, key-value, non-relational, highly concurrent, transactional, high performance, cross-platform

Benefits for us:

- Easy to set up and use.
- No impedance mismatch.
- High performance

Constraints:

- Embedded
- No use of SQL language

On to the demo