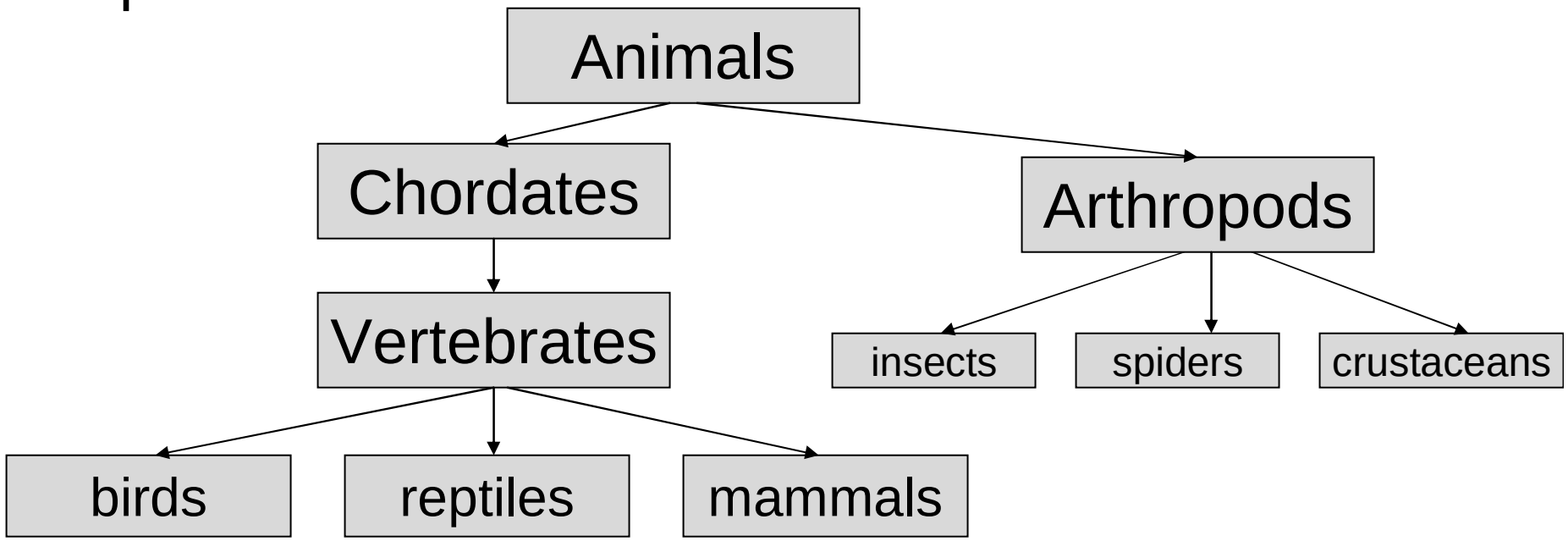
Example: Taxonomy of Organisms

- Hierarchy of categories:
  - Kingdom - phylum – class – order – family – genus - species



  - How would you design a relational schema for this?

Relational Schema for Taxonomy

- Adjacency list
- Idea – each tuple has a parent id

name	Parent_id
Animal	null
Chordates	Animal
Arthropods	Animal
Vertebrates	Chordates
Insects	Arthropods
Spiders	Arthropods
Birds	Vertebrates
Reptiles	Vertebrates

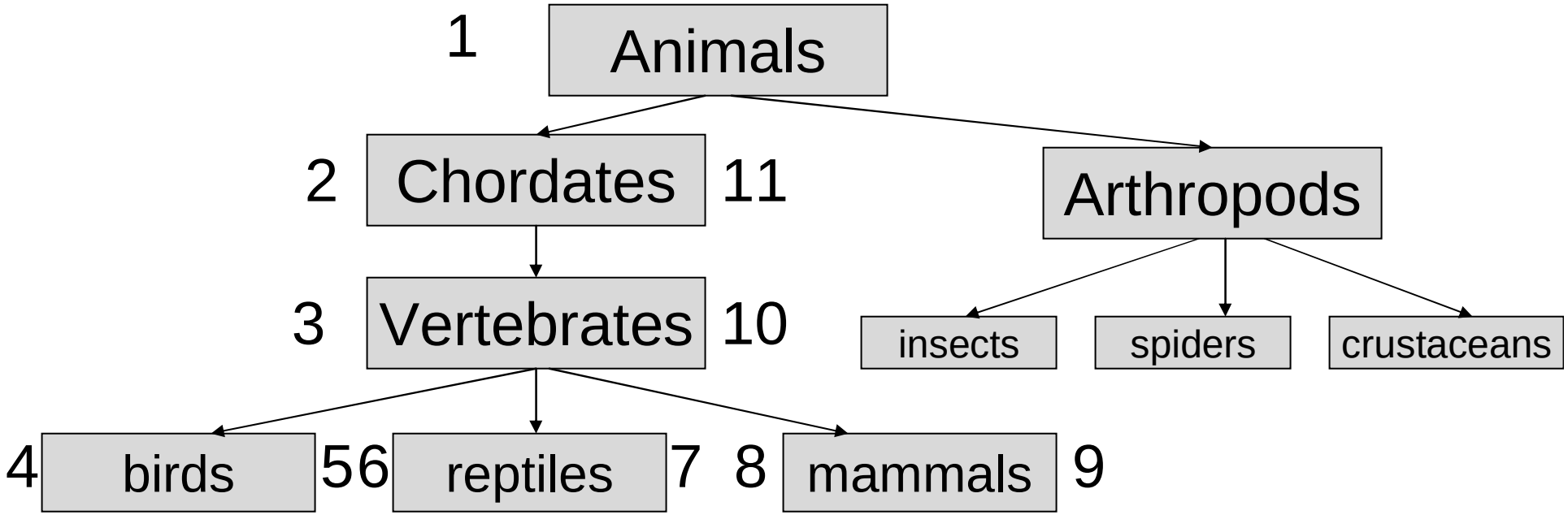
Problem: Recursive Queries

- “Find all ancestors/descendants of X”
- “Find all arthropods that are not crustaceans”
- “Print the whole tree”
- Solutions may require many lines of code and may not be very robust

Another Solution: nested sets

- Idea: traverse entire tree, assign each object in tree a left_id and a right_id
- Each child left_id and right_id are between parent's left_id and right_id
- Label entire tree using depth-first traversal of tree

Nested sets example



name	Left_id	Right_id
Animal	1	20
Chordates	2	11
Vertebrates	3	10
Birds	4	5

Challenges

- Need a script to traverse database
- Script must be rerun every time database is updated
- Update time linear in size of table
  - More costly than B-trees

Multidimensional Arrays

- How would you store an array in a database?
- One option: x, y, z values as columns
- A[5,1,2] = 2.53 becomes:

x	y	z	value
5	1	2	2.53

Problems?

- High storage overhead – each index is now stored as an integer value
- Queries more complex
- Lose proximity information, what is “near” an array element
- Iterating over arrays takes longer
  - A[5,1,2] and A[5,1,3] may not be stored near each other

Should I use a RDBMS?

- Advantages:
  - Leverage years of research and practice
    - E.g., indexes, query optimization
  - Many commercial and open-source products available

Should I use a RDBMS?

- Disadvantages
  - May be difficult to express data/queries using relational model
    - e.g., arrays, hierarchical data
  - Performance issues
- Need to consider these tradeoffs!
- Well-designed database can help
- “Augment” database to support domain