

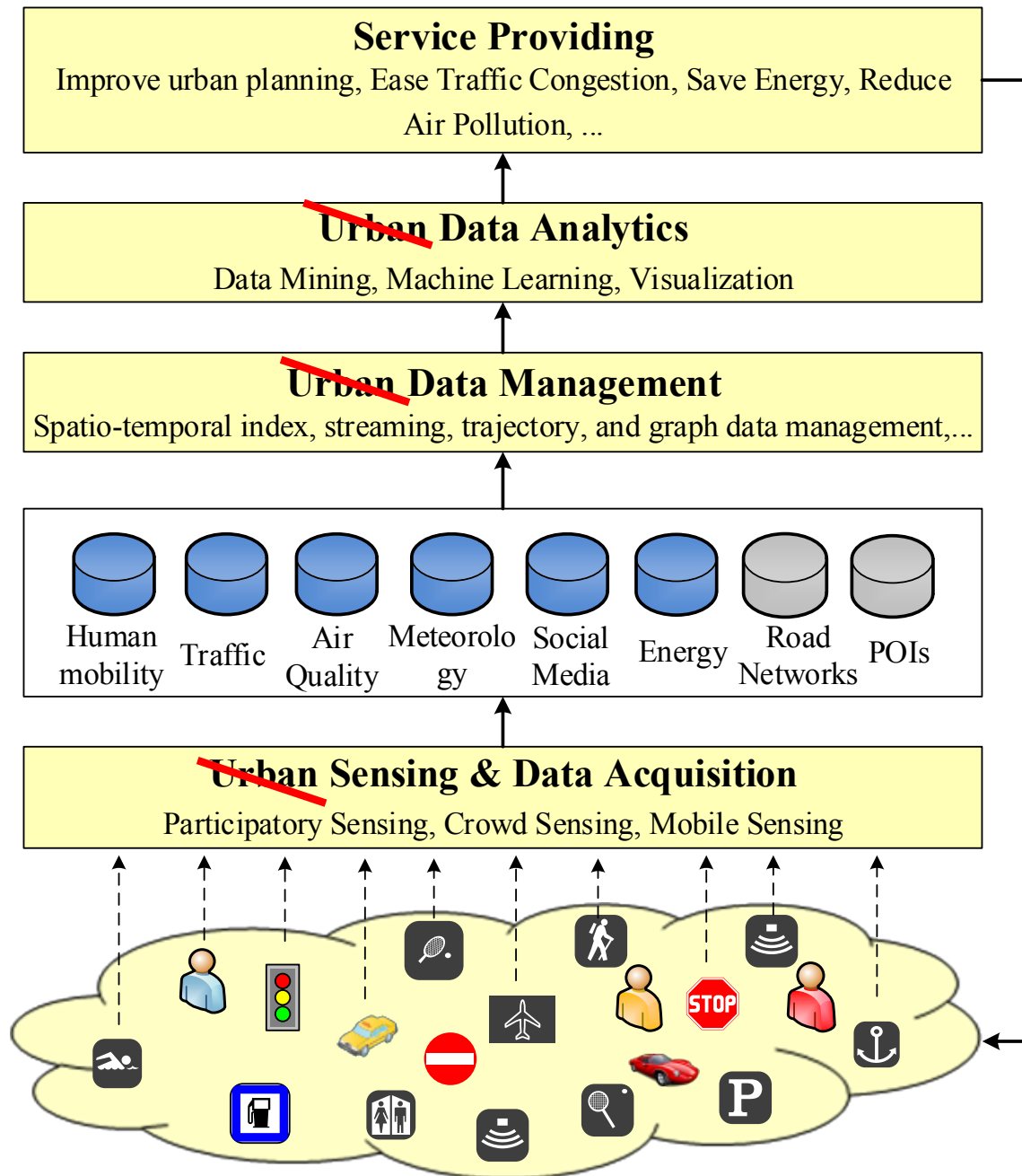
Welcome to
DS3010:
DS-III: Computational Data
Intelligence
Data Management
Prof. Yanhua Li

Time: 11:00am – 12:50pm M & R

Location: HL 114

D-term 2022

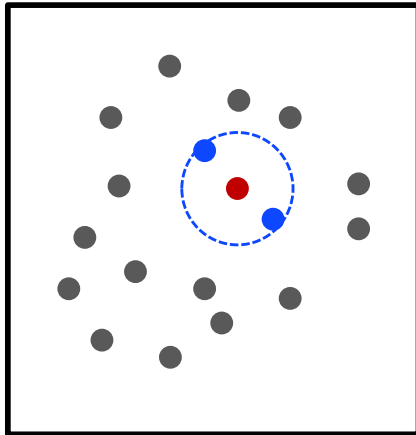
Data pipeline



Urban Computing: concepts, methodologies, and applications.
Zheng, Y., et al. *ACM transactions on Intelligent Systems and Technology*.

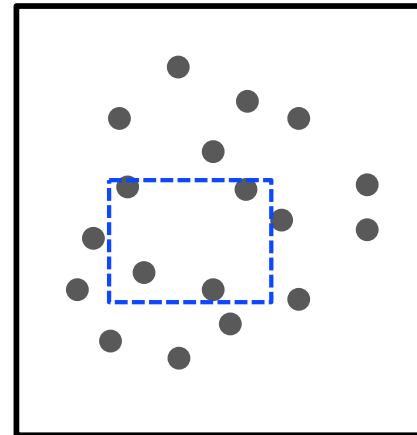
2D-Spatial Queries

K Nearest Neighbour (KNN)
Queries



Given a point or an object,
find the nearest object that
satisfies given conditions

Region (Range) Query



Ask for objects that lie
partially or fully inside a
specified region.

Spatial/Temporal Indexing Structures

❖ Temporal Indexing (1-D data)

- List index
- B-tree

❖ Space Partition-Based Indexing Structures (2-D data)

- Grid-based
- Quad-tree

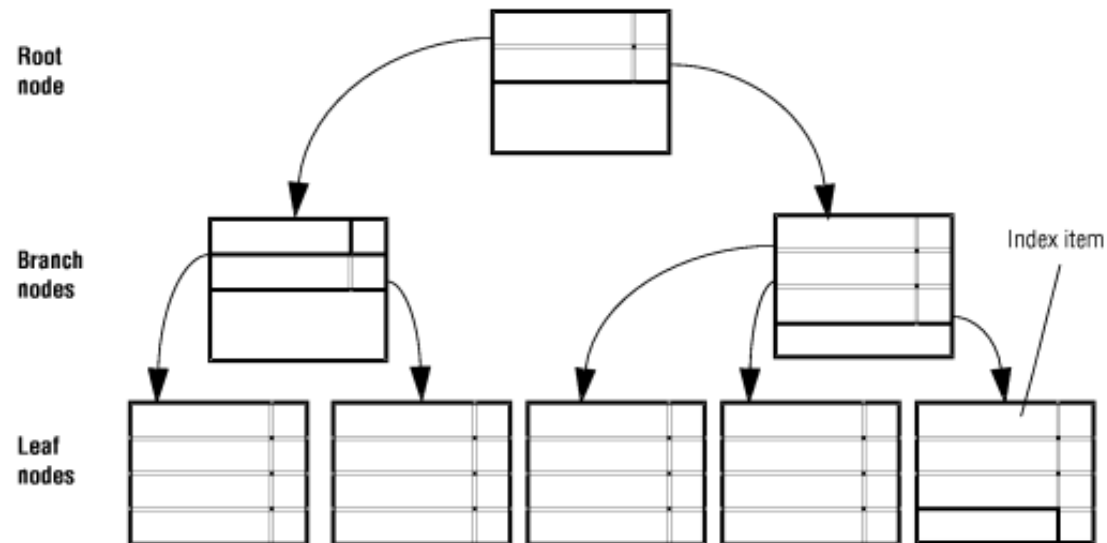
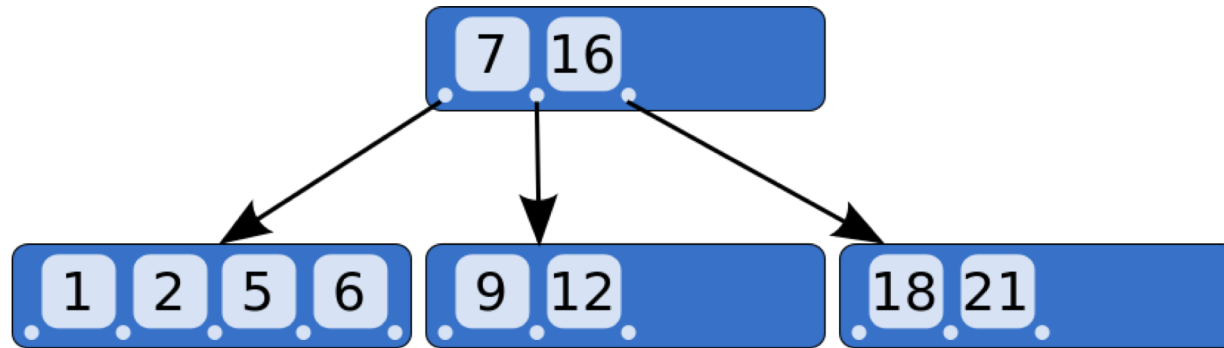
List Index Structure

❖ Example

- From YouTube Prefixes
- To YouTube videos IDs



Full B-Tree Structure



B-Tree Index

- ❖ B-tree is the most commonly used data structures for indexing.
- ❖ It is fully dynamic, that is it can grow and shrink.

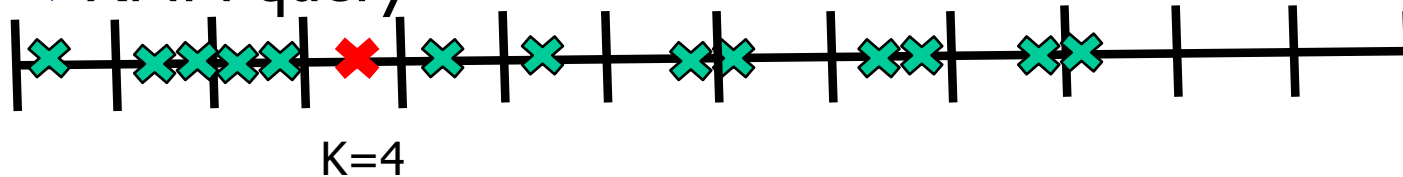
Three Types B-Tree Nodes

- ❖ **Root node** - contains node pointers to branch nodes.
- ❖ **Branch node** - contains pointers to leaf nodes or other branch nodes.
- ❖ **Leaf node** - contains index items

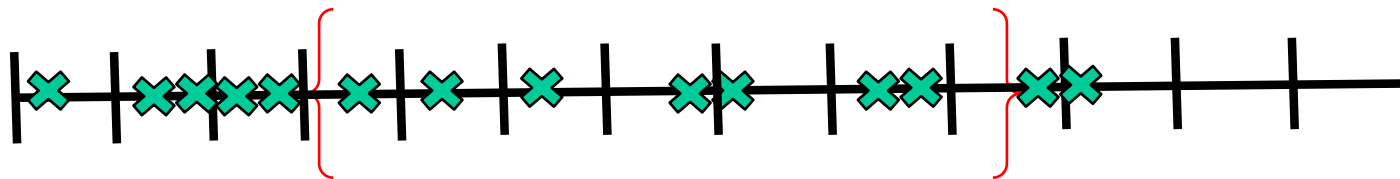
List and B-tree Index for KNN and Range queries

❖ List index

❖ KNN query



❖ Range Query



- Similar for B-tree

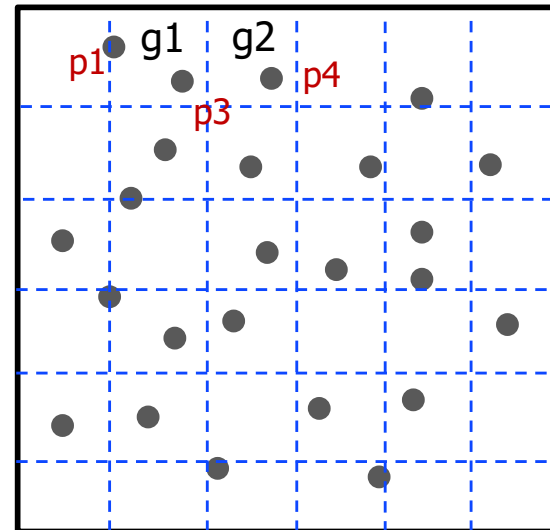
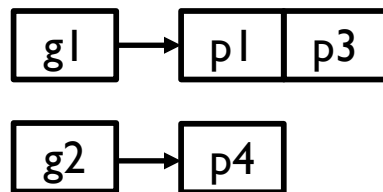
Spatial/Temporal Indexing Structures

- ❖ Temporal Indexing (1-D data)
 - List index
 - B-tree
- ❖ Space Partition-Based Indexing Structures (2-D data)
 - Grid-based
 - Quad-tree

Grid-based Spatial Indexing

❖ Indexing

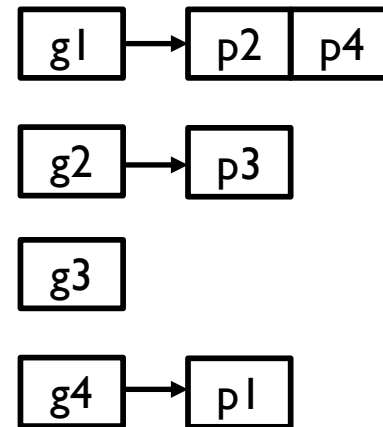
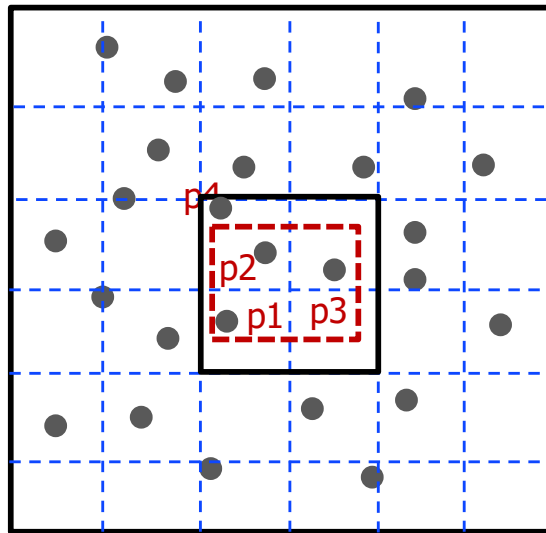
- Partition the space into disjoint and uniform grids
- Build an index between each grid and the points in the grid



Grid-based Spatial Indexing

❖ Range Query

- Find the grids intersecting the range query
- Retrieve the points from the grids and identify the points in the range



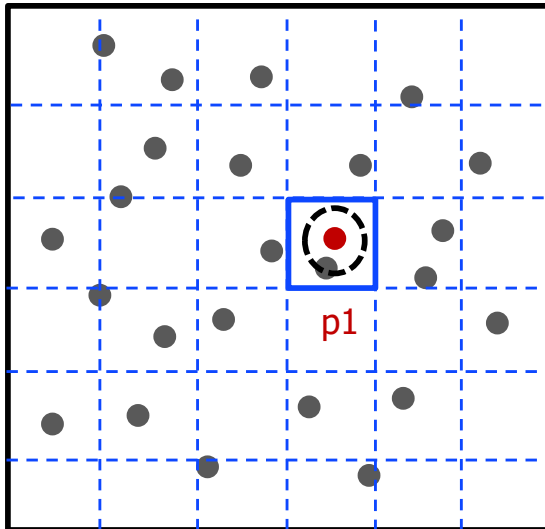


Grid-based Spatial Indexing

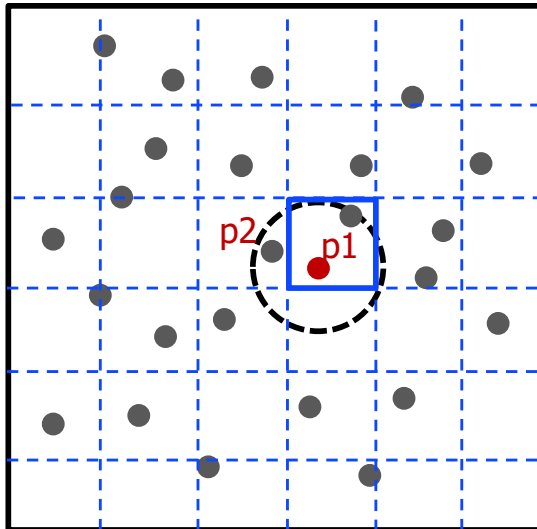
❖ Nearest neighbor query

- Euclidian distance
- Road network distance is quite different

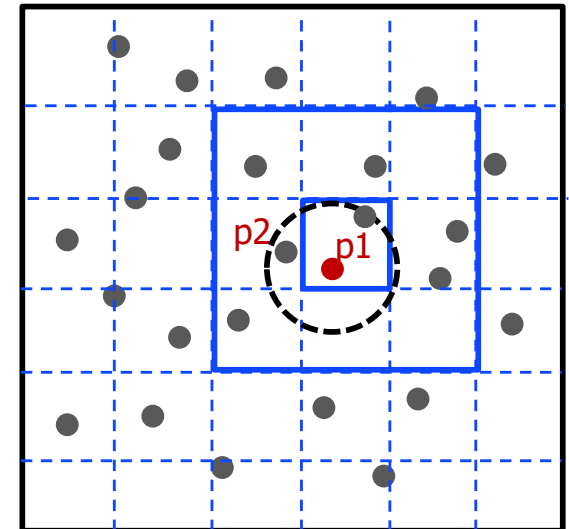
The nearest object is within the grid



The nearest object is outside the grid



Fast approximation





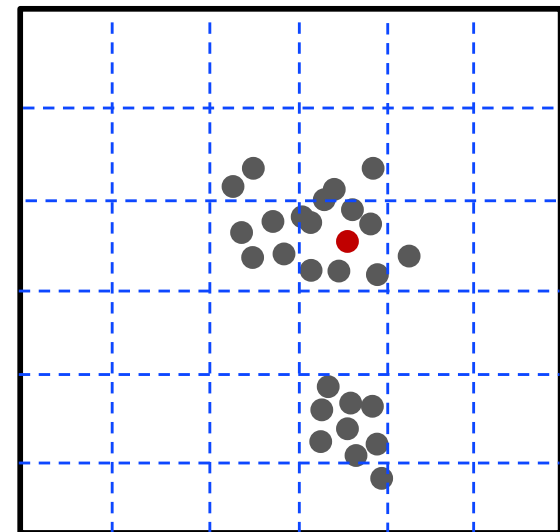
Grid-based Spatial Indexing

❖ Advantages

- Easy to implement and understand
- Very efficient for processing **range and nearest queries**

❖ Disadvantages

- Index size could be big
- Difficult to deal with unbalanced data

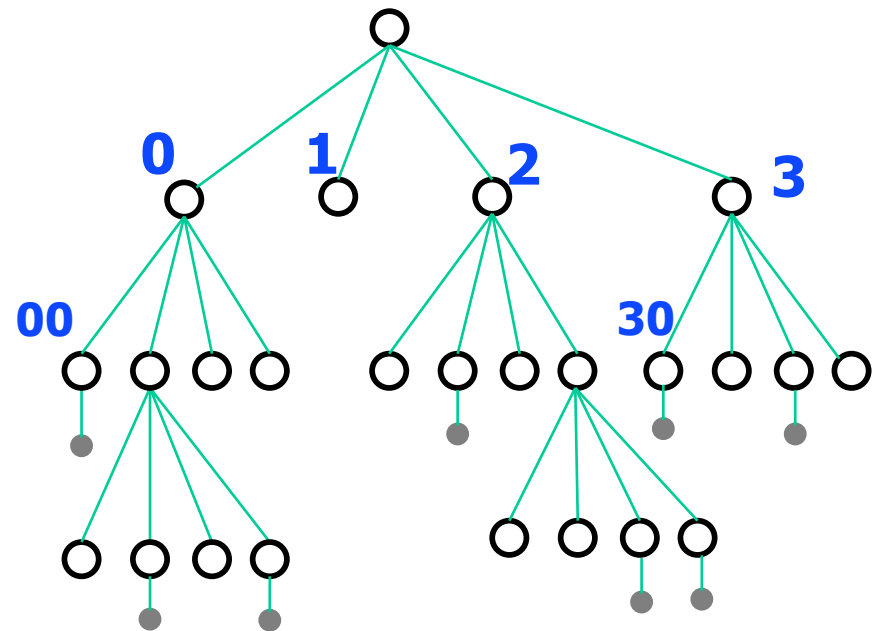
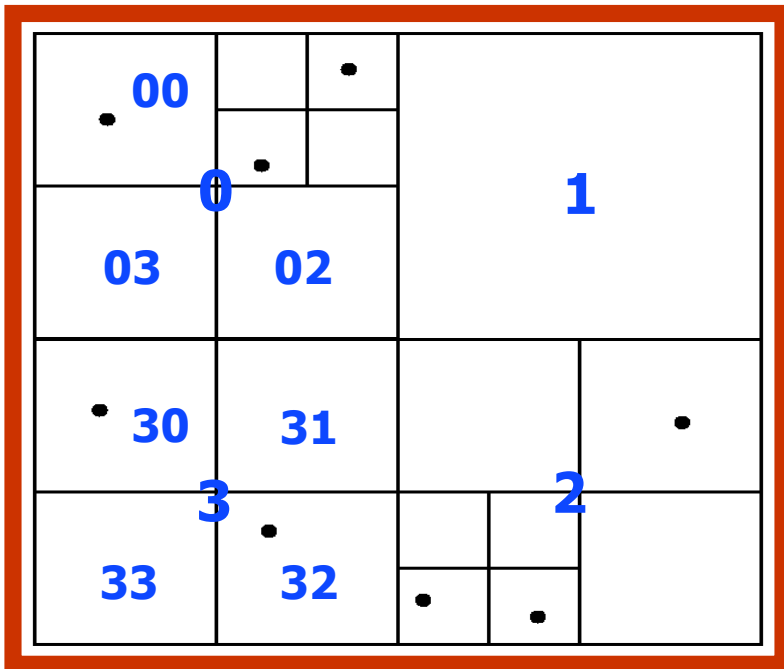


Quad-Tree



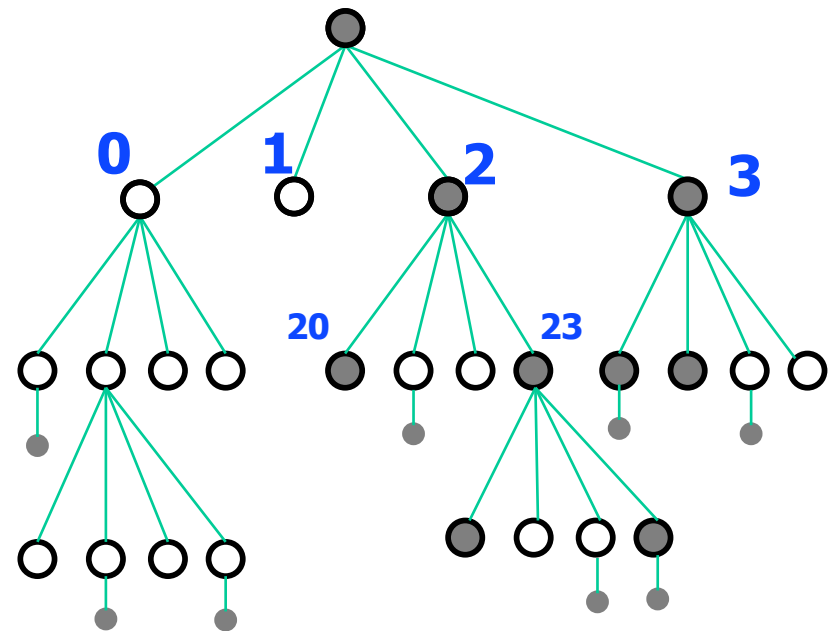
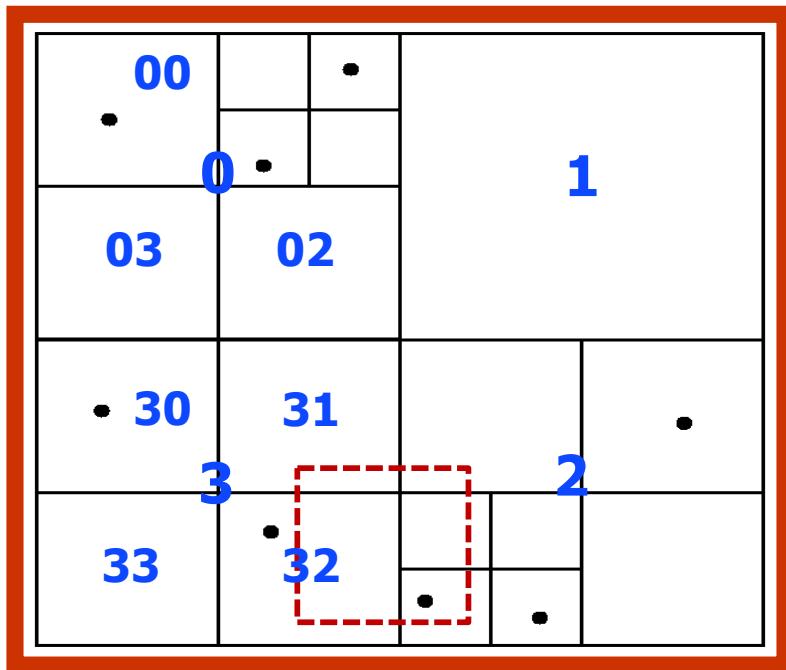
- **Indexing**

- Each node of a quad-tree is associated with a rectangular region of space; the top node is associated with the entire target space.
- Each non-leaf node divides its region into four equal sized quadrants
- Leaf nodes have between zero and some fixed maximum number of points (set to 1 in example).



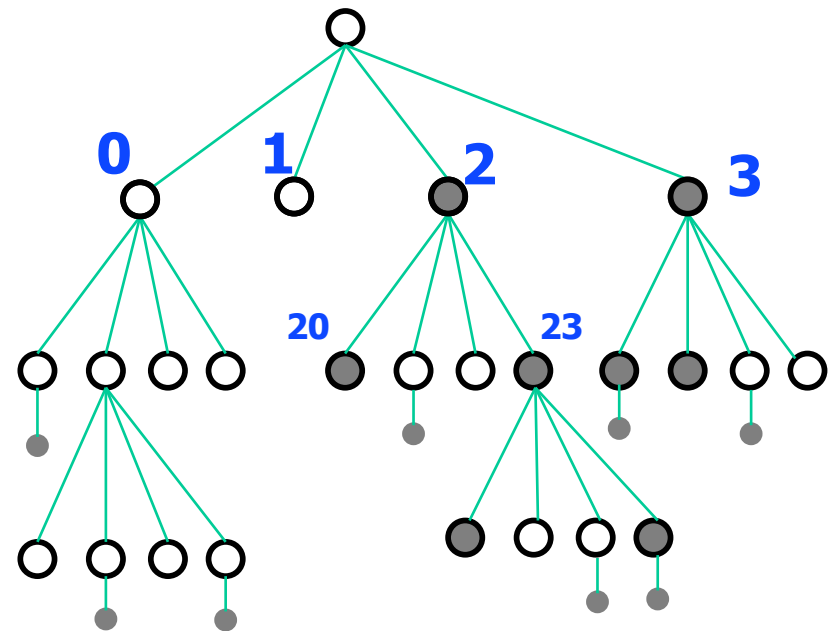
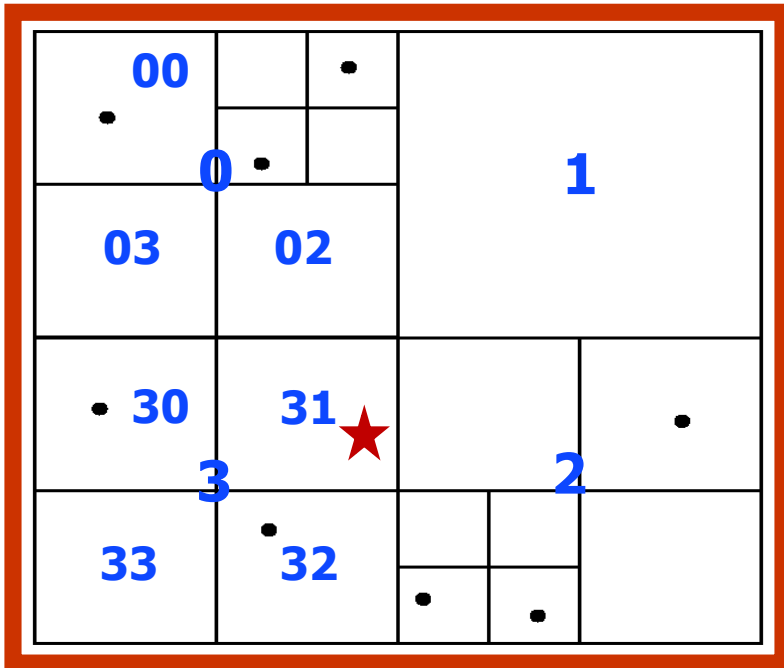
Quad-Tree

- Range query (ok)



Quad-Tree

- Nearest Neighbour Query (hard)



Spatial/Temporal (3D) Indexing Structures

- ❖ Temporal Indexing (1-D data)
 - List index
 - B-tree
- ❖ Space Partition-Based Indexing Structures (2-D data)
 - Grid-based
 - Quad-tree

Questions?