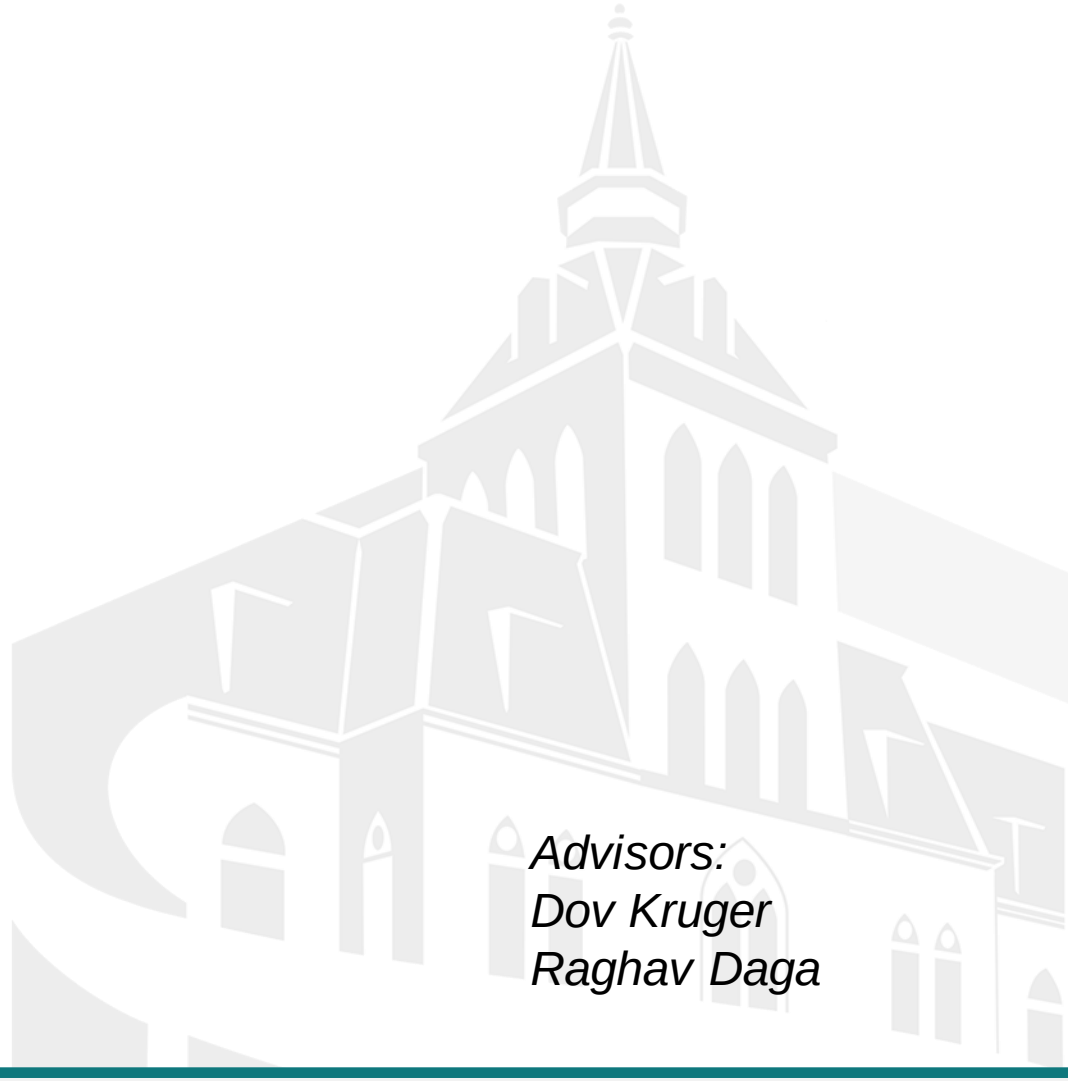
STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Chain

Noah Malhi
Prashant Kumar
Quentin Jimenez



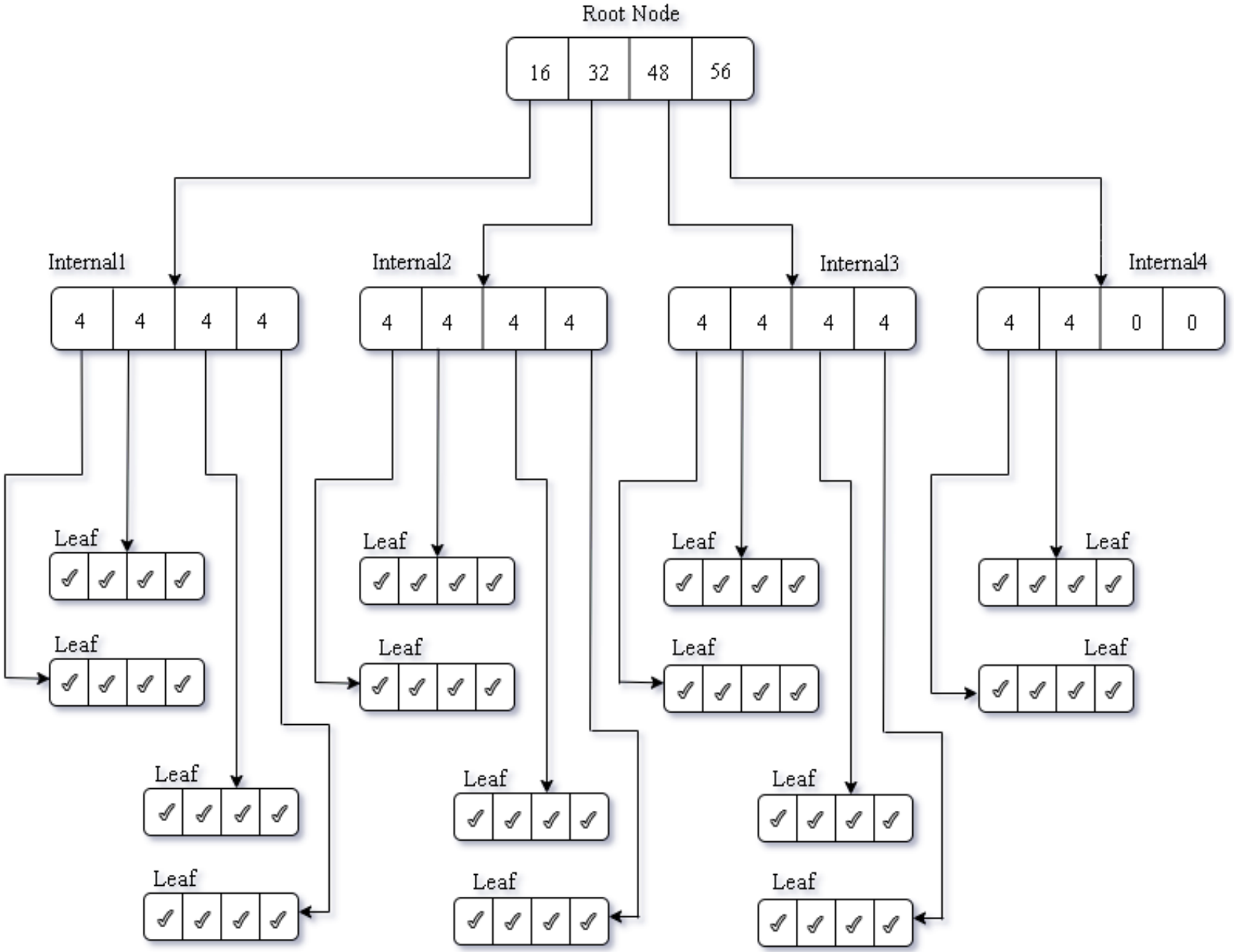
Advisors:
Dov Kruger
Raghav Daga

Problem Statement

- The String Class:
  - Array based object for streaming char
  - Insertion becomes a problem for large objects
- Chain is an efficient tree structure for reading extremely large files.
- Chain is a hybrid concept by Prof. Dov Kruger.
- Leaves of the tree are lines, and each individual line is a conventional string.

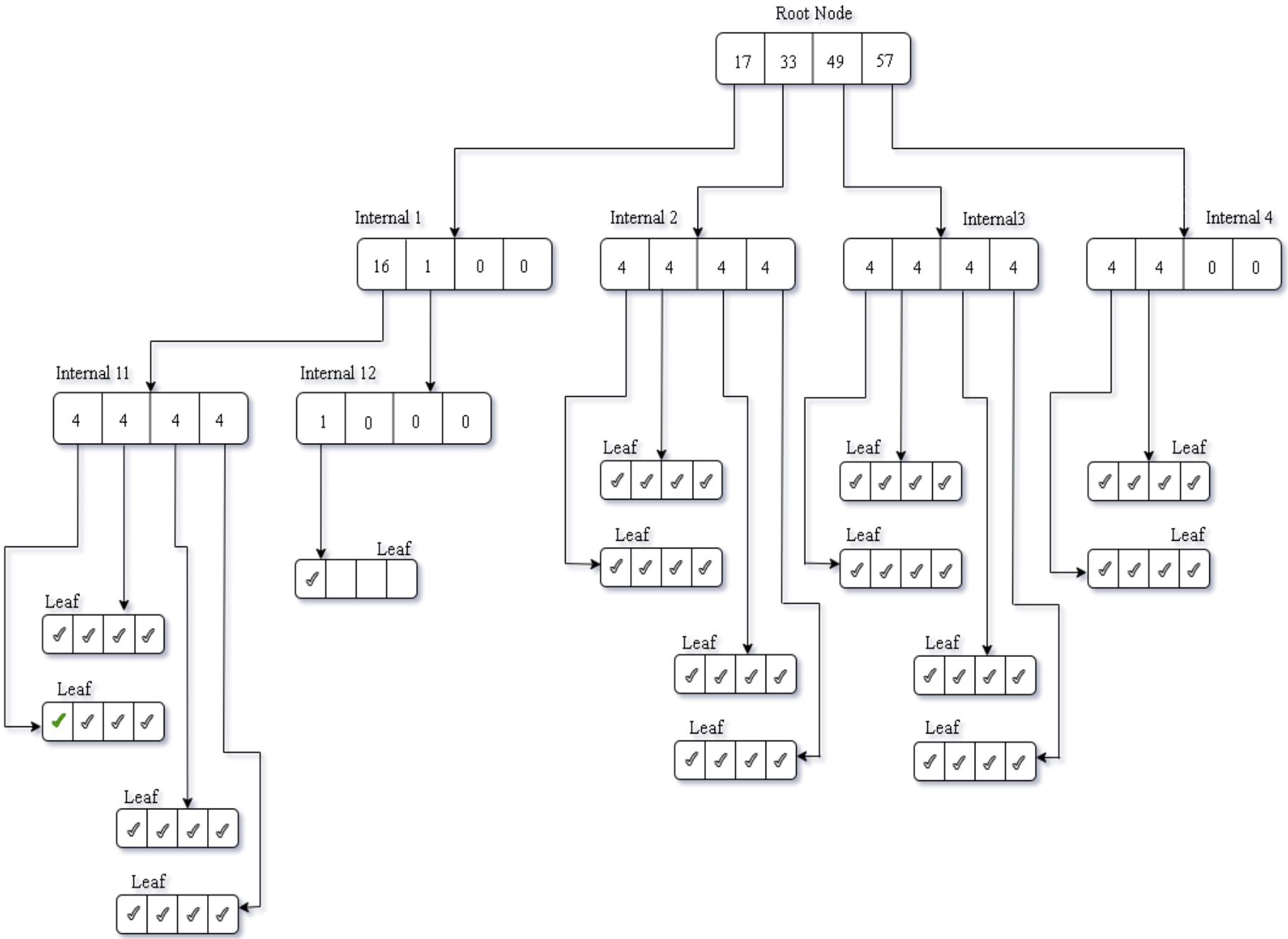
Structure

- M-ary Trees.
- Rules:
  at most m children
  Special cases:
  binary , ternary tree
- Root Node-
  Pointer to internal nodes
  counter for #lines
- Internal Nodes-
  Pointer to internal/leaf
  counter for #lines
- Leaf Nodes
  Pointer to String
  (which is a line)



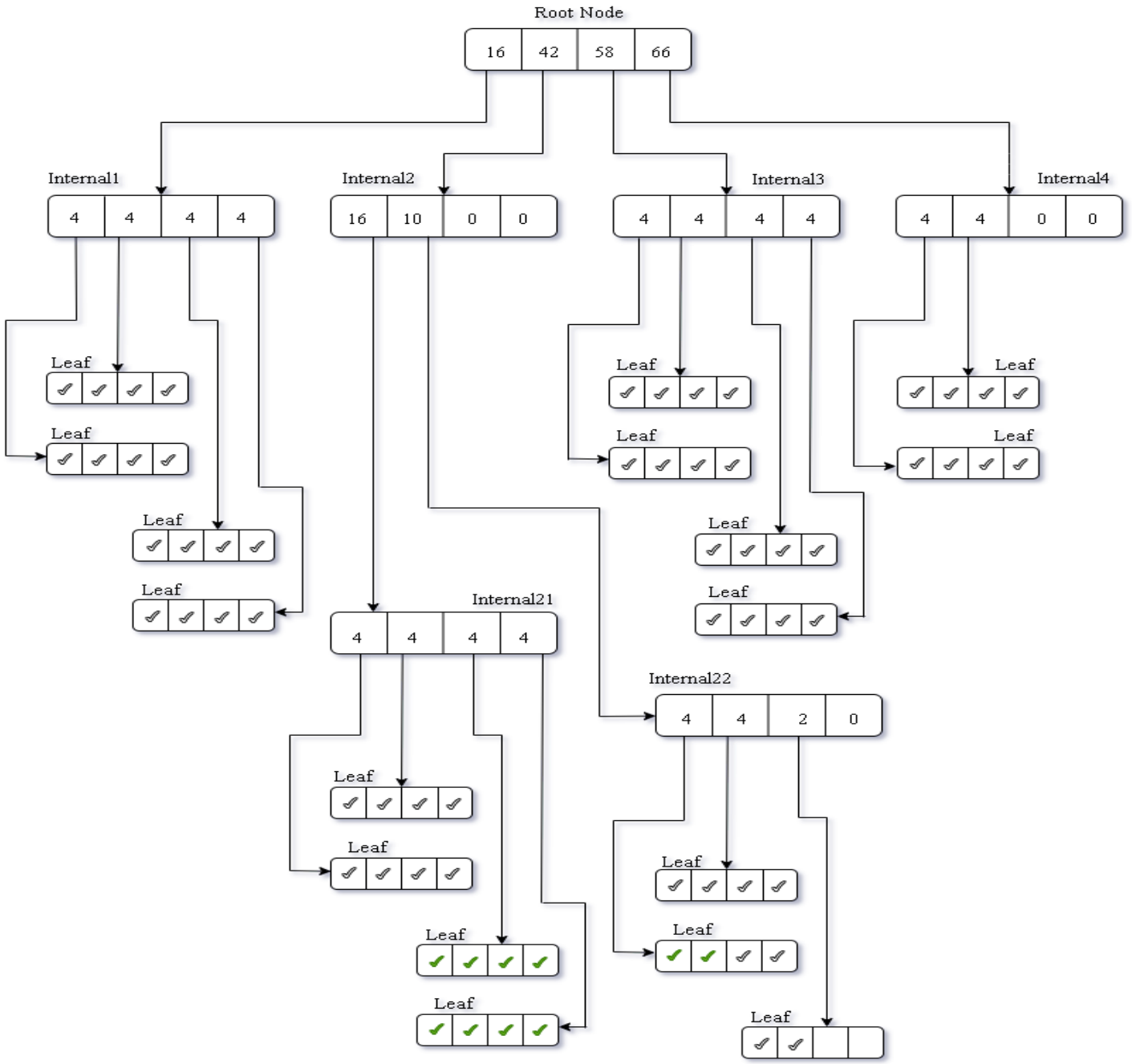
Insert start

- Check space in Internal1
- Split Internal1
- Insert at pos1
- Shift everything to right



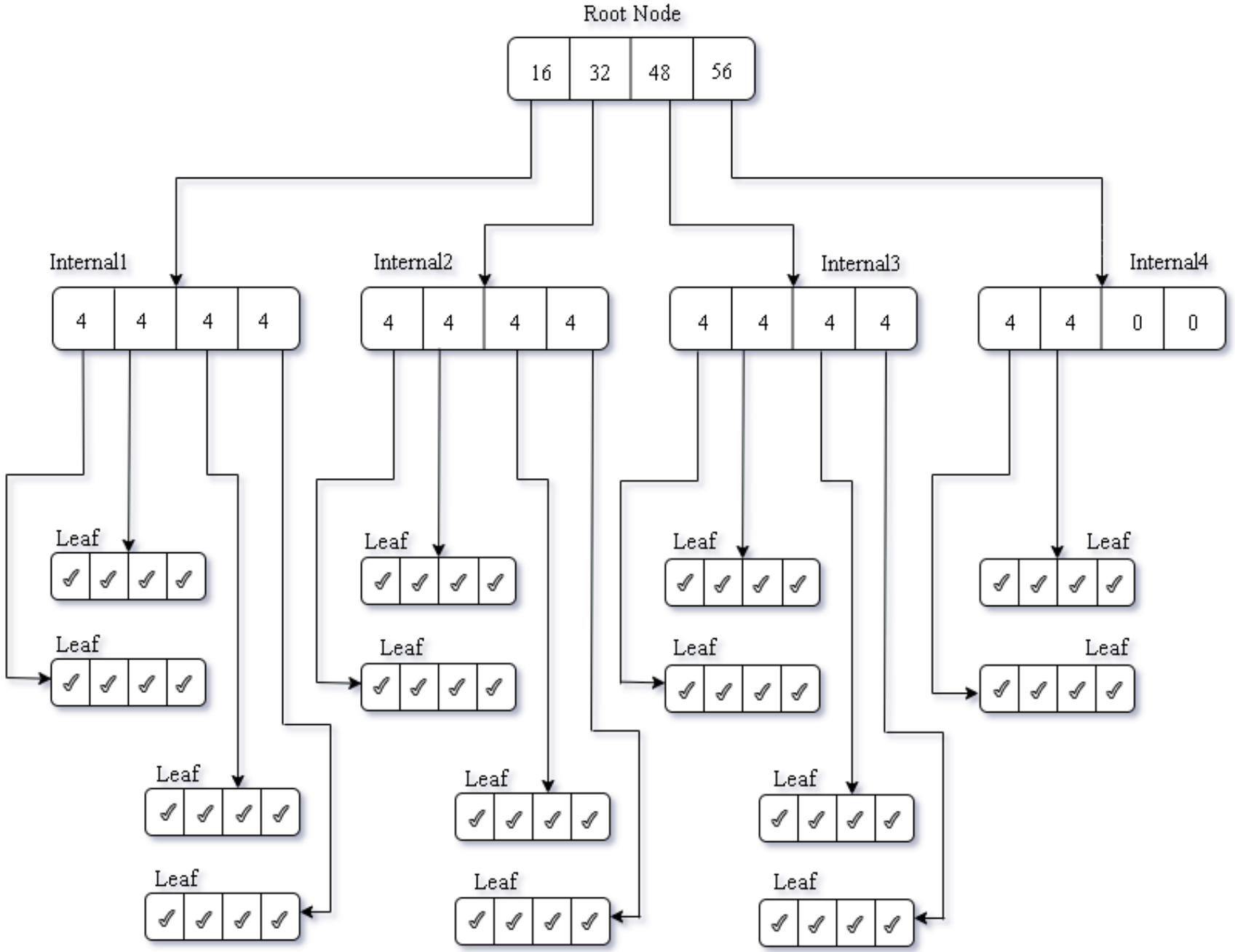
Insert middle

- Insert 10 lines at pos 25



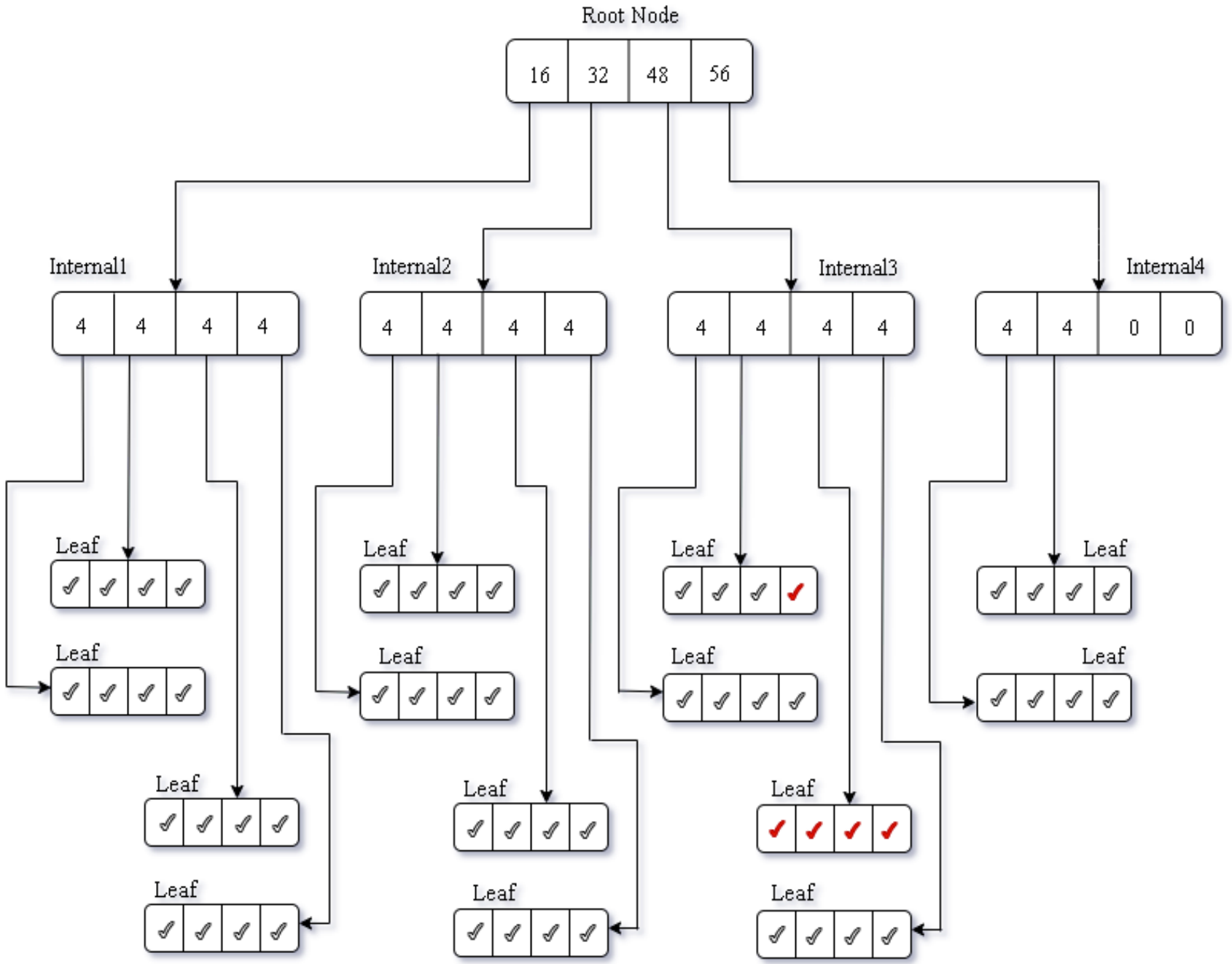
Base Struct: refer.

- M-ary Trees.
- Rules:
  at most m children
  Special cases:
  binary , ternary tree
- Root Node-
  Pointer to internal nodes
  counter for #lines
- Internal Nodes-
  Pointer to internal/leaf
  counter for #lines
- Leaf Nodes
  Pointer to String
  (which is a line)

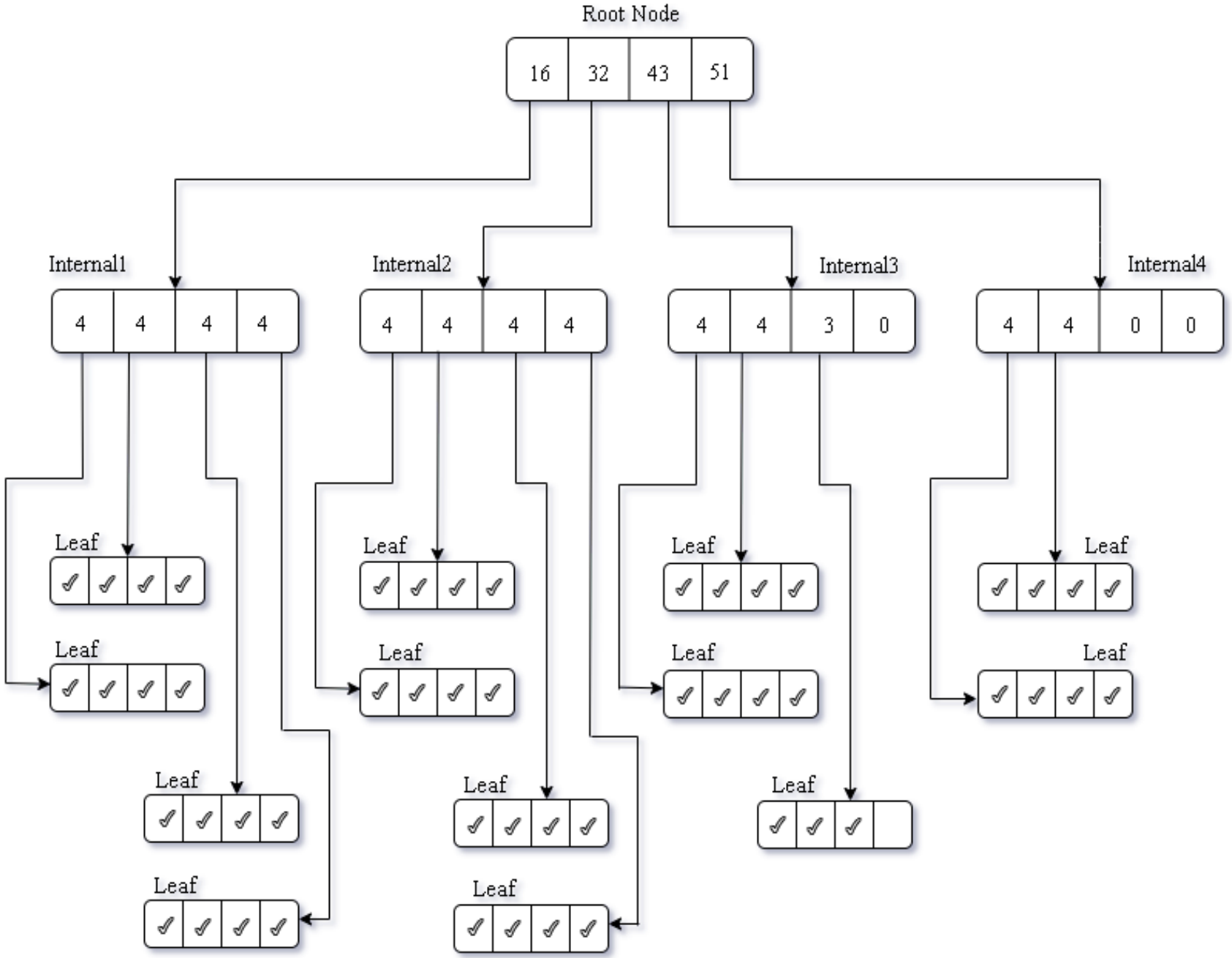


Delete-1

- Delete 5 lines at pos 40

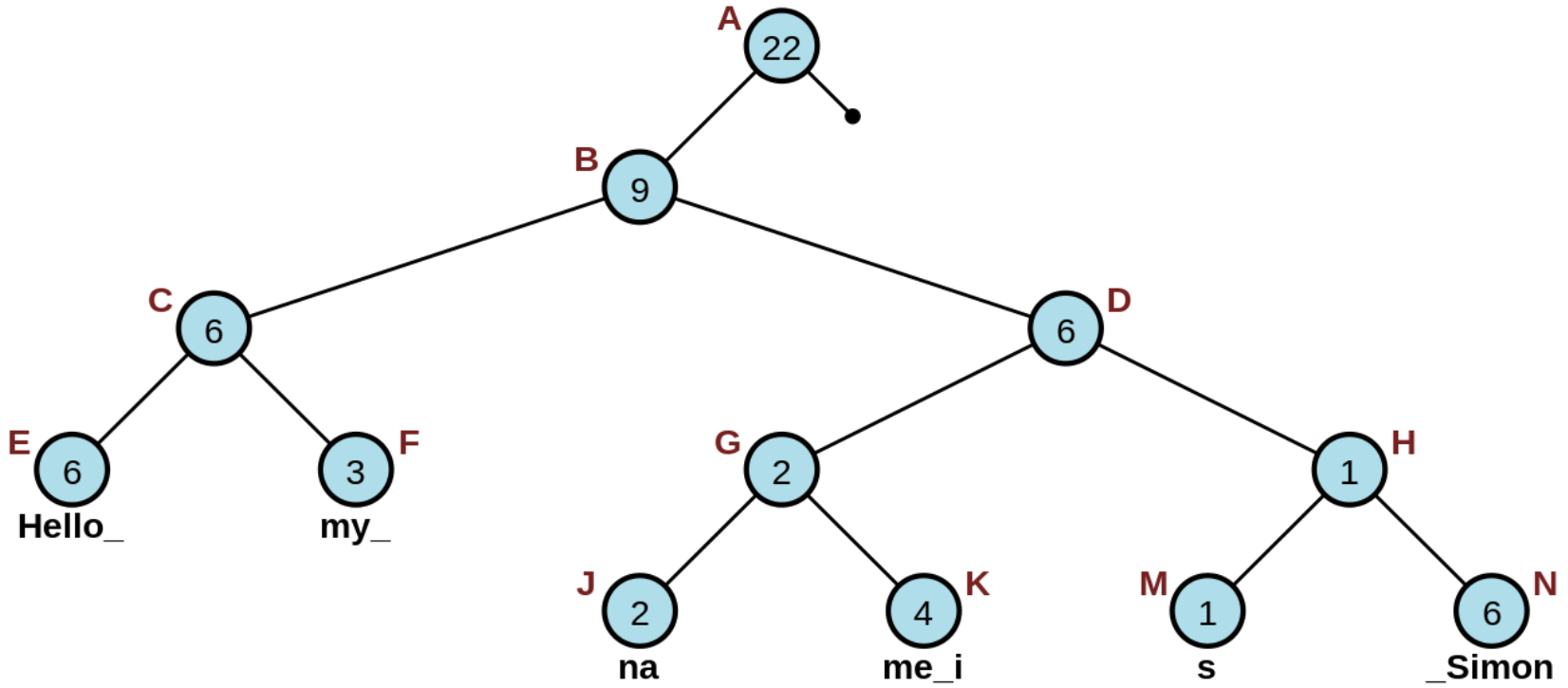


Delete-2


Root Node
16
32
43
51
Internal1
4
4
4
4
Internal2
4
4
4
4
Internal3
4
4
3
0
Internal4
4
4
0
0
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf
Leaf

Literature

- New concept by Professor Dov Kruger
- Ropes: Boehm and Russ
- Binary tree
- Leaf nodes hold strings
- Internal Nodes hold length of the strings to the left
- Sample rope:
  (image: wikipedia)
  String: Hello my name is Simon



How are we different?

- Each line is a node, so we use about an order of magnitude lower number of nodes.
- Ropes are binary trees. The depth cannot be controlled.
- Chain can be higher order trees, leading to very fat tree structures.

Progress

- Issues, TODO

Complexity Analysis

- Assumption: Each word in English language is 8-10 char. Average line length is 8-10 words.
- String Size= N characters
- Rope- n nodes
- M-ary Tree with n nodes (n is diff for rope and chain)
- Example: 100MB text file
  1 char = 1 byte
  1MB = 1 millions char
  100MB = 100 million char/ 10 million words/ 1 million lines.
- #nodes_rope ~ 10 million, #nodes_chain ~ 1 million

	String	Rope	Chain
Insert	O(N)	Olog(n)	Olog(n)
Delete	O(N)	Olog(n)	Olog(n)
Search	O(N)	Olog(n)	Olog(n)

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Thank You!

Questions?

Noah Malhi
Prashant Kumar
Quentin Jimenez