Definition of a B-Tree

A B-tree of order $d$ is a multi-way search tree $T$ with the following properties:

1. The root of $T$ has at least 2 children and at most $d$ children.
2. All internal nodes of $T$ (other than the root) have between $\lceil d/2 \rceil$ and $d$ children.
3. All external nodes of $T$ are at the same level.
Insertion in a B-Tree

Add the new key to the appropriate node

Overflow?

Split the node into two nodes on the same level, and promote median key
Deletion from a B-Tree

Remove the key, swapping it with the smallest key larger than it, if needed.

Underflow?

Transfer a key from a sibling, if possible. Otherwise perform a fusion operation.
External Search

- Main memory is several orders of magnitude faster than disk. Thus the main goal of maintaining a dictionary in disk is to minimize the number of disk accesses.
- To try to minimize the time needed to transfer information from disk to main memory, data items on a disk are grouped into contiguous sections called blocks.
- B-trees are used to implement dictionaries in external memory (disk).
- The order of a B-tree is chosen so that the size of a node is smaller than the size of a disk block, but as close as possible.
Height of a B-tree

- The height of a B-tree gives an upper bound on the maximum number of disk accesses required to access information stored in it.
- The height of a B-tree of order $d$ is
  \[ O(\log[d/2] n), \]
  where $n$ is the number of keys stored in the tree.
- For example, the height of a B-tree of order 200 storing one million keys, is only 3.