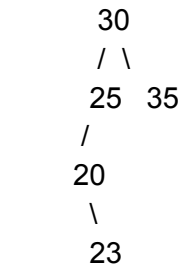
DS 4300 - Spring 2025 Sample Midterm Questions (& HW 04)

Below are the answers to the sample exam questions:

1. **Difference between contiguously allocated and linked lists:**
    - A contiguously allocated list (array) stores elements in a single block of memory, allowing O(1) access time for indexing but requiring resizing if the array fills up. A linked list consists of nodes that are individually allocated and connected via pointers, allowing dynamic memory allocation but requiring O(n) time for indexing.
2. **When are linked lists faster than contiguously-allocated lists?**
    - Linked lists are faster when frequent insertions and deletions occur, particularly in the middle of the list, because they do not require shifting elements like arrays do.
3. **AVL Tree insertion of 23 and imbalance case:**
    - Inserting 23 into the AVL tree:



4.
    - The imbalance occurs at node 25, forming a Left-Right (LR) case. This requires a left rotation on 20 followed by a right rotation on 30 to rebalance.
5. **Why is a B+ Tree better than an AVL tree for indexing large datasets?**
    - A B+ Tree is optimized for disk storage because it has a higher branching factor, reducing the number of disk accesses required. Unlike an AVL tree, all data is stored in leaf nodes, making range queries and sequential access more efficient.
6. **What is disk-based indexing and why is it important for database systems?**
    - Disk-based indexing structures data efficiently on disk to minimize expensive disk I/O operations. It is crucial for databases because accessing disk storage is significantly slower than accessing RAM.
7. **What is a transaction in a relational database system?**

    - A transaction is a sequence of database operations that are executed as a single unit of work, ensuring data integrity.
8. **Four components of ACID transactions:**
    - **Atomicity:** Ensures transactions are all-or-nothing.
    - **Consistency:** Guarantees the database remains in a valid state.
    - **Isolation:** Prevents transactions from interfering with each other.
    - **Durability:** Ensures committed transactions persist even after failures.
9. **Why does the CAP principle not apply to a single-node MongoDB instance?**
    - The CAP theorem applies to distributed systems, where consistency, availability, and partition tolerance must be balanced. A single-node MongoDB instance does not experience network partitions, making CAP irrelevant.
10. **Differences between horizontal and vertical scaling:**
    - **Horizontal scaling:** Adding more machines to distribute the load.
    - **Vertical scaling:** Upgrading the hardware of a single machine.
11. **How a key/value store can be used as a feature store:**
    - A key/value store allows efficient retrieval of precomputed features for machine learning models, ensuring low-latency data access.
12. **When was Redis originally released?**
    - Redis was first released in 2009.
13. **Difference between INC and INCR commands in Redis:**
    - Redis does not have an `INC` command. The `INCR` command atomically increments a key's value by one.
14. **Benefits of BSON over JSON in MongoDB:**
    - BSON supports additional data types, including binary data and timestamps, and is more efficient in parsing and storage than JSON.

MongoDB query for suspense movies released between 2010 and 2015:

```
 db.movies.find({
   release_year: { $gte: 2010, $lte: 2015 },
   genre: "suspense"
}, {
   title: 1,
   _id: 0
});
```

15.

16. **What does the `$nin` operator mean in a MongoDB query?**

    - The `$nin` (not in) operator filters out documents where the specified field does not contain any of the listed values.