Hashing—Collision Resolution – Other Techniques

Double Hashing

- Research shows that with progressive overflow collision resolution records tend to cluster around certain addresses
- With double hashing when a collision occurs, a second hash function is applied to get a second number *c* having no divisors in common with the first hash address.
    - This number is then added to the original address.
    - This process is continued until a free address is found.
- This technique can sometimes moves the records too far away from the original address which can cause to much movement of the disk head

Chained Progressive Overflow

- Use pointers to link overflow records together
- So each set of synonyms is a linked list
- All the issues of tombstones go away and
- Search length will be shorter because we don't have to look at records that couldn't possibly be the one we are looking for
- You have to make sure that any record that will be hashed to doesn't get a synonym
    - remember that there are addresses that will not get hashed to
- Two—pass loading can take care of this if we already have all the records
    - Save the synonyms and reload them after all records have been tried
- There isn't a clever way if we don't have all the records already

Chaining with a Separate Overflow Area

- Move all the overflow records to a separate area
- Chain the overflow records for a specific address together
- This doesn't really work well because often the overflow area ends up on a different cylinder which takes a lot of extra time to access
- You have to use this technique if you expect the file to grow beyond the original address space size