

Hash table

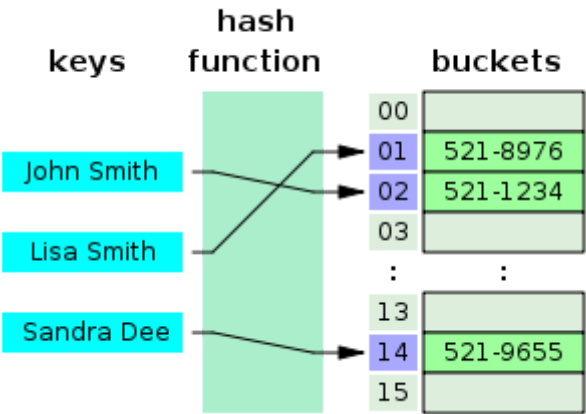
In computing, a **hash table** (**hash map**) is a data structure that implements an associative array abstract data type, a structure that can map keys to values. A hash table uses a hash function to compute an index, also called a *hash code*, into an array of *buckets* or *slots*, from which the desired value can be found. During lookup, the key is hashed and the resulting hash indicates where the corresponding value is stored.

Ideally, the hash function will assign each key to a unique bucket, but most hash table designs employ an imperfect hash function, which might cause hash *collisions* where the hash function generates the same index for more than one key. Such collisions are typically accommodated in some way.

In a well-dimensioned hash table, the average cost (number of instructions) for each lookup is independent of the number of elements stored in the table. Many hash table designs also allow arbitrary insertions and deletions of key-value pairs, at (amortized^[2]) constant average cost per operation.^{[3][4]}

In many situations, hash tables turn out to be on average more efficient than search trees or any other table lookup structure. For this reason, they are widely used in many kinds of computer software, particularly for associative arrays, database indexing, caches, and sets.

Hash table		
<u>Type</u>	Unordered <u>associative array</u>	
Invented	1953	
<u>Time complexity in big O notation</u>		
Algorithm	Average	Worst case
Space	$O(n)$ ^[1]	$O(n)$
Search	$O(1)$	$O(n)$
Insert	$O(1)$	$O(n)$
Delete	$O(1)$	$O(n)$



A small phone book as a hash table

Contents

Hashing

Choosing a hash function

Perfect hash function

Key statistics

Collision resolution

Separate chaining

Separate chaining with linked lists

Separate chaining with list head cells

Separate chaining with other structures

Open addressing

Coalesced hashing

Cuckoo hashing

[Hopscotch hashing](#)

[Robin Hood hashing](#)

[2-choice hashing](#)

[Dynamic resizing](#)

[Resizing by copying all entries](#)

[Alternatives to all-at-once rehashing](#)

[Incremental resizing](#)

[Monotonic keys](#)

[Linear hashing](#)

[Hashing for distributed hash tables](#)

[Performance](#)

[Speed analysis](#)

[Memory utilization](#)

[Features](#)

[Advantages](#)

[Drawbacks](#)

[Uses](#)

[Associative arrays](#)

[Database indexing](#)

[Caches](#)

[Sets](#)

[Object representation](#)

[Unique data representation](#)

[Transposition table](#)

[Implementations](#)

[In programming languages](#)

[History](#)

[See also](#)

[Related data structures](#)

[References](#)

[Further reading](#)

[External links](#)

Hashing

The idea of hashing is to distribute the entries (key/value pairs) across an array of *buckets*. Given a key, the algorithm computes an *index* that suggests where the entry can be found:

```
index = f(key, array_size)
```

Often this is done in two steps:

```
hash = hashfunc(key)
index = hash % array_size
```

In this method, the *hash* is independent of the array size, and it is then *reduced* to an index (a number between 0 and `array_size - 1`) using the modulo operator (%).

In the case that the array size is a power of two, the remainder operation is reduced to masking, which improves speed, but can increase problems with a poor hash function.^[5]

Choosing a hash function

A basic requirement is that the function should provide a uniform distribution of hash values. A non-uniform distribution increases the number of collisions and the cost of resolving them. Uniformity is sometimes difficult to ensure by design, but may be evaluated empirically using statistical tests, e.g., a Pearson's chi-squared test for discrete uniform distributions.^{[6][7]}

The distribution needs to be uniform only for table sizes that occur in the application. In particular, if one uses dynamic resizing with exact doubling and halving of the table size, then the hash function needs to be uniform only when the size is a power of two. Here the index can be computed as some range of bits of the hash function. On the other hand, some hashing algorithms prefer to have the size be a prime number.^[8] The modulus operation may provide some additional mixing; this is especially useful with a poor hash function.

For open addressing schemes, the hash function should also avoid *clustering*, the mapping of two or more keys to consecutive slots. Such clustering may cause the lookup cost to skyrocket, even if the load factor is low and collisions are infrequent. The popular multiplicative hash^[3] is claimed to have particularly poor clustering behavior.^[8]

Cryptographic hash functions are believed to provide good hash functions for any table size, either by modulo reduction or by bit masking. They may also be appropriate if there is a risk of malicious users trying to sabotage a network service by submitting requests designed to generate a large number of collisions in the server's hash tables. However, the risk of sabotage can also be avoided by cheaper methods (such as applying a secret salt to the data, or using a universal hash function). A drawback of cryptographic hashing functions is that they are often slower to compute, which means that in cases where the uniformity for *any* size is not necessary, a non-cryptographic hashing function might be preferable.

Perfect hash function

If all keys are known ahead of time, a perfect hash function can be used to create a perfect hash table that has no collisions. If minimal perfect hashing is used, every location in the hash table can be used as well.

Perfect hashing allows for constant time lookups in all cases. This is in contrast to most chaining and open addressing methods, where the time for lookup is low on average, but may be very large, $O(n)$, for instance when all the keys hash to a few values.

Key statistics

A critical statistic for a hash table is the *load factor*, defined as

$$\text{load factor} = \frac{n}{k},$$

where

- n is the number of entries occupied in the hash table.
- k is the number of buckets.

As the load factor grows larger, the hash table becomes slower, and it may even fail to work (depending on the method used). The expected constant time property of a hash table assumes that the load factor be kept below some bound. For a *fixed* number of buckets, the time for a lookup grows with the number of entries, and therefore the desired constant time is not achieved. In some implementations, the solution is to automatically grow (usually, double) the size of the table when the load factor bound is reached, thus forcing to re-hash all entries. As a real-world example, the default load factor for a HashMap in Java 10 is 0.75, which "offers a good trade-off between time and space costs."^[9]

Second to the load factor, one can examine the variance of number of entries per bucket. For example, two tables both have 1,000 entries and 1,000 buckets; one has exactly one entry in each bucket, the other has all entries in the same bucket. Clearly the hashing is not working in the second one.

A low load factor is not especially beneficial. As the load factor approaches 0, the proportion of unused areas in the hash table increases, but there is not necessarily any reduction in search cost. This results in wasted memory.

Collision resolution

Hash collisions are practically unavoidable when hashing a random subset of a large set of possible keys. For example, if 2,450 keys are hashed into a million buckets, even with a perfectly uniform random distribution, according to the birthday problem there is approximately a 95% chance of at least two of the keys being hashed to the same slot.

Therefore, almost all hash table implementations have some collision resolution strategy to handle such events. Some common strategies are described below. All these methods require that the keys (or pointers to them) be stored in the table, together with the associated values.

Separate chaining

In the method known as *separate chaining*, each bucket is independent, and has some sort of list of entries with the same index. The time for hash table operations is the time to find the bucket (which is constant) plus the time for the list operation.

In most implementations buckets will have few entries, if the hash function is working properly. Therefore, structures that are efficient in time and space for these cases are preferred. Structures that are efficient for a fairly large number of entries per bucket are not needed or desirable. If these cases happen often, the hashing function needs to be fixed.^[10]

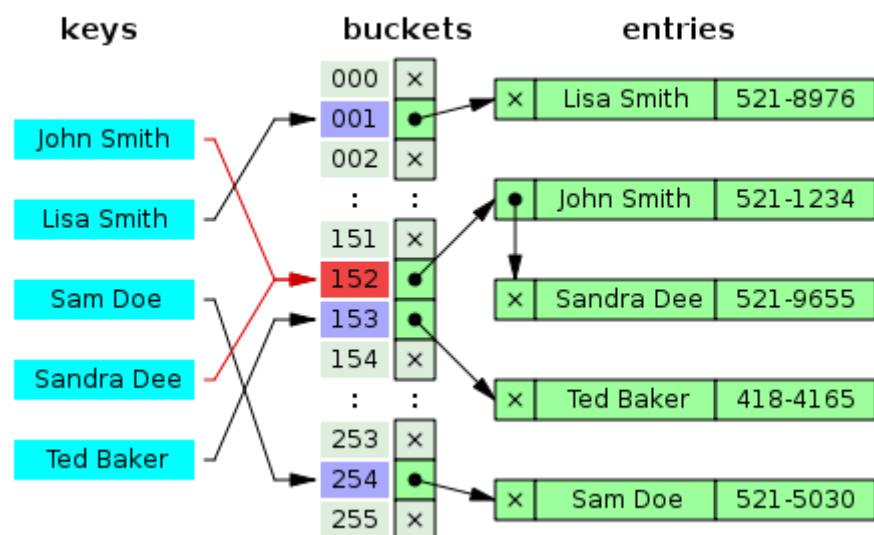
There are some implementations^[11] which give excellent performance for both time and space, with the average number of elements per bucket ranging between 5 and 100.

Separate chaining with linked lists

Chained hash tables with linked lists are popular because they require only basic data structures with simple algorithms, and can use simple hash functions that are unsuitable for other methods.

The cost of a table operation is that of scanning the entries of the selected bucket for the desired key. If the distribution of keys is sufficiently uniform, the *average* cost of a lookup depends only on the average number of keys per bucket—that is, it is roughly proportional to the load factor.

For this reason, chained hash tables remain effective even when the number of table entries n is much higher than the number of slots. For example, a chained hash table with 1000 slots and 10,000 stored keys (load factor 10) is five to ten times slower than a 10,000-slot table (load factor 1); but still 1000 times faster than a plain sequential list.



Hash collision resolved by separate chaining.

For separate-chaining, the worst-case scenario is when all entries are inserted into the same bucket, in which case the hash table is ineffective and the cost is that of searching the bucket data structure. If the latter is a linear list, the lookup procedure may have to scan all its entries, so the worst-case cost is proportional to the number n of entries in the table.

The bucket chains are often searched sequentially using the order the entries were added to the bucket. If the load factor is large and some keys are more likely to come up than others, then rearranging the chain with a move-to-front heuristic may be effective. More sophisticated data structures, such as balanced search trees, are worth considering only if the load factor is large (about 10 or more), or if the hash distribution is likely to be very non-uniform, or if one must guarantee good performance even in a worst-case scenario. However, using a larger table and/or a better hash function may be even more effective in those cases.

Chained hash tables also inherit the disadvantages of linked lists. When storing small keys and values, the space overhead of the `next` pointer in each entry record can be significant. An additional disadvantage is that traversing a linked list has poor cache performance, making the processor cache ineffective.

Separate chaining with list head cells

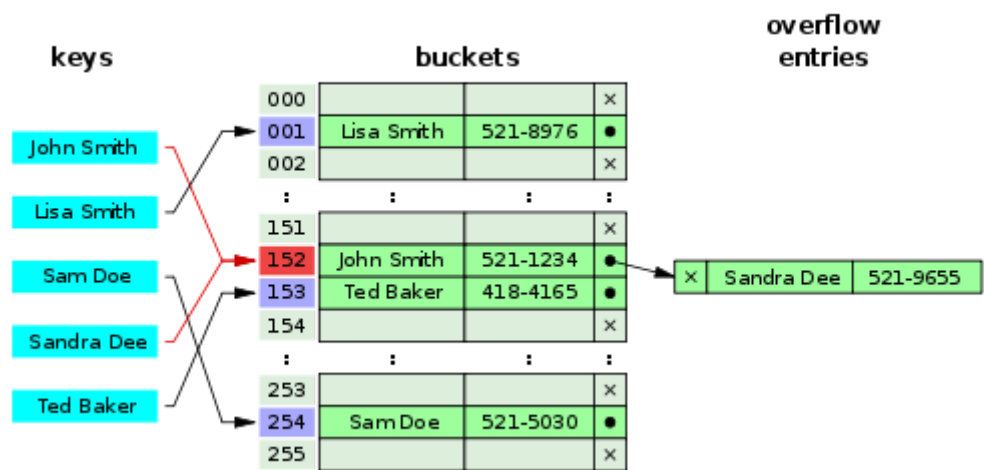
Some chaining implementations store the first record of each chain in the slot array itself.^[4] The number of pointer traversals is decreased by one for most cases. The purpose is to increase cache efficiency of hash table access.

The disadvantage is that an empty bucket takes the same space as a bucket with one entry. To save space, such hash tables often have about as many slots as stored entries, meaning that many slots have two or more entries.

Separate chaining with other structures

Instead of a list, one can use any other data structure that supports the required operations. For example, by using a self-balancing binary search tree, the theoretical worst-case time of common hash table operations (insertion, deletion, lookup) can be brought down to $O(\log n)$ rather than $O(n)$. However, this introduces extra

complexity into the implementation, and may cause even worse performance for smaller hash tables, where the time spent inserting into and balancing the tree is greater than the time needed to perform a linear search on all of the elements of a list.^{[3][12]} A real world example of a hash table that uses a self-balancing binary search tree for buckets is the `HashMap` class in Java version 8.^[13]



Hash collision by separate chaining with head records in the bucket array.

The variant called array hash table uses a dynamic array to store all the entries that hash to the same slot.^{[14][15][16]} Each newly inserted entry gets appended to the end of the dynamic array that is assigned to the slot. The dynamic array is resized in an *exact-fit* manner, meaning it is grown only by as many bytes as needed. Alternative techniques such as growing the array by block sizes or *pages* were found to improve insertion performance, but at a cost in space. This variation makes more efficient use of CPU caching and the translation lookaside buffer (TLB), because slot entries are stored in sequential memory positions. It also dispenses with the `next` pointers that are required by linked lists, which saves space. Despite frequent array resizing, space overheads incurred by the operating system such as memory fragmentation were found to be small.

An elaboration on this approach is the so-called dynamic perfect hashing,^[17] where a bucket that contains k entries is organized as a perfect hash table with k^2 slots. While it uses more memory (n^2 slots for n entries, in the worst case and $n \times k$ slots in the average case), this variant has guaranteed constant worst-case lookup time, and low amortized time for insertion. It is also possible to use a fusion tree for each bucket, achieving constant time for all operations with high probability.^[18]

Open addressing

In another strategy, called open addressing, all entry records are stored in the bucket array itself. When a new entry has to be inserted, the buckets are examined, starting with the hashed-to slot and proceeding in some *probe sequence*, until an unoccupied slot is found. When searching for an entry, the buckets are scanned in the same sequence, until either the target record is found, or an unused array slot is found, which indicates that there is no such key in the table.^[19] The name "open addressing" refers to the fact that the location ("address") of the item is not determined by its hash value. (This method is also called **closed hashing**; it should not be confused with "open hashing" or "closed addressing" that usually mean separate chaining.)

Well-known probe sequences include:

- Linear probing, in which the interval between probes is fixed (usually 1). Because of good CPU cache utilization and high performance this algorithm is most widely used on modern computer architectures in hash table implementations.^[20]
- Quadratic probing, in which the interval between probes is increased by adding the successive outputs of a quadratic polynomial to the starting value given by the original hash computation

- **Double hashing**, in which the interval between probes is computed by a second hash function

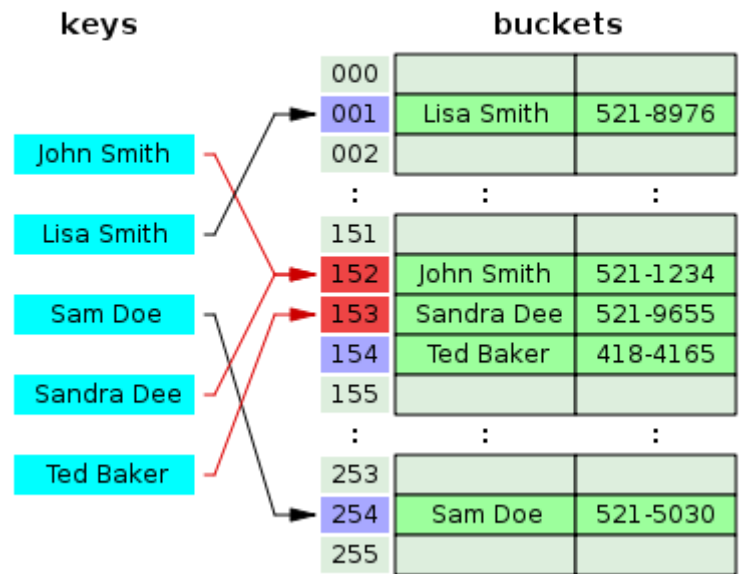
A drawback of all these open addressing schemes is that the number of stored entries cannot exceed the number of slots in the bucket array. In fact, even with good hash functions, their performance dramatically degrades when the load factor grows beyond 0.7 or so. For many applications, these restrictions mandate the use of dynamic resizing, with its attendant costs.

Open addressing schemes also put more stringent requirements on the hash function: besides distributing the keys more uniformly over the buckets, the function must also minimize the clustering of hash values that are consecutive in the probe order. Using separate chaining, the only concern is that too many objects map to the *same* hash value; whether they are adjacent or nearby is completely irrelevant.

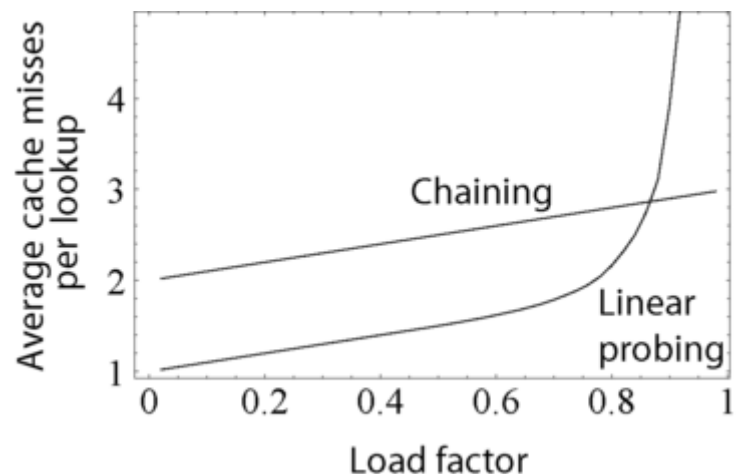
Open addressing only saves memory if the entries are small (less than four times the size of a pointer) and the load factor is not too small. If the load factor is close to zero (that is, there are far more buckets than stored entries), open addressing is wasteful even if each entry is just two words.

Open addressing avoids the time overhead of allocating each new entry record, and can be implemented even in the absence of a memory allocator. It also avoids the extra indirection required to access the first entry of each bucket (that is, usually the only one). It also has better locality of reference, particularly with linear probing. With small record sizes, these factors can yield better performance than chaining, particularly for lookups. Hash tables with open addressing are also easier to serialize, because they do not use pointers.

On the other hand, normal open addressing is a poor choice for large elements, because these elements fill entire CPU cache lines (negating the cache advantage), and a large amount of space is wasted on large empty table slots. If the open addressing table only stores references to elements (external storage), it uses space comparable to chaining even for large records but loses its speed advantage.



Hash collision resolved by open addressing with linear probing (interval=1). Note that "Ted Baker" has a unique hash, but nevertheless collided with "Sandra Dee", that had previously collided with "John Smith".



This graph compares the average number of CPU cache misses required to look up elements in large hash tables (far exceeding size of the cache) with chaining and linear probing. Linear probing performs better due to better locality of reference, though as the table gets full, its performance degrades drastically.

Generally speaking, open addressing is better used for hash tables with small records that can be stored within the table (internal storage) and fit in a cache line. They are particularly suitable for elements of one word or less. If the table is expected to have a high load factor, the records are large, or the data is variable-sized, chained hash tables often perform as well or better.

Coalesced hashing

A hybrid of chaining and open addressing, coalesced hashing links together chains of nodes within the table itself.^[19] Like open addressing, it achieves space usage and (somewhat diminished) cache advantages over chaining. Like chaining, it does not exhibit clustering effects; in fact, the table can be efficiently filled to a high density. Unlike chaining, it cannot have more elements than table slots.

Cuckoo hashing

Another alternative open-addressing solution is cuckoo hashing, which ensures constant lookup and deletion time in the worst case, and constant amortized time for insertions (with low probability that the worst-case will be encountered). It uses two or more hash functions, which means any key/value pair could be in two or more locations. For lookup, the first hash function is used; if the key/value is not found, then the second hash function is used, and so on. If a collision happens during insertion, then the key is re-hashed with the second hash function to map it to another bucket. If all hash functions are used and there is still a collision, then the key it collided with is removed to make space for the new key, and the old key is re-hashed with one of the other hash functions, which maps it to another bucket. If that location also results in a collision, then the process repeats until there is no collision or the process traverses all the buckets, at which point the table is resized. By combining multiple hash functions with multiple cells per bucket, very high space utilization can be achieved.

Hopscotch hashing

Another alternative open-addressing solution is hopscotch hashing,^[21] which combines the approaches of cuckoo hashing and linear probing, yet seems in general to avoid their limitations. In particular it works well even when the load factor grows beyond 0.9. The algorithm is well suited for implementing a resizable concurrent hash table.

The hopscotch hashing algorithm works by defining a neighborhood of buckets near the original hashed bucket, where a given entry is always found. Thus, search is limited to the number of entries in this neighborhood, which is logarithmic in the worst case, constant on average, and with proper alignment of the neighborhood typically requires one cache miss. When inserting an entry, one first attempts to add it to a bucket in the neighborhood. However, if all buckets in this neighborhood are occupied, the algorithm traverses buckets in sequence until an open slot (an unoccupied bucket) is found (as in linear probing). At that point, since the empty bucket is outside the neighborhood, items are repeatedly displaced in a sequence of hops. (This is similar to cuckoo hashing, but with the difference that in this case the empty slot is being moved into the neighborhood, instead of items being moved out with the hope of eventually finding an empty slot.) Each hop brings the open slot closer to the original neighborhood, without invalidating the neighborhood property of any of the buckets along the way. In the end, the open slot has been moved into the neighborhood, and the entry being inserted can be added to it.

Robin Hood hashing

A variation on double-hashing collision resolution is Robin Hood hashing.^{[22][23]} The idea is that a new key may displace a key already inserted, if its probe count is larger than that of the key at the current position. The net effect of this is that it reduces worst case search times in the table. This is similar to ordered hash tables^[24] except that the criterion for bumping a key does not depend on a direct relationship between the keys. Since both the worst case and the variation in the number of probes is reduced dramatically, an interesting variation is to probe the table starting at the expected successful probe value and then expand from that position in both directions.^[25] External Robin Hood hashing is an extension of this algorithm where the table is stored in an external file and each table position corresponds to a fixed-sized page or bucket with B records.^[26]

2-choice hashing

2-choice hashing employs two different hash functions, $h_1(x)$ and $h_2(x)$, for the hash table. Both hash functions are used to compute two table locations. When an object is inserted in the table, it is placed in the table location that contains fewer objects (with the default being the $h_1(x)$ table location if there is equality in bucket size). 2-choice hashing employs the principle of the power of two choices.^[27]

Dynamic resizing

When an insert is made such that the number of entries in a hash table exceeds the product of the load factor and the current capacity then the hash table will need to be *rehashed*.^[9] Rehashing includes increasing the size of the underlying data structure^[9] and mapping existing items to new bucket locations. In some implementations, if the initial capacity is greater than the maximum number of entries divided by the load factor, no rehash operations will ever occur.^[9]

To limit the proportion of memory wasted due to empty buckets, some implementations also shrink the size of the table—followed by a rehash—when items are deleted. From the point of space–time tradeoffs, this operation is similar to the deallocation in dynamic arrays.

Resizing by copying all entries

A common approach is to automatically trigger a complete resizing when the load factor exceeds some threshold r_{\max} . Then a new larger table is allocated, each entry is removed from the old table, and inserted into the new table. When all entries have been removed from the old table then the old table is returned to the free storage pool. Likewise, when the load factor falls below a second threshold r_{\min} , all entries are moved to a new smaller table.

For hash tables that shrink and grow frequently, the resizing downward can be skipped entirely. In this case, the table size is proportional to the maximum number of entries that ever were in the hash table at one time, rather than the current number. The disadvantage is that memory usage will be higher, and thus cache behavior may be worse. For best control, a "shrink-to-fit" operation can be provided that does this only on request.

If the table size increases or decreases by a fixed percentage at each expansion, the total cost of these resizings, amortized over all insert and delete operations, is still a constant, independent of the number of entries n and of the number m of operations performed.

For example, consider a table that was created with the minimum possible size and is doubled each time the load ratio exceeds some threshold. If m elements are inserted into that table, the total number of extra re-insertions that occur in all dynamic resizings of the table is at most $m - 1$. In other words, dynamic resizing roughly doubles the cost of each insert or delete operation.

Alternatives to all-at-once rehashing

Some hash table implementations, notably in real-time systems, cannot pay the price of enlarging the hash table all at once, because it may interrupt time-critical operations. If one cannot avoid dynamic resizing, a solution is to perform the resizing gradually.

Disk-based hash tables almost always use some alternative to all-at-once rehashing, since the cost of rebuilding the entire table on disk would be too high.

Incremental resizing

One alternative to enlarging the table all at once is to perform the rehashing gradually:

- During the resize, allocate the new hash table, but keep the old table unchanged.
- In each lookup or delete operation, check both tables.
- Perform insertion operations only in the new table.
- At each insertion also move r elements from the old table to the new table.
- When all elements are removed from the old table, deallocate it.

To ensure that the old table is completely copied over before the new table itself needs to be enlarged, it is necessary to increase the size of the table by a factor of at least $(r + 1)/r$ during resizing.

Monotonic keys

If it is known that keys will be stored in monotonically increasing (or decreasing) order, then a variation of consistent hashing can be achieved.

Given some initial key k_1 , a subsequent key k_i partitions the key domain $[k_1, \infty)$ into the set $\{[k_1, k_i), [k_i, \infty)\}$. In general, repeating this process gives a finer partition $\{[k_1, k_{i_0}), [k_{i_0}, k_{i_1}), \dots, [k_{i_{n-1}}, k_{i_n}), [k_{i_n}, \infty)\}$ for some sequence of monotonically increasing keys $(k_{i_0}, \dots, k_{i_n})$, where n is the number of refinements. The same process applies, *mutatis mutandis*, to monotonically decreasing keys. By assigning to each subinterval of this partition a different hash function or hash table (or both), and by refining the partition whenever the hash table is resized, this approach guarantees that any key's hash, once issued, will never change, even when the hash table is grown.

Since it is common to grow the overall number of entries by doubling, there will only be $O(\log(N))$ subintervals to check, and binary search time for the redirection will be $O(\log(\log(N)))$.

Linear hashing

Linear hashing^[28] is a hash table algorithm that permits incremental hash table expansion. It is implemented using a single hash table, but with two possible lookup functions.

Hashing for distributed hash tables

Another way to decrease the cost of table resizing is to choose a hash function in such a way that the hashes of most values do not change when the table is resized. Such hash functions are prevalent in disk-based and distributed hash tables, where rehashing is prohibitively costly. The problem of designing a hash such that

most values do not change when the table is resized is known as the distributed hash table problem. The four most popular approaches are rendezvous hashing, consistent hashing, the content addressable network algorithm, and Kademlia distance.

Performance

Speed analysis

In the simplest model, the hash function is completely unspecified and the table does not resize. With an ideal hash function, a table of size k with open addressing has no collisions and holds up to k elements with a single comparison for successful lookup, while a table of size k with chaining and n keys has the minimum $\max(0, n - k)$ collisions and $\Theta(1 + \frac{n}{k})$ comparisons for lookup. With the worst possible hash function, every insertion causes a collision, and hash tables degenerate to linear search, with $\Theta(n)$ amortized comparisons per insertion and up to n comparisons for a successful lookup.

Adding rehashing to this model is straightforward. As in a dynamic array, geometric resizing by a factor of b implies that only $\frac{n}{b^i}$ keys are inserted i or more times, so that the total number of insertions is bounded above by $\frac{bn}{b-1}$, which is $\Theta(n)$. By using rehashing to maintain $n < k$, tables using both chaining and open addressing can have unlimited elements and perform successful lookup in a single comparison for the best choice of hash function.

In more realistic models, the hash function is a random variable over a probability distribution of hash functions, and performance is computed on average over the choice of hash function. When this distribution is uniform, the assumption is called "simple uniform hashing" and it can be shown that hashing with chaining requires $\Theta(1 + \frac{n}{k})$ comparisons on average for an unsuccessful lookup, and hashing with open addressing requires $\Theta\left(\frac{1}{1 - n/k}\right)$.^[29] Both these bounds are constant, if we maintain ' $\frac{n}{k} < c$ ' using table resizing, where c is a fixed constant less than 1.

Two factors affect significantly the latency of operations on a hash table:^[30]

- Cache missing. With the increasing of load factor, the search and insertion performance of hash tables can be degraded a lot due to the rise of average cache missing.
- Cost of resizing. Resizing becomes an extreme time-consuming task when hash tables grow massive.

In latency-sensitive programs, the time consumption of operations on both the average and the worst cases are required to be small, stable, and even predictable. The K hash table ^[31] is designed for a general scenario of low-latency applications, aiming to achieve cost-stable operations on a growing huge-sized table.

Memory utilization

Sometimes the memory requirement for a table needs to be minimized. One way to reduce memory usage in chaining methods is to eliminate some of the chaining pointers or to replace them with some form of abbreviated pointers.

Another technique was introduced by Donald Knuth and is called *quotienting*. For this discussion assume that the key, or a reversibly-hashed version of that key, is an integer m from $\{0, 1, 2, \dots, M-1\}$ and the number of buckets is N . m is divided by N to produce a quotient q and a remainder r . The remainder r is used to select the bucket; in the bucket only the quotient q need be stored. This saves $\log_2(N)$ bits per element, which can be very significant in some applications.

Quotienting works readily with chaining hash tables, or with simple cuckoo hash tables. To apply the technique with ordinary open-addressing hash tables, John G. Cleary introduced a method^[32] where two bits (a *virgin* bit and a *change* bit) are included in each bucket to allow the original bucket index (r) to be reconstructed.

In the scheme just described, $\log_2(M/N) + 2$ bits are used to store each key. It is interesting to note that the theoretical minimum storage would be $\log_2(M/N) + 1.4427$ bits where $1.4427 = \log_2(e)$.

Features

Advantages

- The main advantage of hash tables over other table data structures is speed. This advantage is more apparent when the number of entries is large. Hash tables are particularly efficient when the maximum number of entries can be predicted in advance, so that the bucket array can be allocated once with the optimum size and never resized.
- If the set of key-value pairs is fixed and known ahead of time (so insertions and deletions are not allowed), one may reduce the average lookup cost by a careful choice of the hash function, bucket table size, and internal data structures. In particular, one may be able to devise a hash function that is collision-free, or even perfect. In this case the keys need not be stored in the table.

Drawbacks

- Although operations on a hash table take constant time on average, the cost of a good hash function can be significantly higher than the inner loop of the lookup algorithm for a sequential list or search tree. Thus hash tables are not effective when the number of entries is very small. (However, in some cases the high cost of computing the hash function can be mitigated by saving the hash value together with the key.)
- For certain string processing applications, such as spell-checking, hash tables may be less efficient than tries, finite automata, or Judy arrays. Also, if there are not too many possible keys to store—that is, if each key can be represented by a small enough number of bits—then, instead of a hash table, one may use the key directly as the index into an array of values. Note that there are no collisions in this case.
- The entries stored in a hash table can be enumerated efficiently (at constant cost per entry), but only in some pseudo-random order. Therefore, there is no efficient way to locate an entry whose key is *nearest* to a given key. Listing all n entries in some specific order generally requires a separate sorting step, whose cost is proportional to $\log(n)$ per entry. In comparison, ordered search trees have lookup and insertion cost proportional to $\log(n)$, but allow finding the nearest key at about the same cost, and *ordered* enumeration of all entries at constant cost per entry. However, a LinkingHashMap can be made to create a hash table with a non-random sequence.^[33]
- If the keys are not stored (because the hash function is collision-free), there may be no easy way to enumerate the keys that are present in the table at any given moment.

- Although the *average* cost per operation is constant and fairly small, the cost of a single operation may be quite high. In particular, if the hash table uses dynamic resizing, an insertion or deletion operation may occasionally take time proportional to the number of entries. This may be a serious drawback in real-time or interactive applications.
- Hash tables in general exhibit poor locality of reference—that is, the data to be accessed is distributed seemingly at random in memory. Because hash tables cause access patterns that jump around, this can trigger microprocessor cache misses that cause long delays. Compact data structures such as arrays searched with linear search may be faster, if the table is relatively small and keys are compact. The optimal performance point varies from system to system.
- Hash tables become quite inefficient when there are many collisions. While extremely uneven hash distributions are extremely unlikely to arise by chance, a malicious adversary with knowledge of the hash function may be able to supply information to a hash that creates worst-case behavior by causing excessive collisions, resulting in very poor performance, e.g., a denial of service attack.^{[34][35][36]} In critical applications, a data structure with better worst-case guarantees can be used; however, universal hashing—a randomized algorithm that prevents the attacker from predicting which inputs cause worst-case behavior—may be preferable.^[37] The hash function used by the hash table in the Linux routing table cache was changed with Linux version 2.4.2 as a countermeasure against such attacks.^[38]

Uses

Associative arrays

Hash tables are commonly used to implement many types of in-memory tables. They are used to implement associative arrays (arrays whose indices are arbitrary strings or other complicated objects), especially in interpreted programming languages like Ruby, Python, and PHP.

When storing a new item into a multimap and a hash collision occurs, the multimap unconditionally stores both items.

When storing a new item into a typical associative array and a hash collision occurs, but the actual keys themselves are different, the associative array likewise stores both items. However, if the key of the new item exactly matches the key of an old item, the associative array typically erases the old item and overwrites it with the new item, so every item in the table has a unique key.

Database indexing

Hash tables may also be used as disk-based data structures and database indices (such as in dbm) although B-trees are more popular in these applications. In multi-node database systems, hash tables are commonly used to distribute rows amongst nodes, reducing network traffic for hash joins.

Caches

Hash tables can be used to implement caches, auxiliary data tables that are used to speed up the access to data that is primarily stored in slower media. In this application, hash collisions can be handled by discarding one of the two colliding entries—usually erasing the old item that is currently stored in the table and overwriting it with the new item, so every item in the table has a unique hash value.

Sets

Besides recovering the entry that has a given key, many hash table implementations can also tell whether such an entry exists or not.

Those structures can therefore be used to implement a set data structure,^[39] which merely records whether a given key belongs to a specified set of keys. In this case, the structure can be simplified by eliminating all parts that have to do with the entry values. Hashing can be used to implement both static and dynamic sets.

Object representation

Several dynamic languages, such as Perl, Python, JavaScript, Lua, and Ruby, use hash tables to implement objects. In this representation, the keys are the names of the members and methods of the object, and the values are pointers to the corresponding member or method.

Unique data representation

Hash tables can be used by some programs to avoid creating multiple character strings with the same contents. For that purpose, all strings in use by the program are stored in a single *string pool* implemented as a hash table, which is checked whenever a new string has to be created. This technique was introduced in Lisp interpreters under the name hash consing, and can be used with many other kinds of data (expression trees in a symbolic algebra system, records in a database, files in a file system, binary decision diagrams, etc.).

Transposition table

A transposition table is a complex Hash Table which stores information about each section that has been searched.^[40]

Implementations

In programming languages

Many programming languages provide hash table functionality, either as built-in associative arrays or as standard library modules. In C++11, for example, the unordered_map class provides hash tables for keys and values of arbitrary type.

The Java programming language (including the variant which is used on Android) includes the HashSet, HashMap, LinkedHashSet, and LinkedHashMap generic collections.^[41]

In PHP 5 and 7, the Zend 2 engine and the Zend 3 engine (respectively) use one of the hash functions from Daniel J. Bernstein to generate the hash values used in managing the mappings of data pointers stored in a hash table. In the PHP source code, it is labelled as DJBX33A (Daniel J. Bernstein, Times 33 with Addition).

Python's built-in hash table implementation, in the form of the `dict` type, as well as Perl's hash type (%) are used internally to implement namespaces and therefore need to pay more attention to security, i.e., collision attacks. Python sets also use hashes internally, for fast lookup (though they store only keys, not values).^[42] CPython 3.6+ uses an insertion-ordered variant of the hash table, implemented by splitting out the value storage into an array and having the vanilla hash table only store a set of indices.^[43]

In the .NET Framework, support for hash tables is provided via the non-generic `Hashtable` and generic `Dictionary` classes, which store key-value pairs, and the generic `HashSet` class, which stores only values.

In Ruby the hash table uses the open addressing model from Ruby 2.4 onwards.^{[44][45]}

In Rust's standard library, the generic `HashMap` and `HashSet` structs use linear probing with Robin Hood bucket stealing.

ANSI Smalltalk defines the classes `Set` / `IdentitySet` and `Dictionary` / `IdentityDictionary`. All Smalltalk implementations provide additional (not yet standardized) versions of `WeakSet`, `WeakKeyDictionary` and `WeakValueDictionary`.

Tcl array variables are hash tables, and Tcl dictionaries are immutable values based on hashes. The functionality is also available as C library functions `Tcl_InitHashTable` et al. (<http://www.tcl.tk/man/tcl8.6/TclLib/Hash.htm>) (for generic hash tables) and `Tcl_NewDictObj` et al. (<http://www.tcl.tk/man/tcl8.6/TclLib/DictObj.htm>) (for dictionary values). The performance has been independently benchmarked as extremely competitive.^[46]

In Wolfram language supports hash tables since version 10. They are implemented under the name `Association`.

Common Lisp provides the `hash-table` class for efficient mappings. In spite of its naming, the language standard does not mandate the actual adherence to any hashing technique for implementations.^[47]

History

The idea of hashing arose independently in different places. In January 1953, Hans Peter Luhn wrote an internal IBM memorandum that used hashing with chaining.^[48] Gene Amdahl, Elaine M. McGraw, Nathaniel Rochester, and Arthur Samuel implemented a program using hashing at about the same time. Open addressing with linear probing (relatively prime stepping) is credited to Amdahl, but Ershov (in Russia) had the same idea.^[48]

See also

- Rabin–Karp string search algorithm
- Stable hashing
- Consistent hashing
- Extendible hashing
- Lazy deletion
- Pearson hashing
- PhotoDNA
- Search data structure
- Concurrent hash table
- Record (computer science)

Related data structures

There are several data structures that use hash functions but cannot be considered special cases of hash tables:

- Bloom filter, memory efficient data-structure designed for constant-time approximate lookups; uses hash function(s) and can be seen as an approximate hash table.
- Distributed hash table (DHT), a resilient dynamic table spread over several nodes of a network.
- Hash array mapped trie, a trie structure, similar to the array mapped trie, but where each key is hashed first.

References

1. Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2009). *Introduction to Algorithms* (3rd ed.). Massachusetts Institute of Technology. pp. 253–280. ISBN 978-0-262-03384-8.
2. Charles E. Leiserson, *Amortized Algorithms, Table Doubling, Potential Method* (http://videolectures.net/mit6046jf05_leiserson_lec13/) Archived (https://web.archive.org/web/20090807022046/http://videolectures.net/mit6046jf05_leiserson_lec13/) August 7, 2009, at the Wayback Machine Lecture 13, course MIT 6.046J/18.410J Introduction to Algorithms—Fall 2005
3. Knuth, Donald (1998). *The Art of Computer Programming. 3: Sorting and Searching* (2nd ed.). Addison-Wesley. pp. 513–558. ISBN 978-0-201-89685-5.
4. Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2001). "Chapter 11: Hash Tables". *Introduction to Algorithms* (2nd ed.). MIT Press and McGraw-Hill. pp. 221 (https://archive.org/details/introductiontoal00corm_691/page/n243)–252. ISBN 978-0-262-53196-2.
5. "JDK HashMap Hashcode implementation" (<http://hg.openjdk.java.net/jdk8/jdk8/jdk/file/687fd7c7986d/src/share/classes/java/util/HashMap.java#l322>). Archived (<https://web.archive.org/web/20170521033827/http://hg.openjdk.java.net/jdk8/jdk8/jdk/file/687fd7c7986d/src/share/classes/java/util/HashMap.java#l322>) from the original on May 21, 2017.
6. Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". *Philosophical Magazine*. Series 5. **50** (302). pp. 157–175. doi:10.1080/14786440009463897 (<https://doi.org/10.1080%2F14786440009463897>).
7. Plackett, Robin (1983). "Karl Pearson and the Chi-Squared Test". *International Statistical Review (International Statistical Institute (ISI))*. **51** (1). pp. 59–72. doi:10.2307/1402731 (<https://doi.org/10.2307%2F1402731>). JSTOR 1402731 (<https://www.jstor.org/stable/1402731>).
8. Wang, Thomas (March 1997). "Prime Double Hash Table" (<https://web.archive.org/web/19990903133921/http://www.concentric.net/~Ttwang/tech/primehash.htm>). Archived from the original (<https://www.concentric.net/~Ttwang/tech/primehash.htm>) on September 3, 1999. Retrieved May 10, 2015.
9. Javadoc for HashMap in Java 10 <https://docs.oracle.com/javase/10/docs/api/java/util/HashMap.html>
10. "CS Hash Table" (https://everythingcomputerscience.com/discrete_mathematics/Data_Structures/Hash_Table.html). *everythingcomputerscience.com*.
11. Jaco Geldenhuys; Antti Valmari. "Nearly Memory-optimal Data Structure" (<https://dl.acm.org/doi/10.5555/1767111.1767120>). *ACM Digital Library*.
12. Probst, Mark (April 30, 2010). "Linear vs Binary Search" (<https://schani.wordpress.com/2010/04/30/linear-vs-binary-search/>). Archived (<https://web.archive.org/web/20161120154213/https://schani.wordpress.com/2010/04/30/linear-vs-binary-search/>) from the original on November 20, 2016. Retrieved November 20, 2016.
13. "How does a HashMap work in JAVA" (http://coding-geek.com/how-does-a-hashmap-work-in-java/#JAVA_8_improvements). *coding-geek.com*. Archived (https://web.archive.org/web/20161119095443/http://coding-geek.com/how-does-a-hashmap-work-in-java/#JAVA_8_improvements) from the original on November 19, 2016.

14. Askitis, Nikolas; Zobel, Justin (October 2005). *Cache-conscious Collision Resolution in String Hash Tables. Proceedings of the 12th International Conference, String Processing and Information Retrieval (SPIRE 2005)*. 3772/2005. pp. 91–102. doi:10.1007/11575832_11 (http://doi.org/10.1007%2F11575832_11). ISBN 978-3-540-29740-6.
15. Askitis, Nikolas; Sinha, Ranjan (2010). "Engineering scalable, cache and space efficient tries for strings". *The VLDB Journal*. **17** (5): 633–660. doi:10.1007/s00778-010-0183-9 (<https://doi.org/10.1007%2Fs00778-010-0183-9>). ISSN 1066-8888 (<https://www.worldcat.org/issn/1066-8888>). S2CID 432572 (<https://api.semanticscholar.org/CorpusID:432572>).
16. Askitis, Nikolas (2009). *Fast and Compact Hash Tables for Integer Keys* (<https://web.archive.org/web/20110216180225/http://crpit.com/confpapers/CRPITV91Askitis.pdf>) (PDF). *Proceedings of the 32nd Australasian Computer Science Conference (ACSC 2009)*. **91**. pp. 113–122. ISBN 978-1-920682-72-9. Archived from the original (<http://crpit.com/confpapers/CRPITV91Askitis.pdf>) (PDF) on February 16, 2011. Retrieved June 13, 2010.
17. Erik Demaine, Jeff Lind. 6.897: Advanced Data Structures. MIT Computer Science and Artificial Intelligence Laboratory. Spring 2003. "Archived copy" (http://courses.csail.mit.edu/6.897/spring03/scribe_notes/L2/lecture2.pdf) (PDF). Archived (https://web.archive.org/web/20100615203901/http://courses.csail.mit.edu/6.897/spring03/scribe_notes/L2/lecture2.pdf) (PDF) from the original on June 15, 2010. Retrieved June 30, 2008.
18. Willard, Dan E. (2000). "Examining computational geometry, van Emde Boas trees, and hashing from the perspective of the fusion tree". *SIAM Journal on Computing*. **29** (3): 1030–1049. doi:10.1137/S0097539797322425 (<https://doi.org/10.1137%2FS0097539797322425>). MR 1740562 (<https://www.ams.org/mathscinet-getitem?mr=1740562>).
19. Tenenbaum, Aaron M.; Langsam, Yedidyah; Augenstein, Moshe J. (1990). *Data Structures Using C*. Prentice Hall. pp. 456–461, p. 472. ISBN 978-0-13-199746-2.
20. Pagh, Rasmus; Rodler, Flemming Friche (2001). "Cuckoo Hashing". *Algorithms — ESA 2001. Lecture Notes in Computer Science*. **2161**. pp. 121–133. CiteSeerX 10.1.1.25.4189 (<https://cite-seerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.4189>). doi:10.1007/3-540-44676-1_10 (http://doi.org/10.1007%2F3-540-44676-1_10). ISBN 978-3-540-42493-2.
21. Herlihy, Maurice; Shavit, Nir; Tzafrir, Moran (2008). "Hopscotch Hashing". *DISC '08: Proceedings of the 22nd international symposium on Distributed Computing*. Berlin, Heidelberg: Springer-Verlag. pp. 350–364. CiteSeerX 10.1.1.296.8742 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.296.8742>).
22. Celis, Pedro (1986). *Robin Hood hashing* (<https://cs.uwaterloo.ca/research/tr/1986/CS-86-14.pdf>) (PDF) (Technical report). Computer Science Department, University of Waterloo. CS-86-14. Archived (<https://web.archive.org/web/20140717220245/https://cs.uwaterloo.ca/research/tr/1986/CS-86-14.pdf>) (PDF) from the original on July 17, 2014.
23. Goossaert, Emmanuel (2013). "Robin Hood hashing" (<http://codecapsule.com/2013/11/11/robin-hood-hashing/>). Archived (<https://web.archive.org/web/20140321175355/http://codecapsule.com/2013/11/11/robin-hood-hashing/>) from the original on March 21, 2014.
24. Amble, Ole; Knuth, Don (1974). "Ordered hash tables" (<https://doi.org/10.1093/comjnl/17.2.135>). *Computer Journal*. **17** (2): 135. doi:10.1093/comjnl/17.2.135 (<https://doi.org/10.1093%2Fcomjnl%2F17.2.135>).
25. Viola, Alfredo (October 2005). "Exact distribution of individual displacements in linear probing hashing". *Transactions on Algorithms (TALG)*. **1** (2): 214–242. doi:10.1145/1103963.1103965 (<https://doi.org/10.1145%2F1103963.1103965>). S2CID 11183559 (<https://api.semanticscholar.org/CorpusID:11183559>).
26. Celis, Pedro (March 1988). *External Robin Hood Hashing* (Technical report). Computer Science Department, Indiana University. TR246.

27. Mitzenmacher, Michael; Richa, Andréa W.; Sitaraman, Ramesh (2001). "The Power of Two Random Choices: A Survey of Techniques and Results" (<http://www.eecs.harvard.edu/~michaelm/postscripts/handbook2001.pdf>) (PDF). Harvard University. Archived (<https://web.archive.org/web/20150325175258/http://www.eecs.harvard.edu/%7Emichaelm/postscripts/handbook2001.pdf>) (PDF) from the original on March 25, 2015. Retrieved April 10, 2015.
28. Litwin, Witold (1980). "Linear hashing: A new tool for file and table addressing". *Proc. 6th Conference on Very Large Databases*. pp. 212–223.
29. Doug Dunham. CS 4521 Lecture Notes (<http://www.duluth.umn.edu/~ddunham/cs4521s09/notes/ch11.txt>) Archived (<https://web.archive.org/web/20090722072750/http://www.duluth.umn.edu/~ddunham/cs4521s09/notes/ch11.txt>) July 22, 2009, at the *Wayback Machine*. University of Minnesota Duluth. Theorems 11.2, 11.6. Last modified April 21, 2009.
30. Andy Ke. Inside the latency of hash table operations (<https://www.linkedin.com/pulse/inside-latency-hash-table-operations-li-andy-ke/>) Last modified December 30, 2019.
31. Andy Ke. The K hash table, a design for low-latency applications (<https://www.linkedin.com/pulse/k-hash-table-design-low-latency-applications-li-andy-ke/>) Last modified December 20, 2019.
32. Clerry (1984). "Compact Hash Tables Using Bidirectional Linear Probing" (<https://www.semanticscholar.org/paper/Compact-Hash-Tables-Using-Bidirectional-Linear-Clerry/13781d147663c2bf60072a86d3e4a2adc4bb244b>). *IEEE Transactions on Computers* (9): 828–834. doi:10.1109/TC.1984.1676499 (<https://doi.org/10.1109/2FTC.1984.1676499>). S2CID 195908955 (<https://api.semanticscholar.org/CorpusID:195908955>).
33. "LinkedHashMap (Java Platform SE 7)" (<https://docs.oracle.com/javase/7/docs/api/java/util/LinkedHashMap.html>). *docs.oracle.com*. Retrieved May 1, 2020.
34. Alexander Klink and Julian Wälde's *Efficient Denial of Service Attacks on Web Application Platforms* (https://events.ccc.de/congress/2011/Fahrplan/attachments/2007_28C3_Effective_DoS_on_web_application_platforms.pdf) Archived (https://web.archive.org/web/20160916135747/https://events.ccc.de/congress/2011/Fahrplan/attachments/2007_28C3_Effective_DoS_on_web_application_platforms.pdf) September 16, 2016, at the *Wayback Machine*, December 28, 2011, 28th Chaos Communication Congress. Berlin, Germany.
35. Mike Lennon "Hash Table Vulnerability Enables Wide-Scale DDoS Attacks" (<http://www.securityweek.com/hash-table-collision-attacks-could-trigger-ddos-massive-scale>) Archived (<https://web.archive.org/web/20160919171346/http://www.securityweek.com/hash-table-collision-attacks-could-trigger-ddos-massive-scale>) September 19, 2016, at the *Wayback Machine*. 2011.
36. "Hardening Perl's Hash Function" (<http://blog.booking.com/hardening-perls-hash-function.html>). November 6, 2013. Archived (<https://web.archive.org/web/20160916233851/http://blog.booking.com/hardening-perls-hash-function.html>) from the original on September 16, 2016.
37. Crosby and Wallach. *Denial of Service via Algorithmic Complexity Attacks* (<https://www.usenix.org/conference/12th-usenix-security-symposium/denial-service-algorithmic-complexity-attacks>) Archived (<https://web.archive.org/web/20160304121231/https://www.usenix.org/conference/12th-usenix-security-symposium/denial-service-algorithmic-complexity-attacks>) March 4, 2016, at the *Wayback Machine*. quote: "modern universal hashing techniques can yield performance comparable to commonplace hash functions while being provably secure against these attacks." "Universal hash functions ... are ... a solution suitable for adversarial environments. ... in production systems."
38. Bar-Yosef, Noa; Wool, Avishai (2007). *Remote algorithmic complexity attacks against randomized hash tables* *Proc. International Conference on Security and Cryptography (SECRYPT)* (https://www.eng.tau.ac.il/~yash/C2_039_Wool.pdf) (PDF). p. 124. Archived (https://web.archive.org/web/20140916120456/http://www.eng.tau.ac.il/%7Eyash/C2_039_Wool.pdf) (PDF) from the original on September 16, 2014.
39. "Set (Java Platform SE 7)" (<https://docs.oracle.com/javase/7/docs/api/java/util/Set.html>). *docs.oracle.com*. Retrieved May 1, 2020.

40. "Transposition Table - Chessprogramming wiki" (https://www.chessprogramming.org/Transposition_Table). *chessprogramming.org*. Retrieved May 1, 2020.
41. "Lesson: Implementations (The Java™ Tutorials > Collections)" (<https://docs.oracle.com/javase/tutorial/collections/implementations/index.html>). *docs.oracle.com*. Archived (<https://web.archive.org/web/20170118041252/https://docs.oracle.com/javase/tutorial/collections/implementations/index.html>) from the original on January 18, 2017. Retrieved April 27, 2018.
42. "Python: List vs Dict for look up table" (<https://stackoverflow.com/questions/513882/python-list-vs-dict-for-look-up-table>). *stackoverflow.com*. Archived (<https://web.archive.org/web/20171202040407/https://stackoverflow.com/questions/513882/python-list-vs-dict-for-look-up-table>) from the original on December 2, 2017. Retrieved April 27, 2018.
43. Dimitris Fasarakis Hilliard. "Are dictionaries ordered in Python 3.6+?" (<https://stackoverflow.com/a/39980744/>). *Stack Overflow*.
44. Dmitriy Vasin (June 19, 2018). "Do You Know How Hash Table Works? (Ruby Examples)" (<http://anadea.info/blog/how-hash-table-works-ruby-examples>). *anadea.info*. Retrieved July 3, 2019.
45. Jonan Scheffler (December 25, 2016). "Ruby 2.4 Released: Faster Hashes, Unified Integers and Better Rounding" (<https://blog.heroku.com/ruby-2-4-features-hashes-integers-rounding#hash-changes>). *heroku.com*. Retrieved July 3, 2019.
46. Wing, Eric. "Hash Table Shootout 2: Rise of the Interpreter Machines" (<https://web.archive.org/web/20131014184050/http://playcontrol.net/opensource/LuaHashMap/benchmarks.html>). *LuaHashMap: An easy to use hash table library for C*. PlayControl Software. Archived from the original (<https://playcontrol.net/opensource/LuaHashMap/benchmarks.html>) on October 14, 2013. Retrieved October 24, 2019. "Did Tcl win? In any case, these benchmarks showed that these interpreter implementations have very good hash implementations and are competitive with our reference benchmark of the STL unordered_map. Particularly in the case of Tcl and Lua, they were extremely competitive and often were within 5%-10% of unordered_map when they weren't beating it." (On 2019-10-24, the original site still has the text, but the figures appear to be broken, whereas they are intact in the archive.)
47. "CLHS: System Class HASH-TABLE" (http://www.lispworks.com/documentation/HyperSpec/Body/t_hash_t.htm#hash-table). *lispworks.com/documentation/HyperSpec/Front/index.htm*. Archived (https://web.archive.org/web/20191022044020/http://www.lispworks.com/documentation/HyperSpec/Body/t_hash_t.htm) from the original on October 22, 2019. Retrieved May 18, 2020.
48. Mehta, Dinesh P.; Sahni, Sartaj (October 28, 2004). *Handbook of Datastructures and Applications*. p. 9-15. ISBN 978-1-58488-435-4.

Further reading

- Tamassia, Roberto; Goodrich, Michael T. (2006). "Chapter Nine: Maps and Dictionaries". *Data structures and algorithms in Java : [updated for Java 5.0]* (https://archive.org/details/datastructur-esal00good_183) (4th ed.). Hoboken, NJ: Wiley. pp. 369 (https://archive.org/details/datastructur-esal00good_183/page/n368)–418. ISBN 978-0-471-73884-8.
- McKenzie, B. J.; Harries, R.; Bell, T. (February 1990). "Selecting a hashing algorithm". *Software Practice & Experience*. **20** (2): 209–224. doi:10.1002/spe.4380200207 (<https://doi.org/10.1002%2Fspe.4380200207>). hdl:10092/9691 (<https://hdl.handle.net/10092%2F9691>).

External links

- A Hash Function for Hash Table Lookup (<http://www.burtleburtle.net/bob/hash/doobs.html>) by Bob Jenkins.
- Hash functions (<http://www.azillionmonkeys.com/qed/hash.html>) by Paul Hsieh

- [Design of Compact and Efficient Hash Tables for Java \(https://web.archive.org/web/20110505033634/http://blog.griddynamics.com/2011/03/ultimate-sets-and-maps-for-java-part-i.html\)](https://web.archive.org/web/20110505033634/http://blog.griddynamics.com/2011/03/ultimate-sets-and-maps-for-java-part-i.html)
 - [NIST entry on hash tables \(https://xlinux.nist.gov/dads/HTML/hashtab.html\)](https://xlinux.nist.gov/dads/HTML/hashtab.html)
 - [Lecture on Hash Tables from Stanford's CS106A \(https://web.stanford.edu/class/archive/cs/cs106a/cs106a.1178/lectures/Lecture20/Lecture20.pdf\)](https://web.stanford.edu/class/archive/cs/cs106a/cs106a.1178/lectures/Lecture20/Lecture20.pdf)
 - [Open Data Structures – Chapter 5 – Hash Tables \(http://opendatastructures.org/versions/edition-0.1e/ods-java/5_Hash_Tables.html\)](http://opendatastructures.org/versions/edition-0.1e/ods-java/5_Hash_Tables.html), Pat Morin
 - [MIT's Introduction to Algorithms: Hashing 1 \(http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-7-hashing-hash-functions/\)](http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-7-hashing-hash-functions/) MIT OCW lecture Video
 - [MIT's Introduction to Algorithms: Hashing 2 \(http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-8-universal-hashing-perfect-hashing/\)](http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-8-universal-hashing-perfect-hashing/) MIT OCW lecture Video
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Hash_table&oldid=992474320"

This page was last edited on 5 December 2020, at 13:16 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.