# CSC3170 Tutorial 5

School of Data Science

The Chinese University of Hong Kong, Shenzhen

# Outline
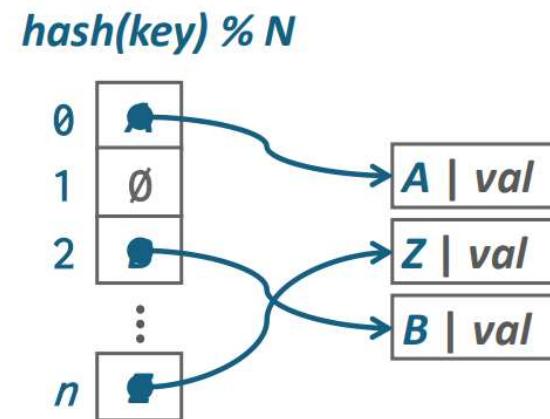
- Hash Tables
  - Concepts
  - Exercise

# Hash Tables

- A hash table implements an unordered associative array that maps keys to values.

- It uses a hash function to compute an offset into this array for a given key, from which the desired value can be found.
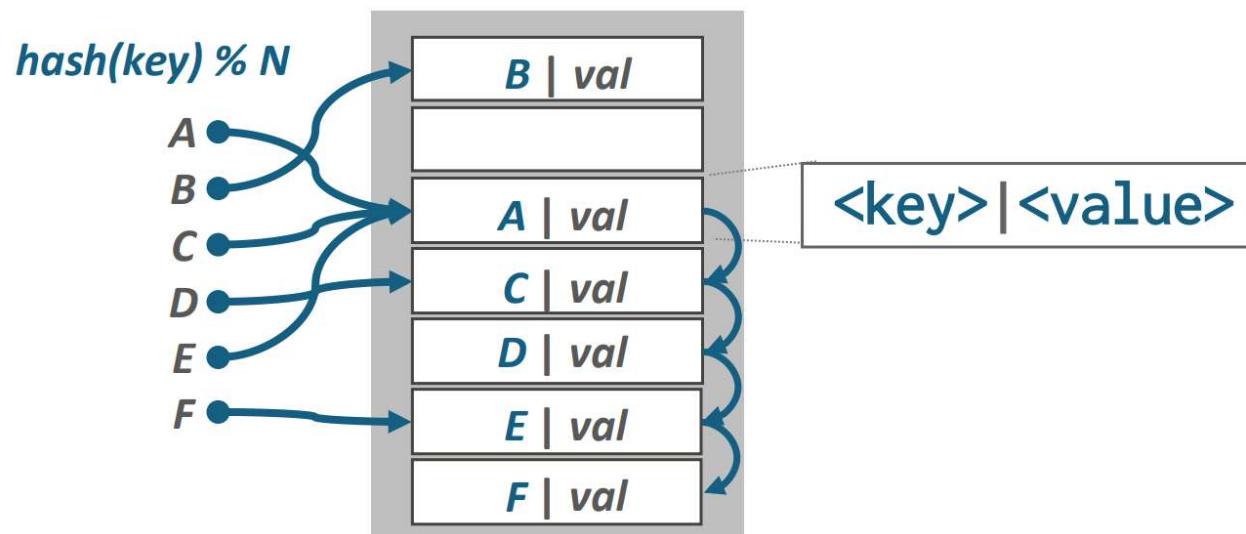
# Static Hash Table

- Allocate a giant array that has one slot for every element you need to store.

- Hash function: key % N

- Unrealistic assumptions

hash(key) % N

| | |
|---|---|
| 0 | 🔔 |
| 1 | Ø |
| 2 | 💬 |
| ⋮ | |
| n | ✊ |

A | val

Z | val

B | val
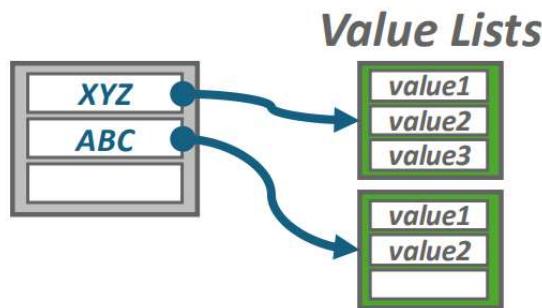
# Linear Probe Hashing

- Similar to the static hash table.
- Resolve collisions by linearly searching for the next free slot in the table.

# Non-Unique keys

- Separated Linked List
    - Store values in separate storage area of each key.
    - For instance, we have value lists for tuples with key = "XYZ" and



**Value Lists**

| XYZ | → | value1 |
| ABC | → | value2 |
|     |   | value3 |

| value1 |
| value2 |
|        |

- Redundant keys
    - store duplicate key entries together in the hash table.

# Cuckoo Hashing

- Use multiple hash functions to find multiple locations in the hash table to insert records.
  - On insert, check multiple locations and pick the one that is empty.
  - If no location is available, evict the element from one of them and then rehash it find a new location.
- Look-ups and deletions are always $O(1)$ because only one location per hash table is checked.

# Exercise

Consider the following cuckoo hashing schema:

1. Both tables have a size of 4.

2. The hashing function of the first table returns the fourth and third least significant bits:
   $h_1(x) = $ `(x >> 2) & 0b11`.

3. The hashing function of the second table returns the least significant two bits:
   $h_2(x) = $ `x & 0b11`.

4. When inserting, try table 1 first.

5. When replacement is necessary, first select an element in the second table.

6. The original entries in the table are shown in the figure below.

# Exercise

- Which elements is first inserted, 11 or 9?

- Let's now do some insertion and deletion:
  - insert 16
  - delete 11
  - insert 17
  - insert 10
  - insert 33
  - delete 16
- After those operations, what is the smallest key that potentially cause an infinite loop?
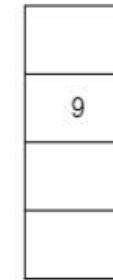


Figure 1: Initial contents of the hash tables.

# Q&A