Tutorial 7: Disk Storage, Basic File Structures, and Hashing

CS3402 Database Systems

Question 1

- If the database consists of 512 records and the blocking factor is 8. On average, how many blocks needed to be searched to if the records are in unordered format and ordered format?

Question 1 (Answer)

- 512/8 = 64 blocks
- Sequential search = 64/2 = 32
- Binary search = $\log_2 64 = \log_2 2^6 = 6$

Question 2

- Under which case the performance of hashing is better: (1) the values of hash key are uniformly distributed; and (2) the values of the hash key are in normal distribution.

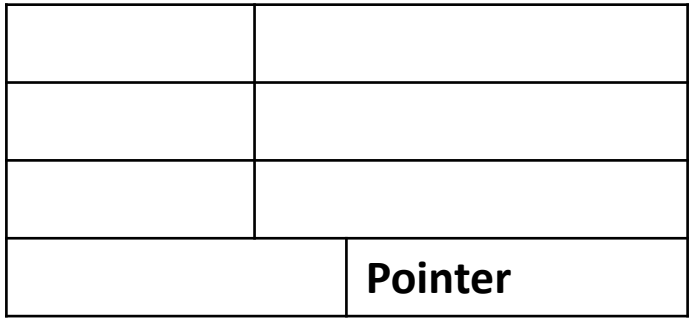
Question 2 (Answer)

- (1) is better. For (2), the overflow problem is more serious since many records are grouped into some buckets.
- References:
  - Uniform distribution https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)
  - Normal distribution https://en.wikipedia.org/wiki/Normal_distribution

Question 3 (1/2)

- Suppose that we use hashing to organize a *PRODUCT* file containing records with the following product# values: 2369, 3760, 4692, 4871, 5659, 1821, 1074, and 7115.
- (a) Let the hash function be ***h(product#) = product# mod 5***, show the *static hash* structure for this file. Assuming that each bucket can hold at most three records as shown below, and records in each bucket is unordered.

Bucket i



Question 3 (2/2)

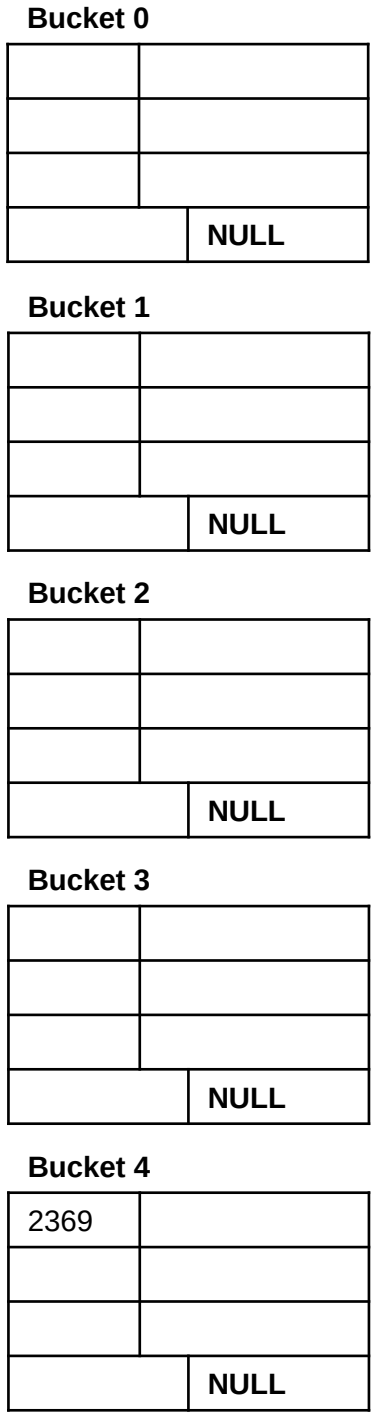
- (b) Some new records are inserted into the file with product# values: 1620, 2428, 3945, 4759, 6975, 4981, and 9206. Show the updated hash structure for this file when using *chaining* for *collision resolution*. That is, if collision occurs, new records are inserted in overflow buckets and pointers are set from the original buckets to the overflow buckets. Assuming that each overflow bucket can hold at most three records, as shown below.

Overflow Bucket

		Pointer
		Pointer
		Pointer

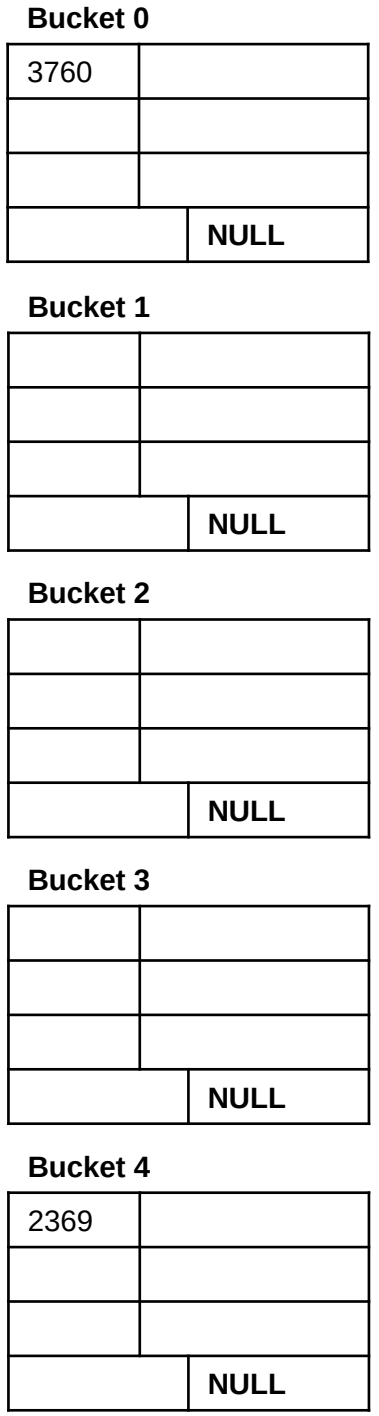
Question 3 (Answer) (1/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 =
  - 4692 % 5 =
  - 4871 % 5 =
  - 5659 % 5 =
  - 1821 % 5 =
  - 1074 % 5 =
  - 7115 % 5 =



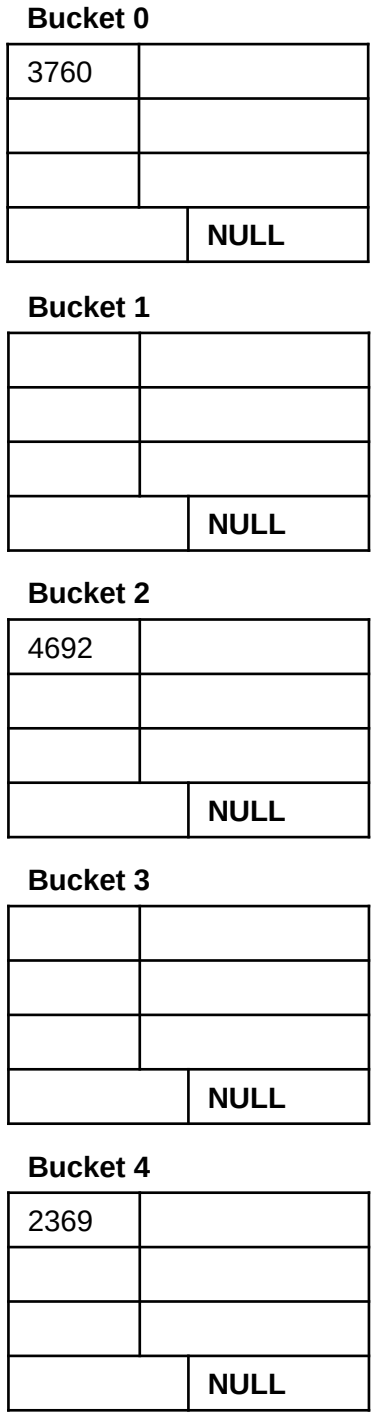
Question 3 (Answer) (2/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 =
  - 4871 % 5 =
  - 5659 % 5 =
  - 1821 % 5 =
  - 1074 % 5 =
  - 7115 % 5 =



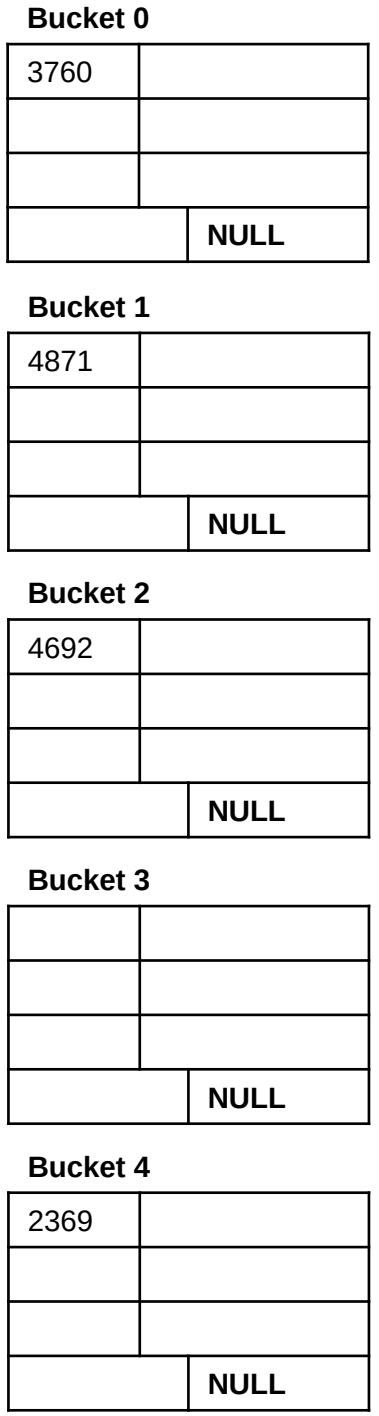
Question 3 (Answer) (3/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 =
  - 5659 % 5 =
  - 1821 % 5 =
  - 1074 % 5 =
  - 7115 % 5 =



Question 3 (Answer) (4/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 = 1
  - 5659 % 5 =
  - 1821 % 5 =
  - 1074 % 5 =
  - 7115 % 5 =



Question 3 (Answer) (5/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 = 1
  - 5659 % 5 = 4
  - 1821 % 5 =
  - 1074 % 5 =
  - 7115 % 5 =

Bucket 0

3760	
	NULL

Bucket 1

4871	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

	NULL

Bucket 4

2369	
5659	
	NULL

Question 3 (Answer) (6/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 = 1
  - 5659 % 5 = 4
  - 1821 % 5 = 1
  - 1074 % 5 =
  - 7115 % 5 =

Bucket 0

3760	
	NULL

Bucket 1

4871	
1821	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

	NULL

Bucket 4

2369	
5659	
	NULL

Question 3 (Answer) (7/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 = 1
  - 5659 % 5 = 4
  - 1821 % 5 = 1
  - 1074 % 5 = 4
  - 7115 % 5 =

Bucket 0

3760	
	NULL

Bucket 1

4871	
1821	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

	NULL

Bucket 4

2369	
5659	
1074	
	NULL

Question 3 (Answer) (8/8)

- (a) Static hashing with 5 buckets, each of which contains at most 3 records
  - 2369 % 5 = 4
  - 3760 % 5 = 0
  - 4692 % 5 = 2
  - 4871 % 5 = 1
  - 5659 % 5 = 4
  - 1821 % 5 = 1
  - 1074 % 5 = 4
  - 7115 % 5 = 0

Bucket 0

3760	
7115	
	NULL

Bucket 1

4871	
1821	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

	NULL

Bucket 4

2369	
5659	
1074	
	NULL

Question 3 (Answer) (1/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 =
  - 3945 % 5 =
  - 4759 % 5 =
  - 6975 % 5 =
  - 4981 % 5 =
  - 9206 % 5 =

Bucket 0

3760	
7115	
1620	
	NULL

Bucket 1

4871	
1821	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

	NULL

Bucket 4

2369	
5659	
1074	
	NULL

Question 3 (Answer) (2/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 =
  - 4759 % 5 =
  - 6975 % 5 =
  - 4981 % 5 =
  - 9206 % 5 =

Bucket 0

3760	
7115	
1620	
	NULL

Bucket 1

4871	
1821	
	NULL

Bucket 2

4692	
	NULL

Bucket 3

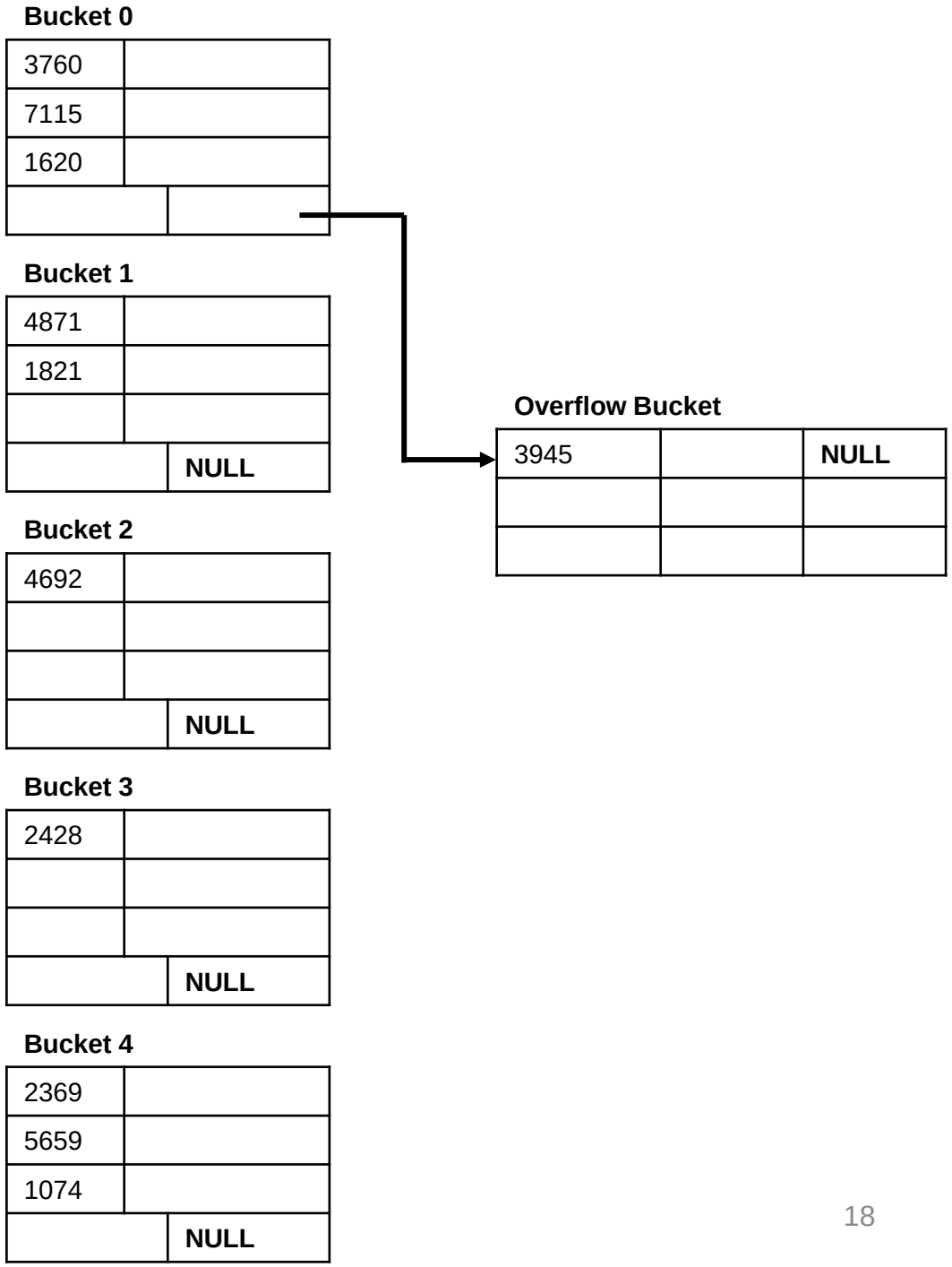
2428	
	NULL

Bucket 4

2369	
5659	
1074	
	NULL

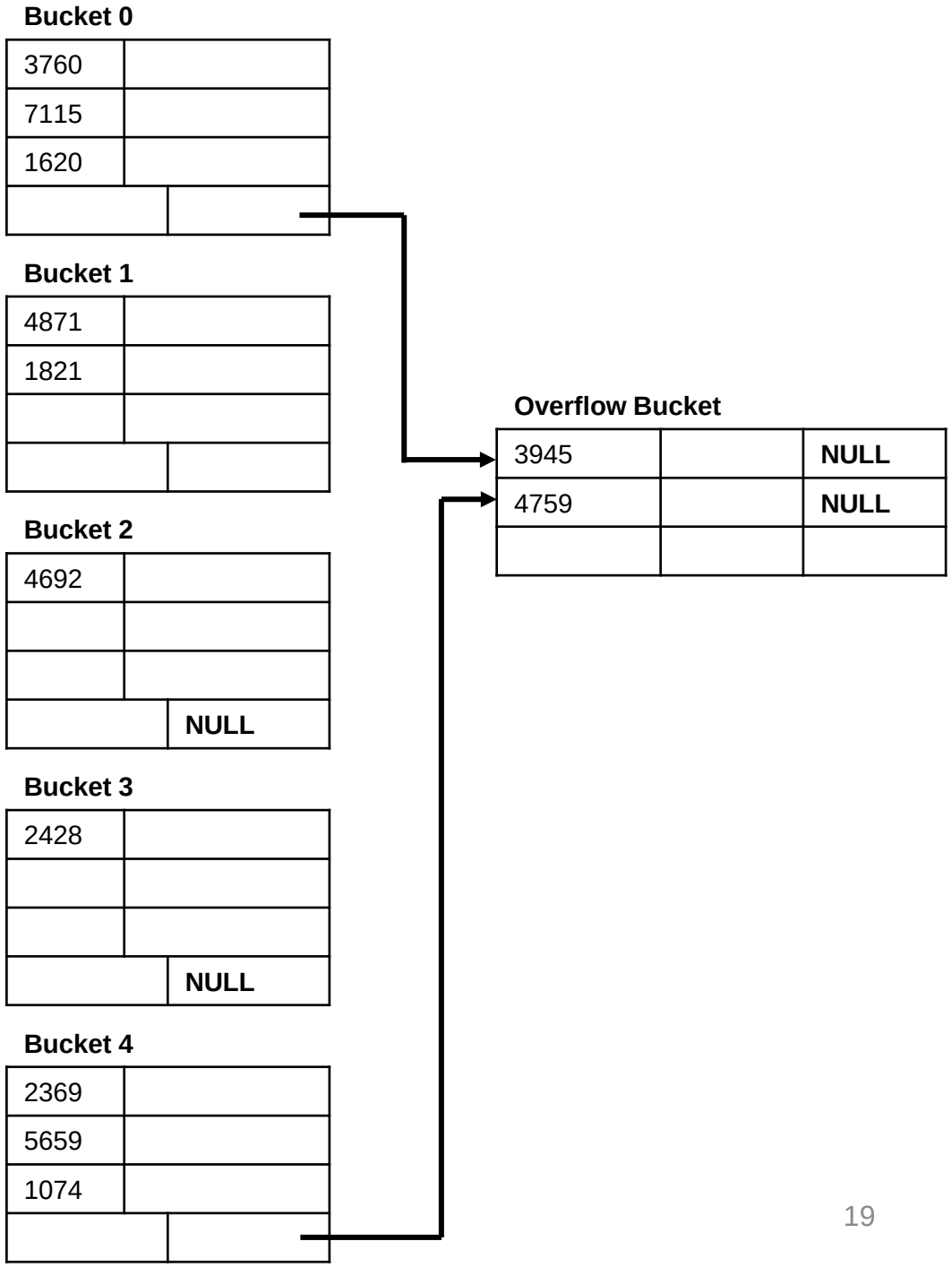
Question 3 (Answer) (3/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 = 0
  - 4759 % 5 =
  - 6975 % 5 =
  - 4981 % 5 =
  - 9206 % 5 =



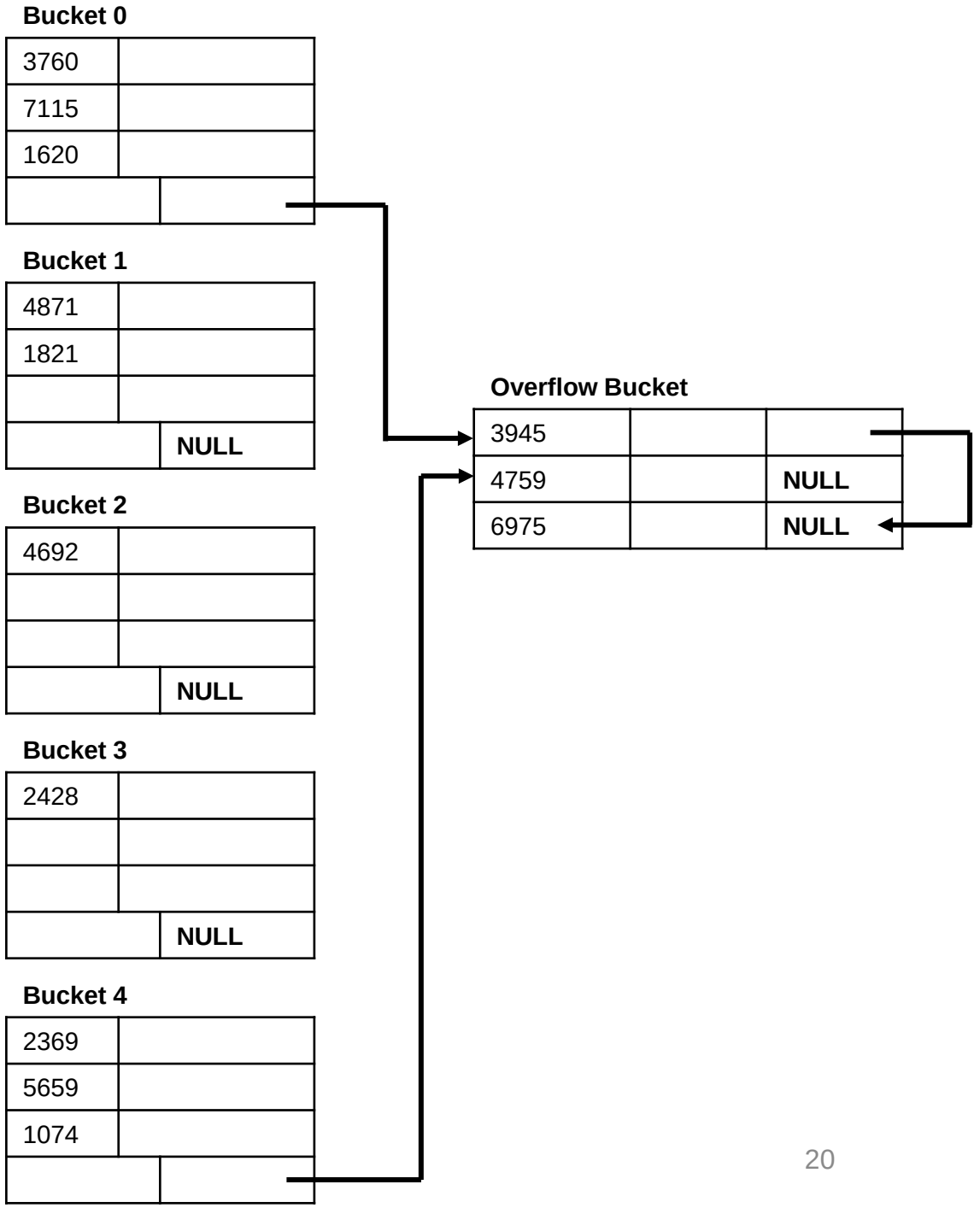
Question 3 (Answer) (4/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 = 0
  - 4759 % 5 = 4
  - 6975 % 5 =
  - 4981 % 5 =
  - 9206 % 5 =



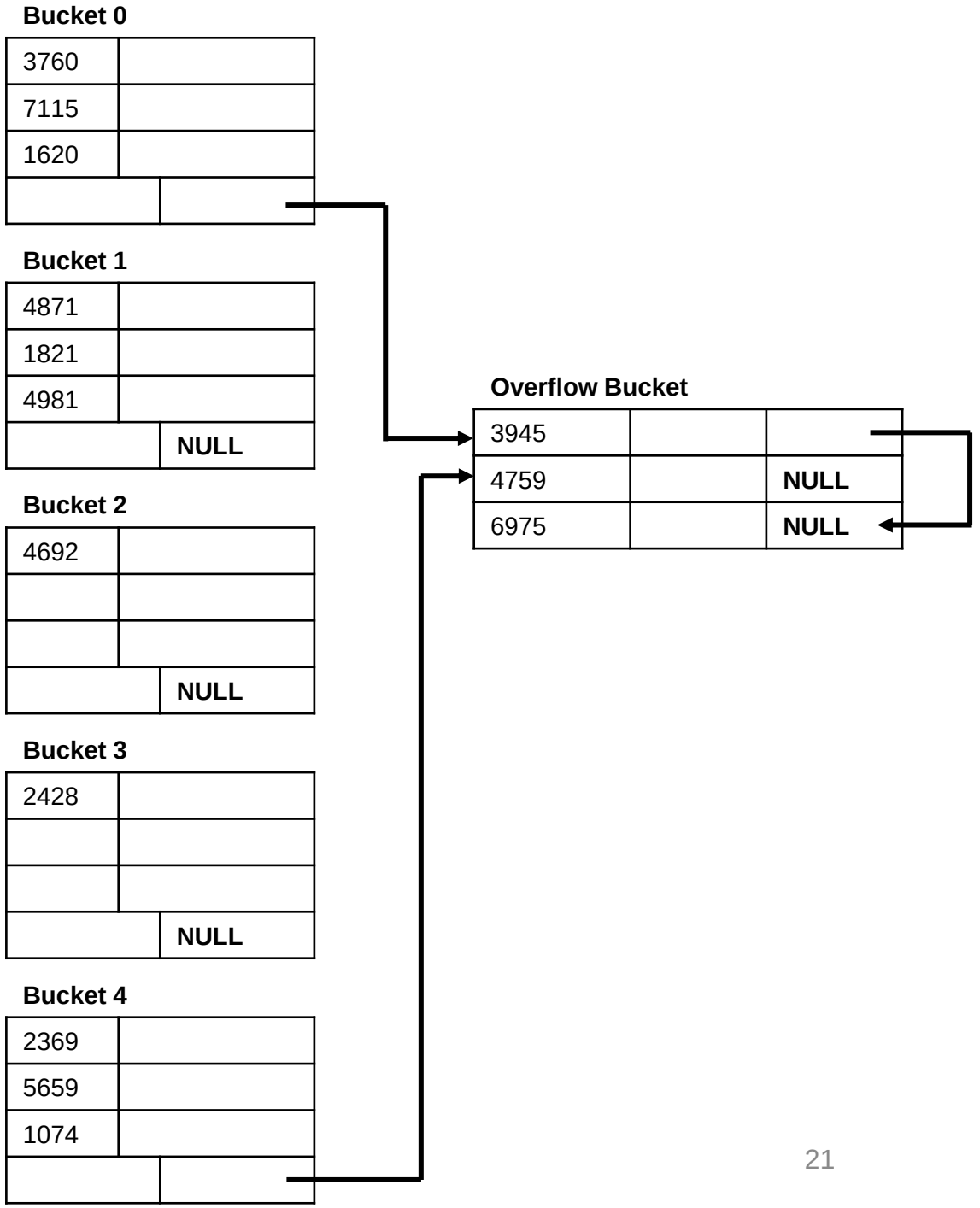
Question 3 (Answer) (5/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 = 0
  - 4759 % 5 = 4
  - 6975 % 5 = 0
  - 4981 % 5 =
  - 9206 % 5 =



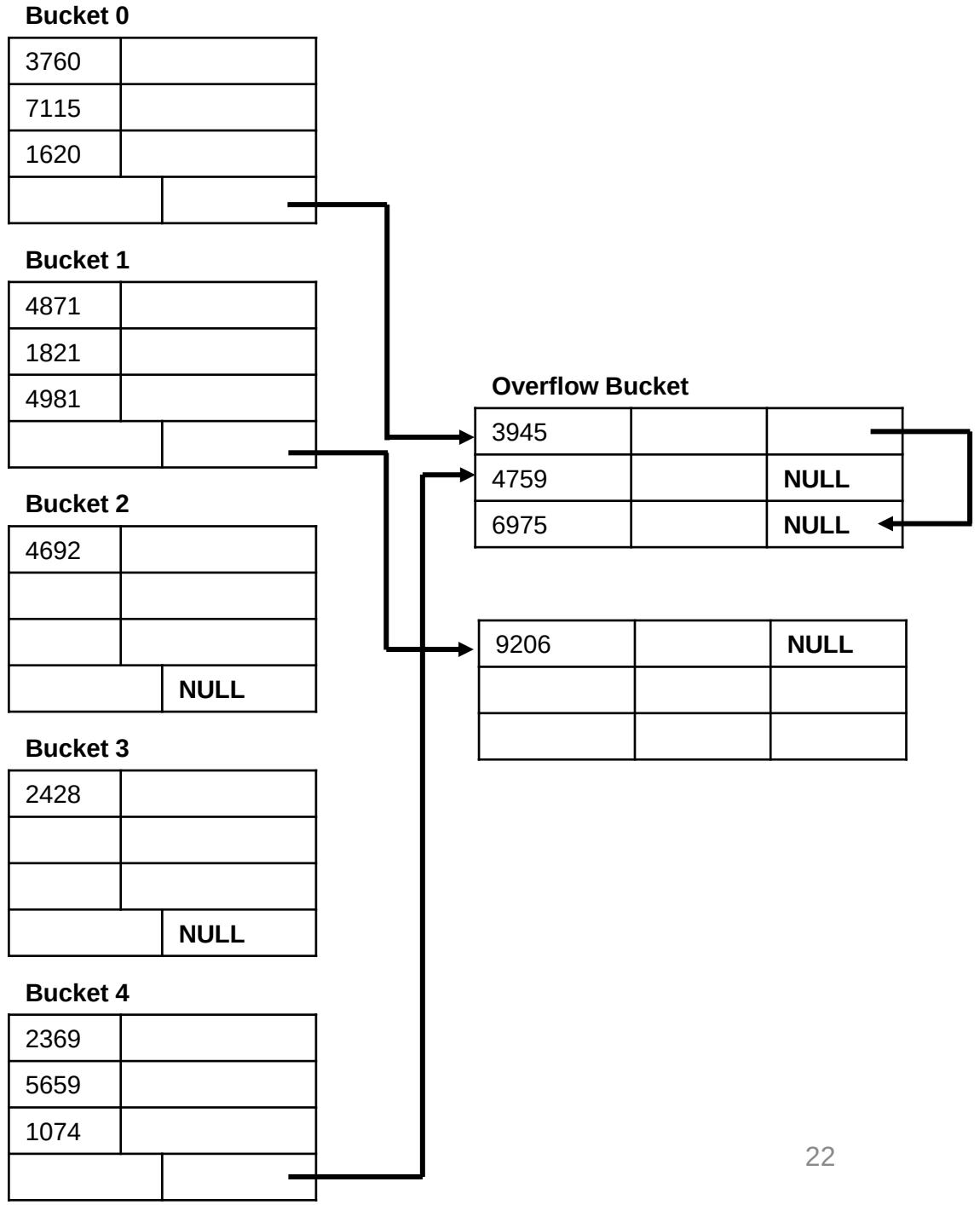
Question 3 (Answer) (6/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 = 0
  - 4759 % 5 = 4
  - 6975 % 5 = 0
  - 4981 % 5 = 1
  - 9206 % 5 =



Question 3 (Answer) (7/7)

- (b) Overflow handling
  - 1620 % 5 = 0
  - 2428 % 5 = 3
  - 3945 % 5 = 0
  - 4759 % 5 = 4
  - 6975 % 5 = 0
  - 4981 % 5 = 1
  - 9206 % 5 = 1



Question 4

- In extendible hashing, how many hash codes can you have in maximum if the global depth is 3?

Question 4 (Answer)

- 8 hash codes (000, 001, 010, 100, 011, 110, 101, 111)