CSC212

Data Structure

Lecture 20

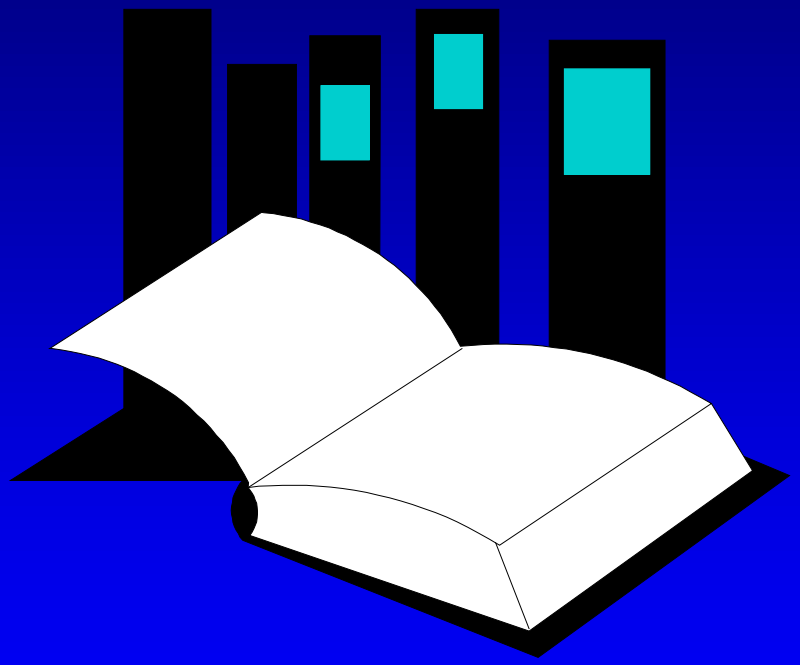
Hashing

Instructor: George Wolberg

Department of Computer Science

City College of New York

Hash Tables

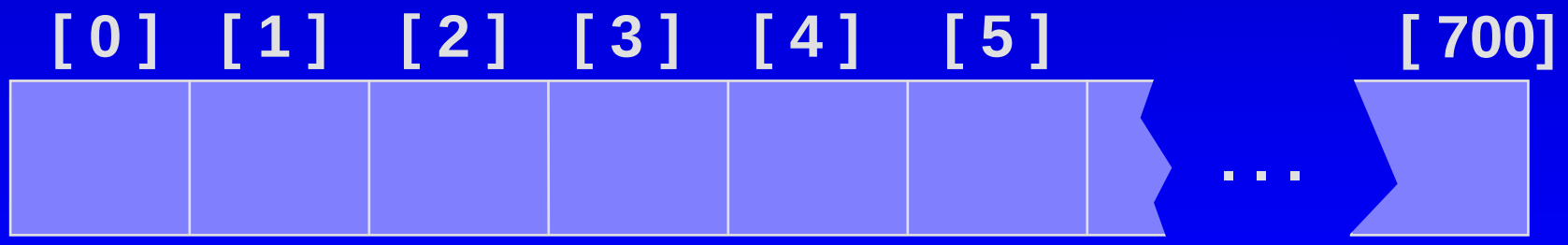


Data Structures and Other Objects Using C++

- Chapter 12 discusses several ways of storing information in an array, and later searching for the information.
- Hash tables are a common approach to the storing/searching problem.
- This presentation introduces hash tables.

What is a Hash Table ?

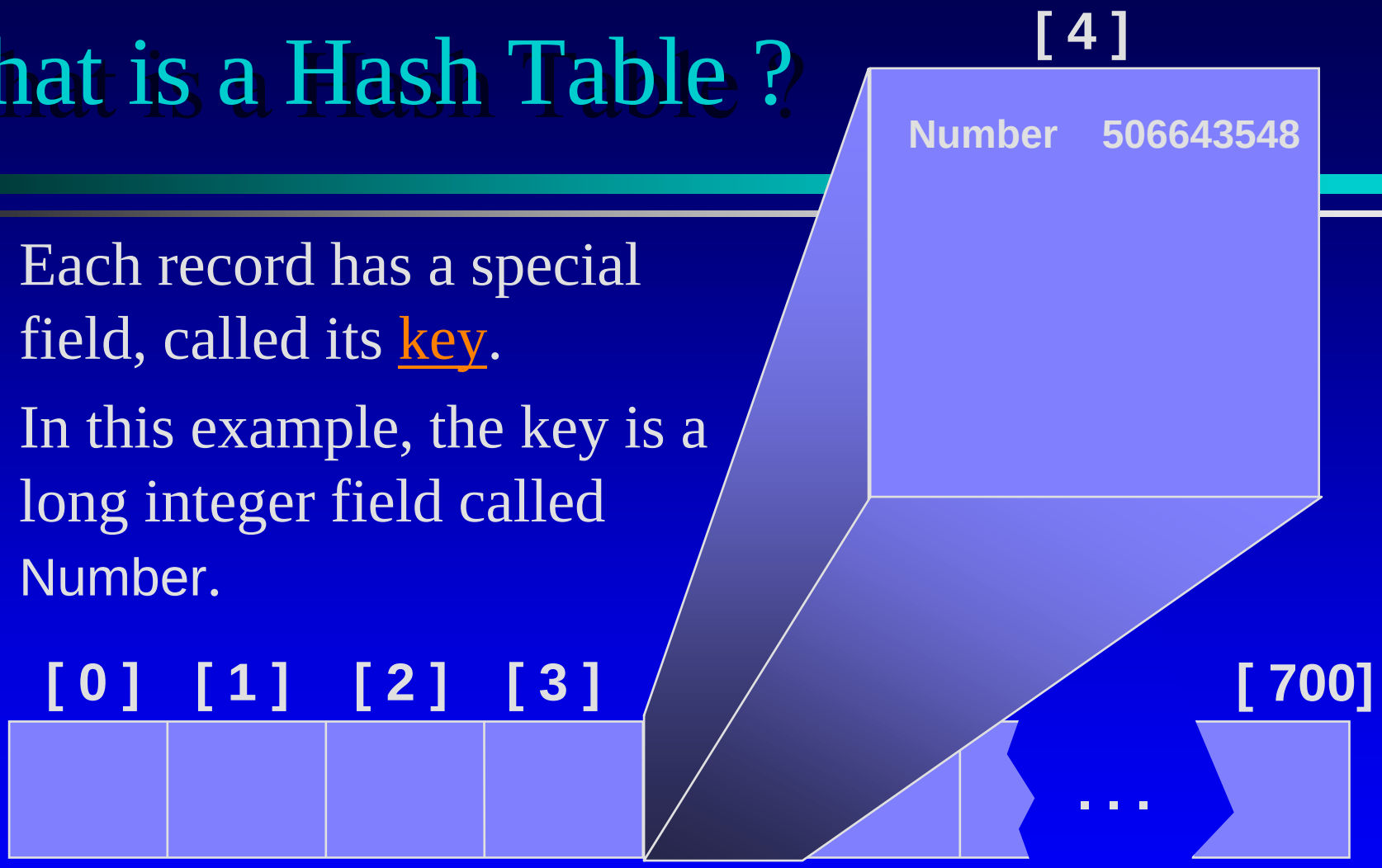
- The simplest kind of hash table is an array of records.
- This example has 701 records.



An array of records

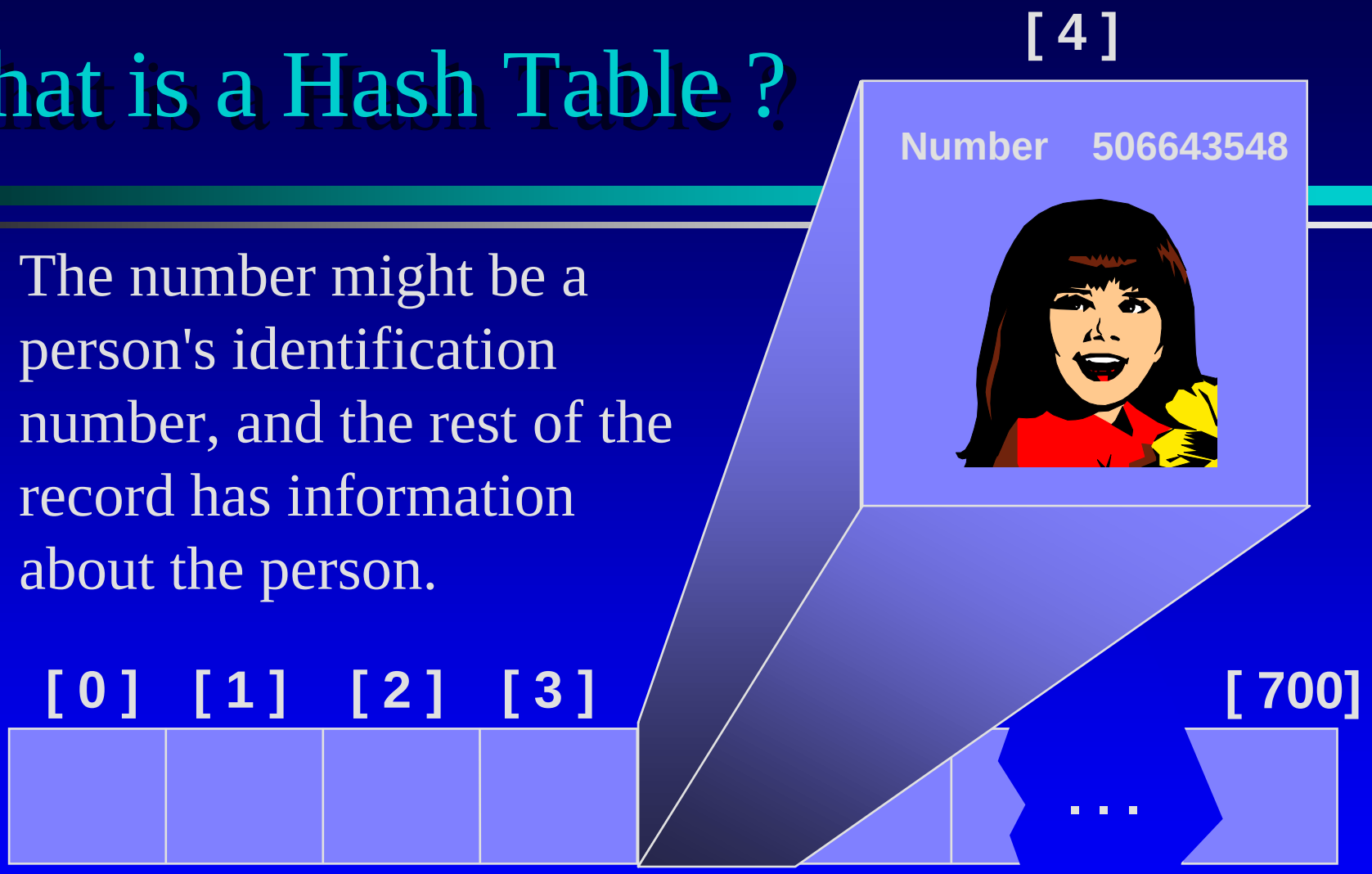
What is a Hash Table ?

- Each record has a special field, called its key.
- In this example, the key is a long integer field called Number.



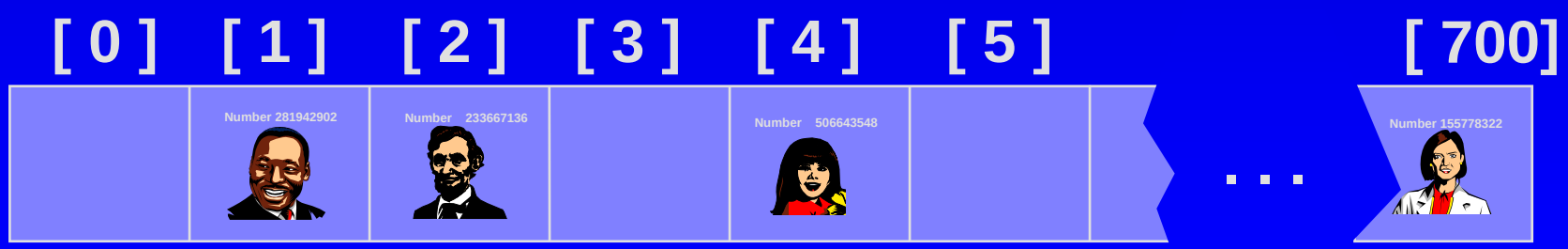
What is a Hash Table ?

- The number might be a person's identification number, and the rest of the record has information about the person.



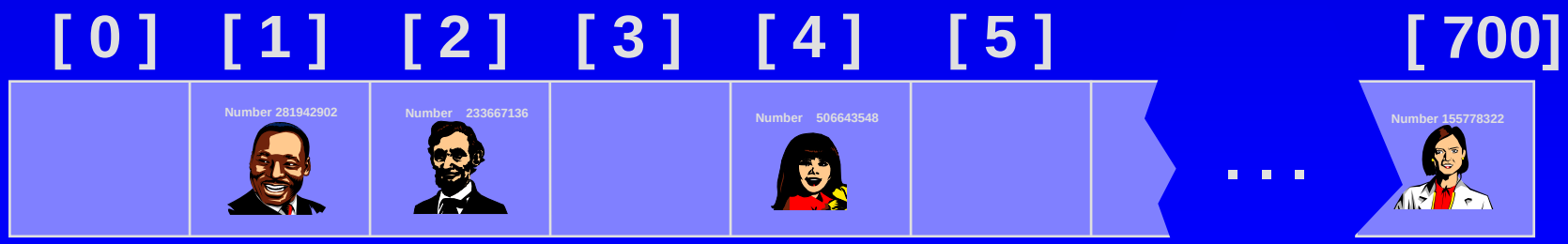
What is a Hash Table ?

- When a hash table is in use, some spots contain valid records, and other spots are "empty".



Inserting a New Record

- In order to insert a new record, the **key** must somehow be **converted to** an array **index**.
- The index is called the **hash value** of the key.

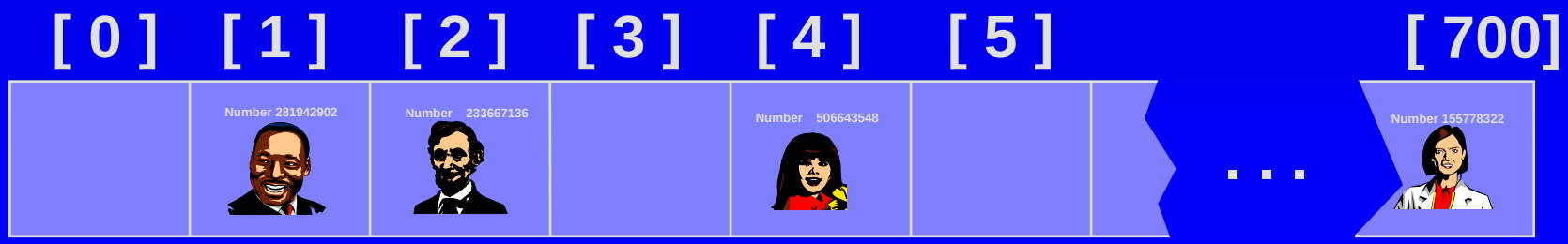


Inserting a New Record

- Typical way create a hash value:

  (Number mod 701)

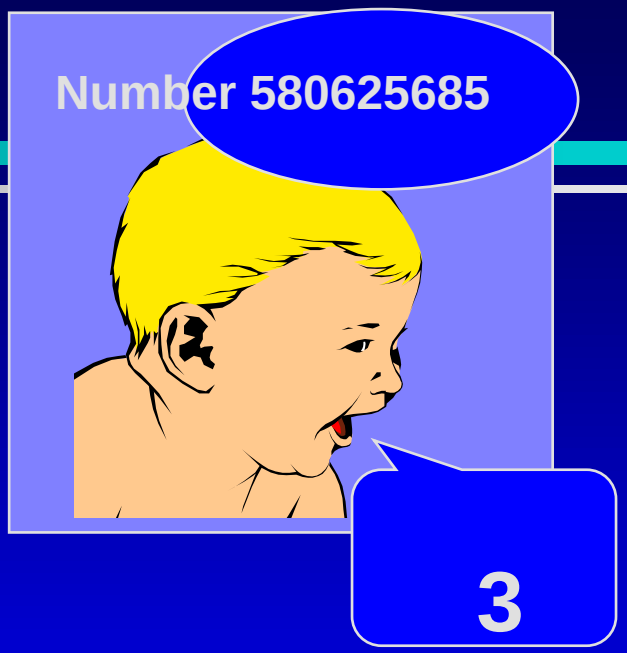
What is (580625685 mod 701) ?



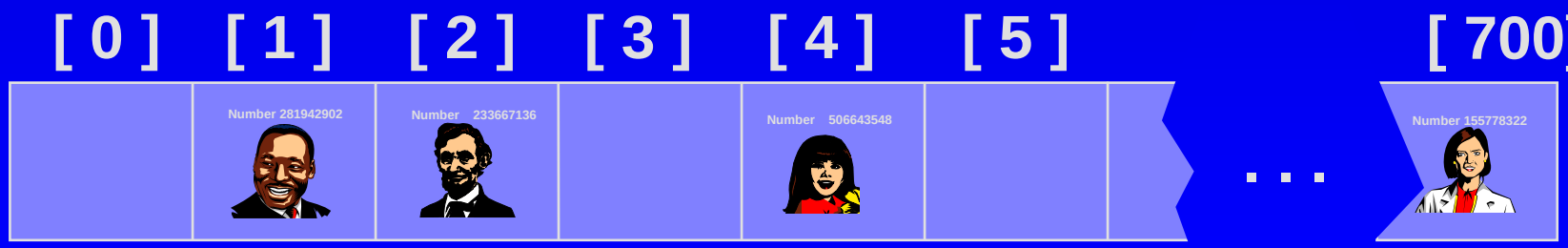
Inserting a New Record

- Typical way to create a hash value:

  (Number mod 701)

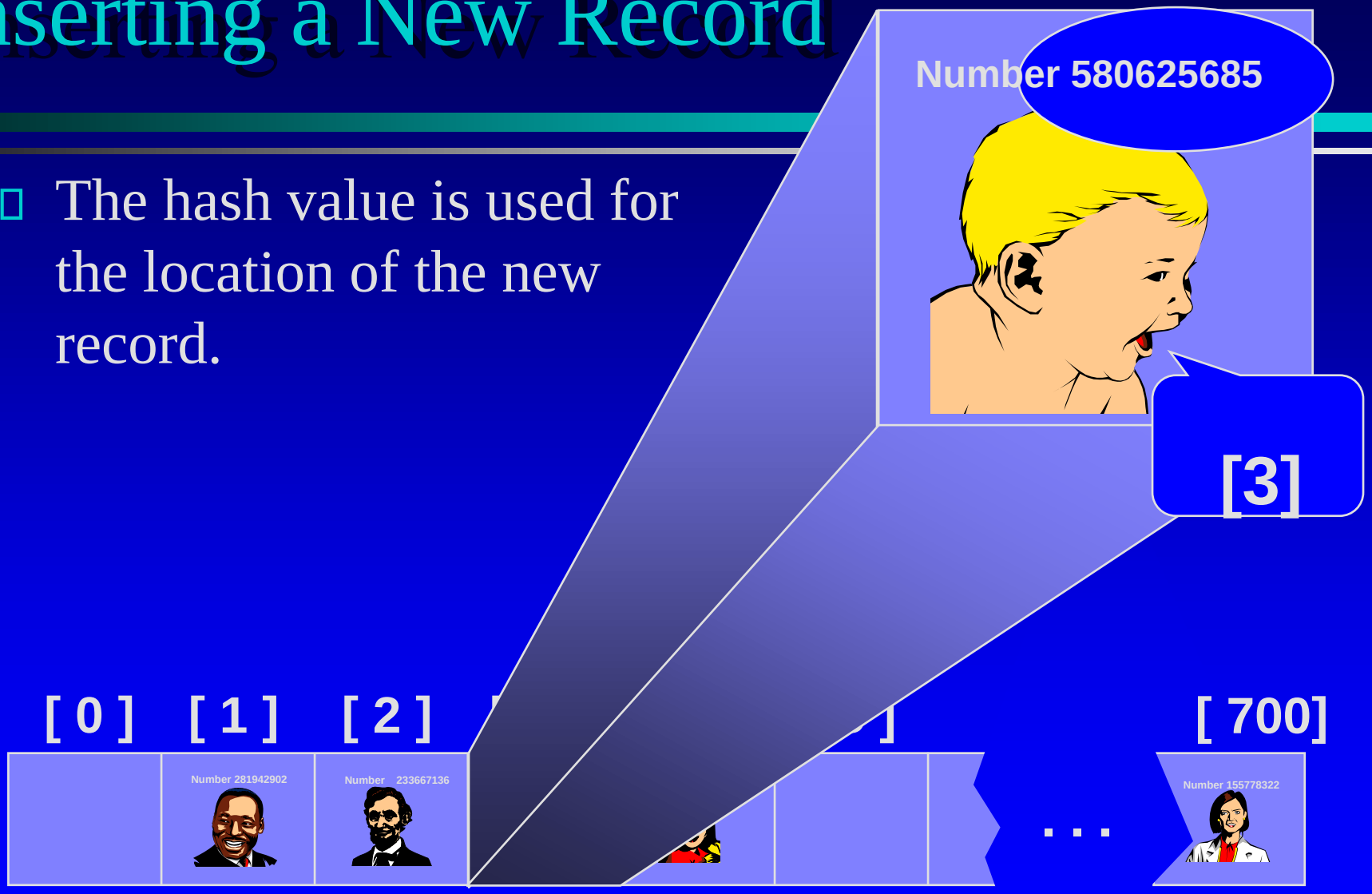


What is (580625685 mod 701) ?



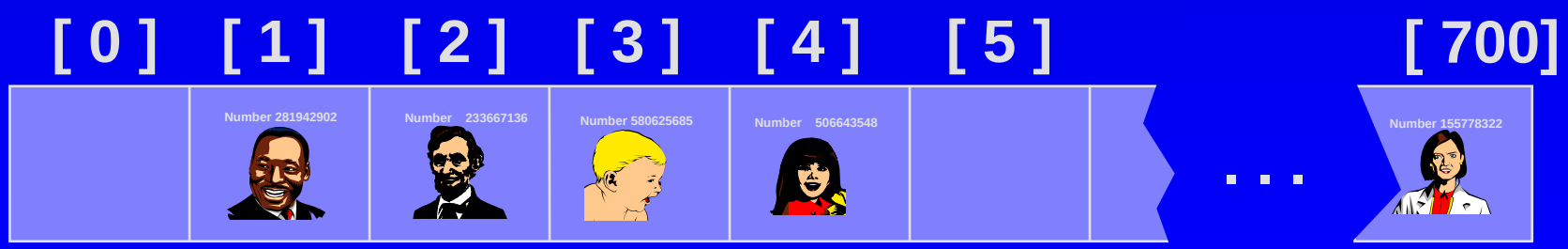
Inserting a New Record

- The hash value is used for the location of the new record.



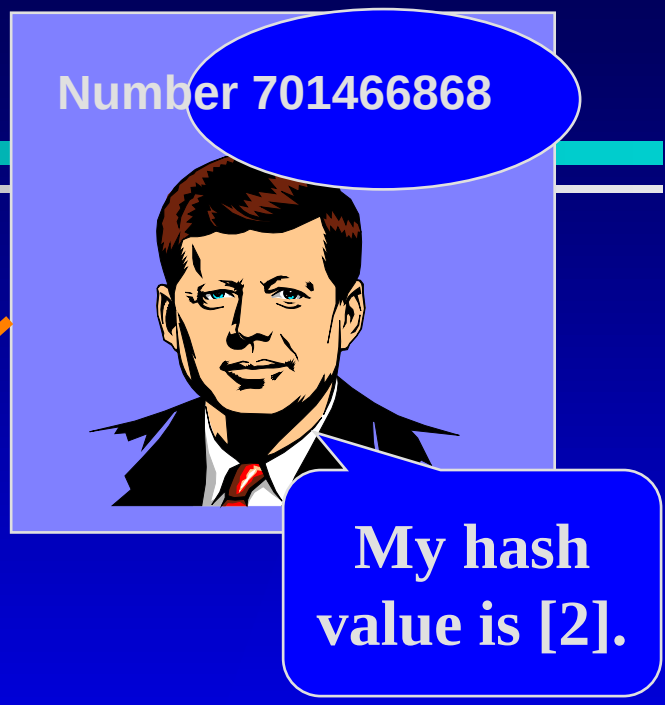
Inserting a New Record

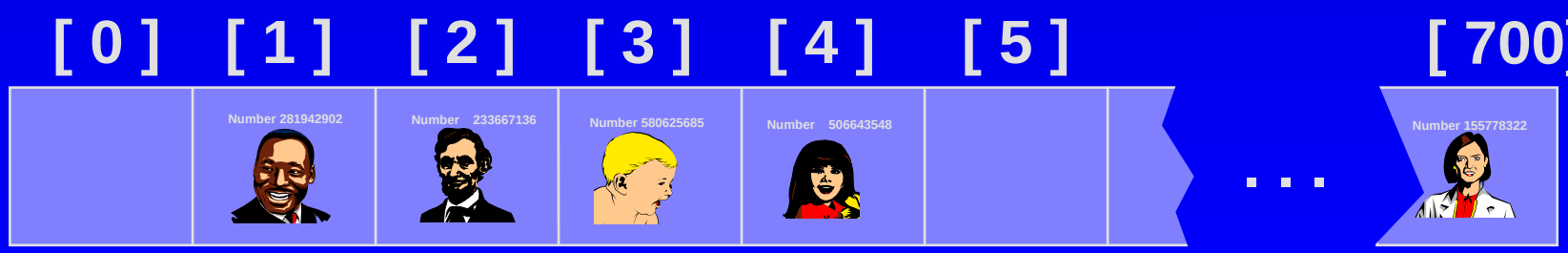
- The hash value is used for the location of the new record.



Collisions

- Here is another new record to insert, with a hash value of 2.



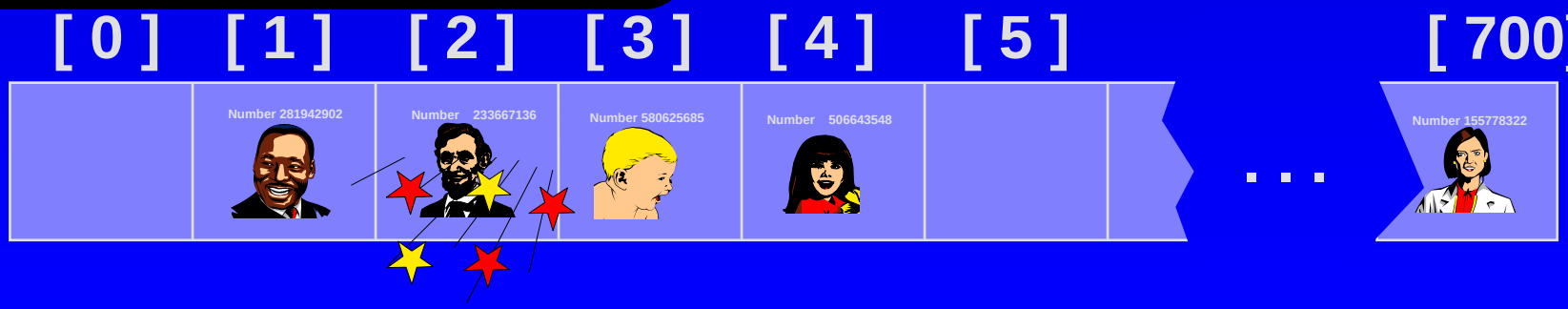


Collisions

- Here is another new record to insert, with a hash value of 2.

My hash value is [2].

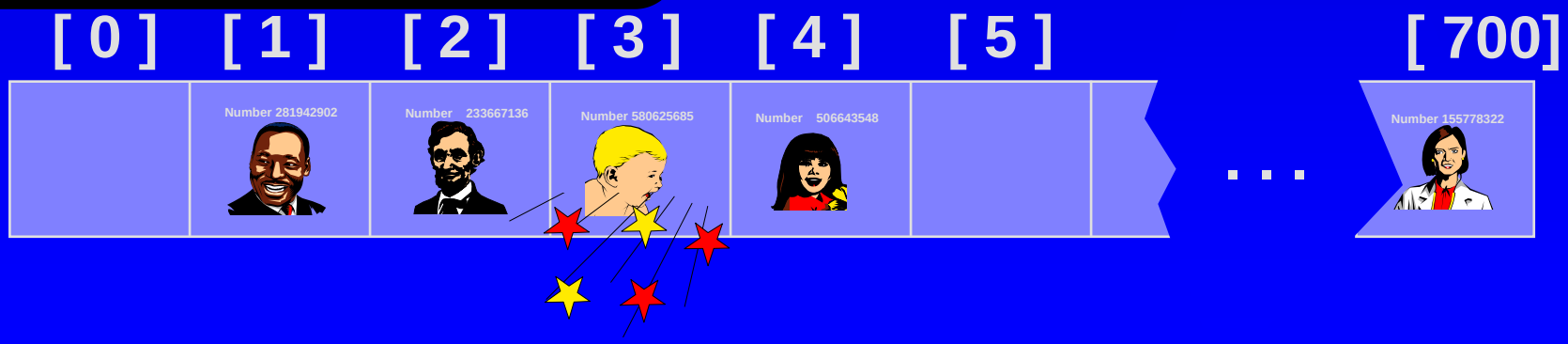
When a collision occurs, move forward until you find an empty spot.



Collisions

- This is called a **collision**, because there is already another valid record at [2].

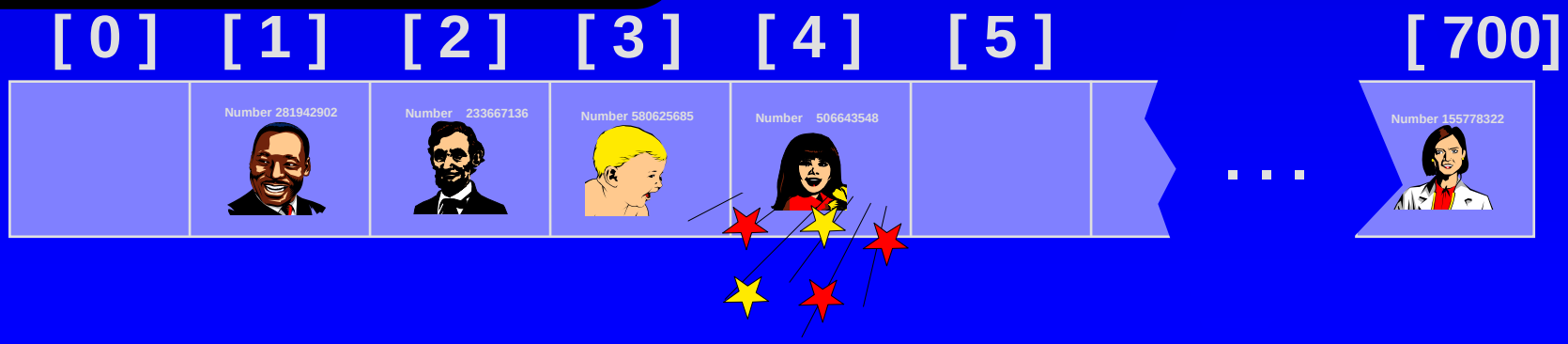
When a collision occurs, move forward until you find an empty spot.



Collisions

- This is called a **collision**, because there is already another valid record at [2].

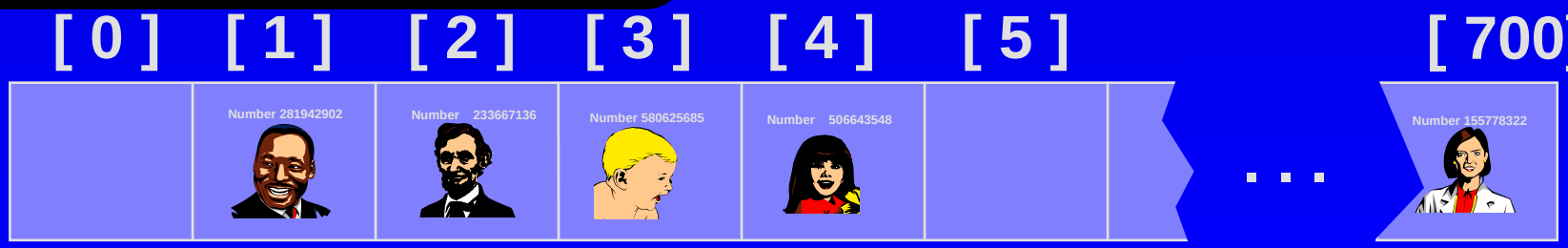
When a collision occurs, move forward until you find an empty spot.



Collisions

- This is called a **collision**, because there is already another valid record at [2].

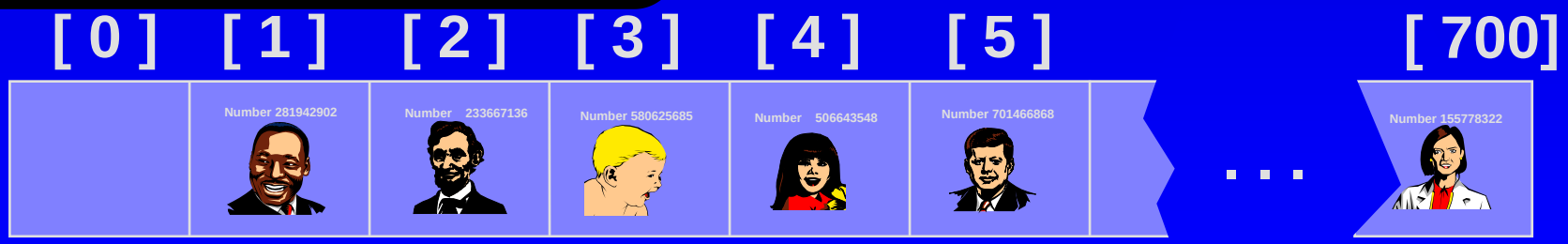
When a collision occurs, move forward until you find an empty spot.



Collisions

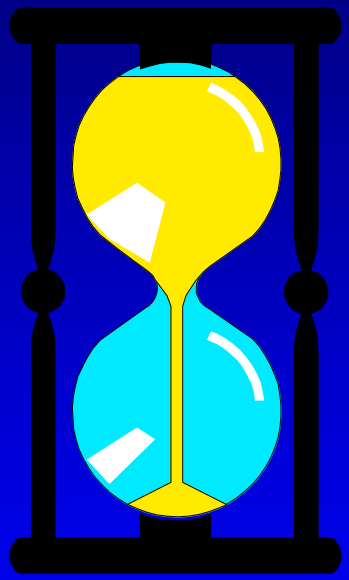
- This is called a **collision**, because there is already another valid record at [2].

The new record goes in the empty spot.



A Quiz

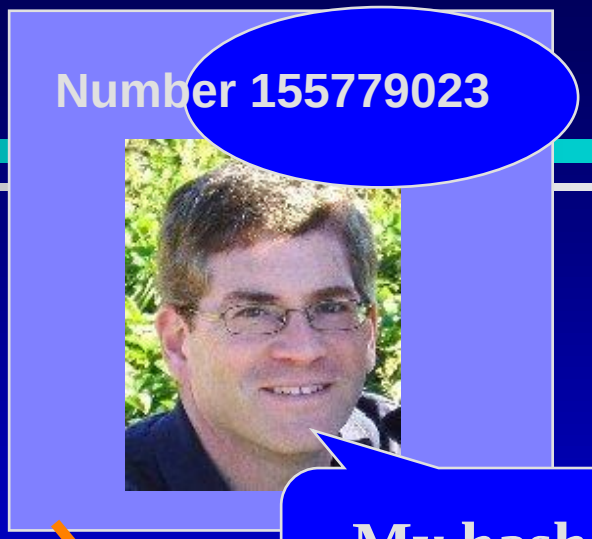
Where would you be placed in this table, if there is no collision? Use your social security number or some other favorite number.



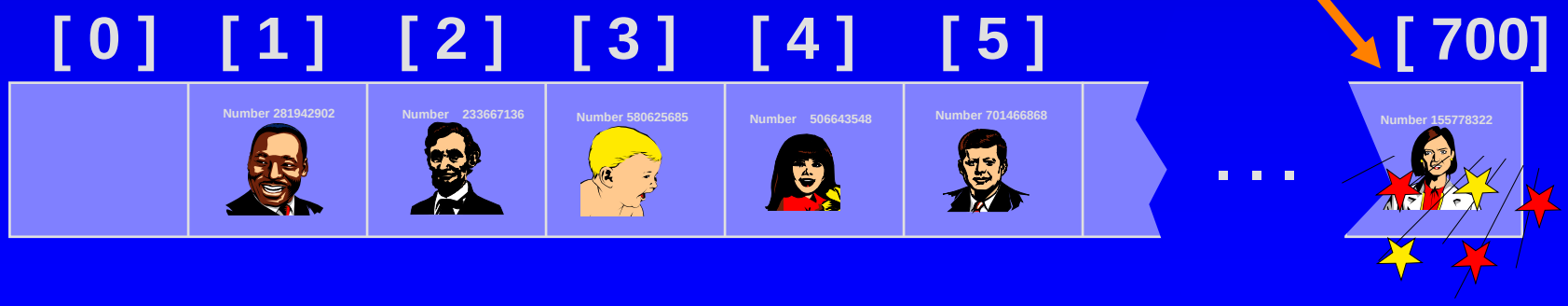
[0]	[1]	[2]	[3]	[4]	[5]		[700]
	Number 281942902 	Number 233667136 	Number 580625685 	Number 506643548 	Number 701466868 	. . .	Number 155778322 

Another Kind of Collision

Where would you be placed in this table, if there is no collision? Use your social security number or some other favorite number.

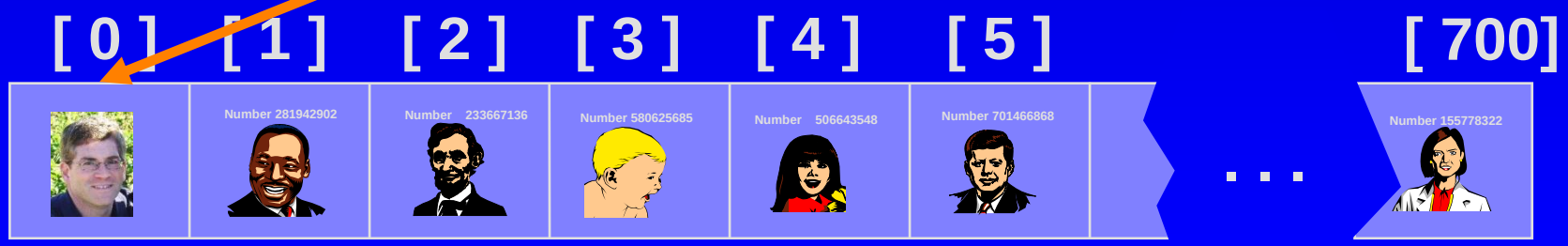


My hash value is [700].



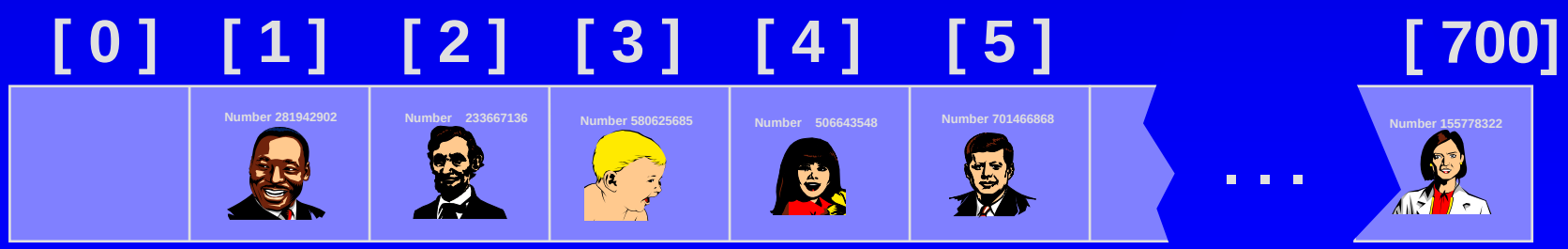
Another Kind of Collision

Where would you be placed in this table, if there is no collision? Use your social security number or some other favorite number.



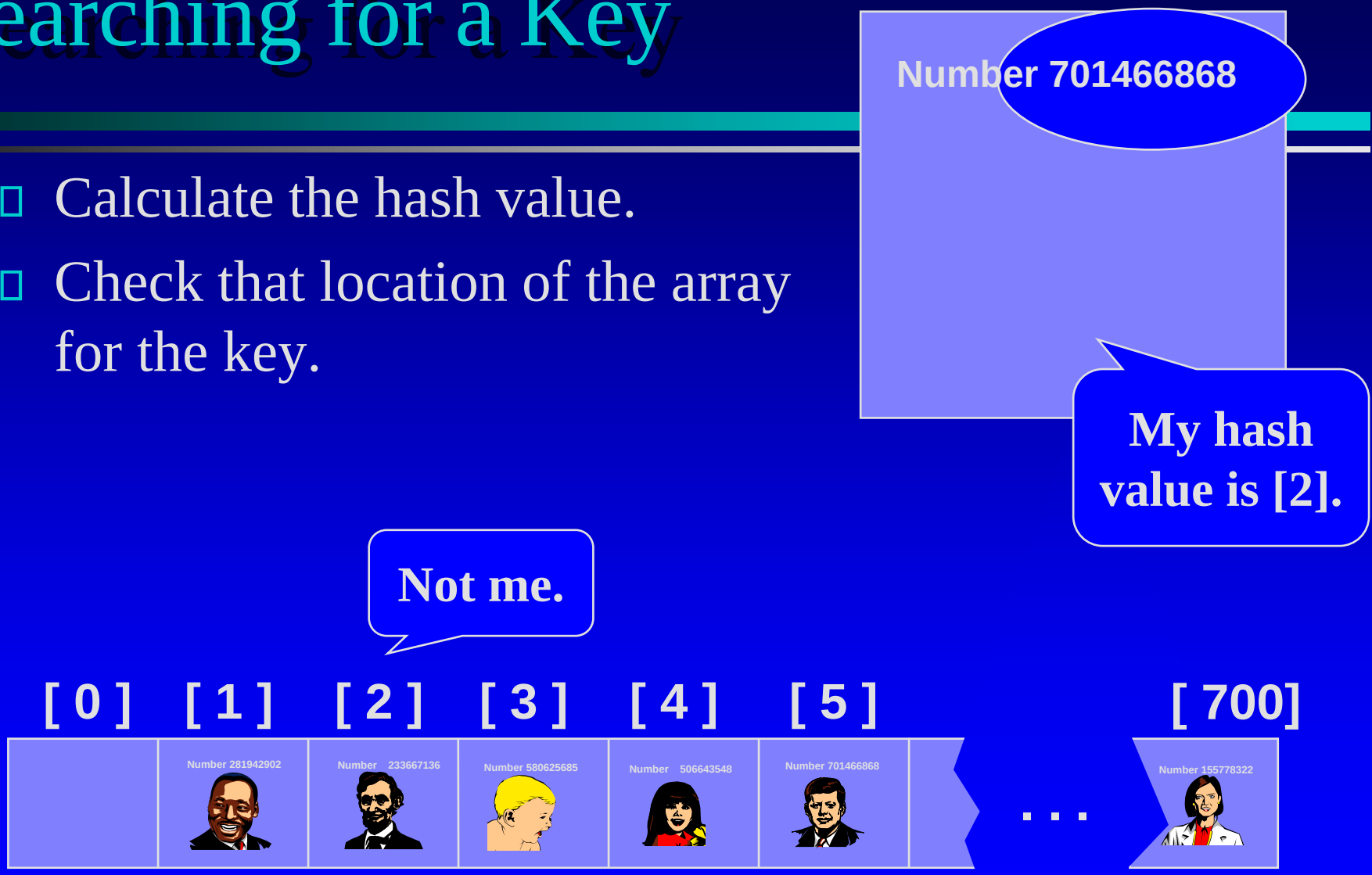
Searching for a Key

- The data that's attached to a key can be found fairly quickly.



Searching for a Key

- Calculate the hash value.
- Check that location of the array for the key.

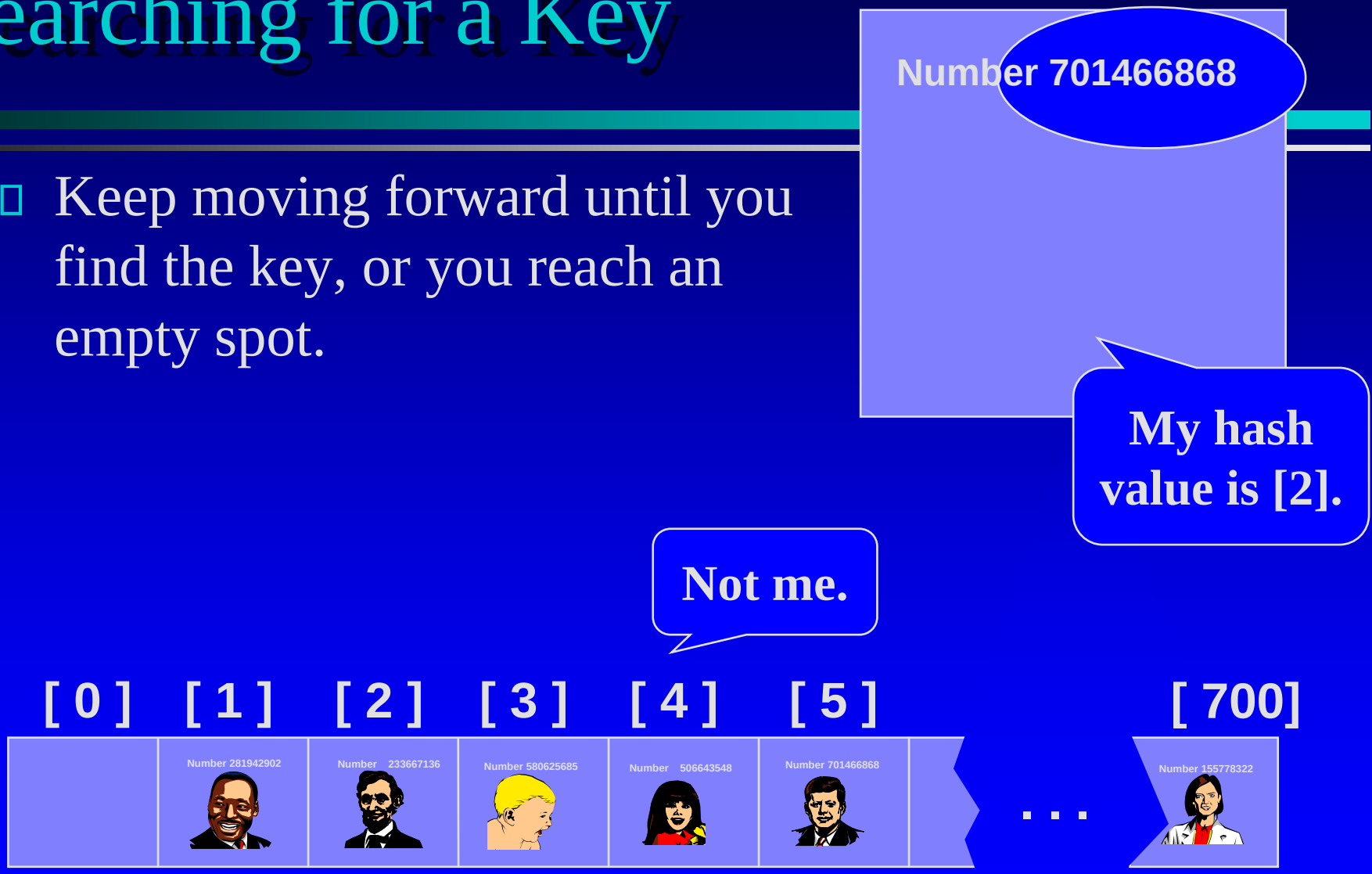


Searching for a Key

- Keep moving forward until you find the key, or you reach an empty spot.

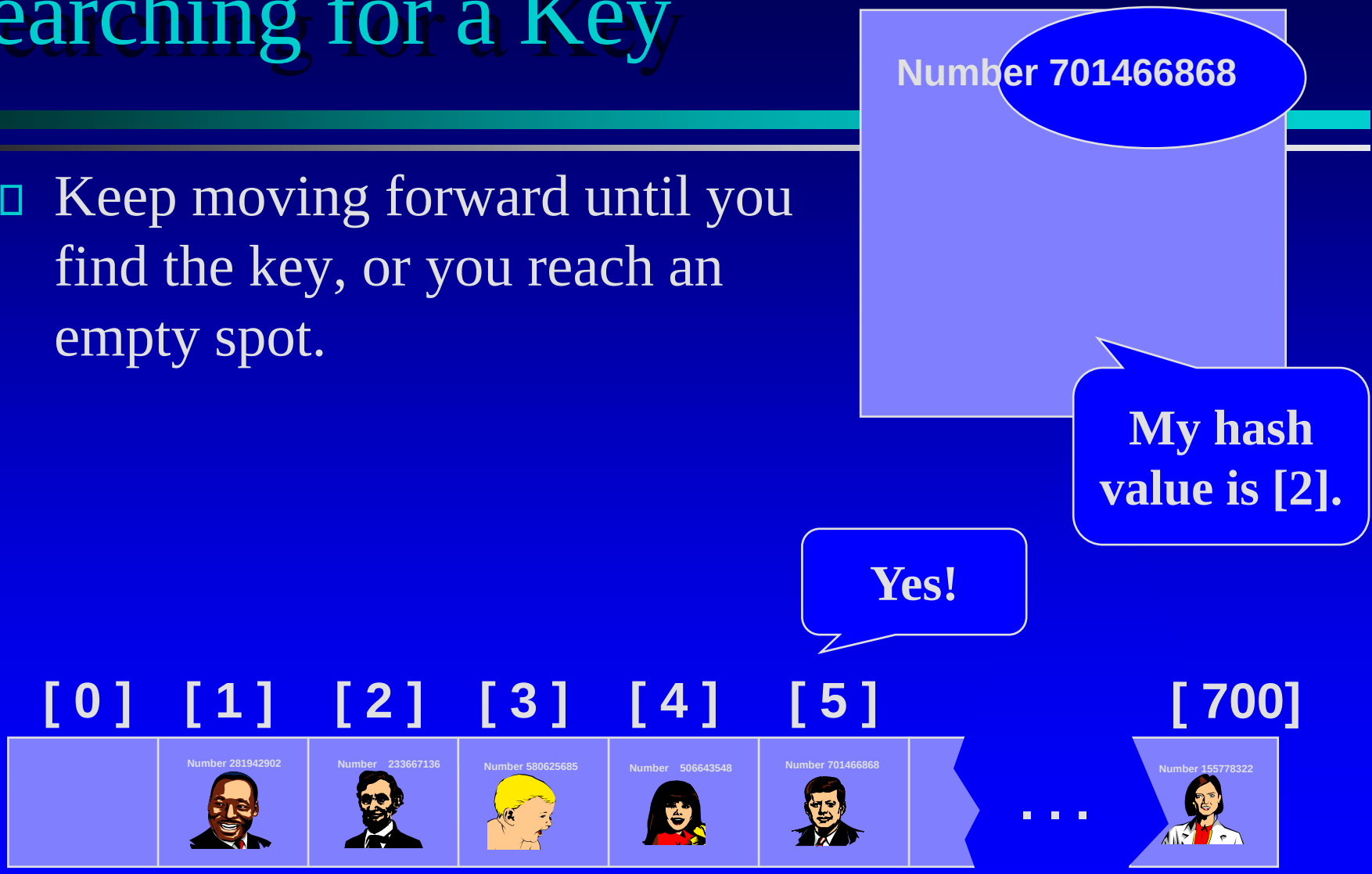
Searching for a Key

- Keep moving forward until you find the key, or you reach an empty spot.



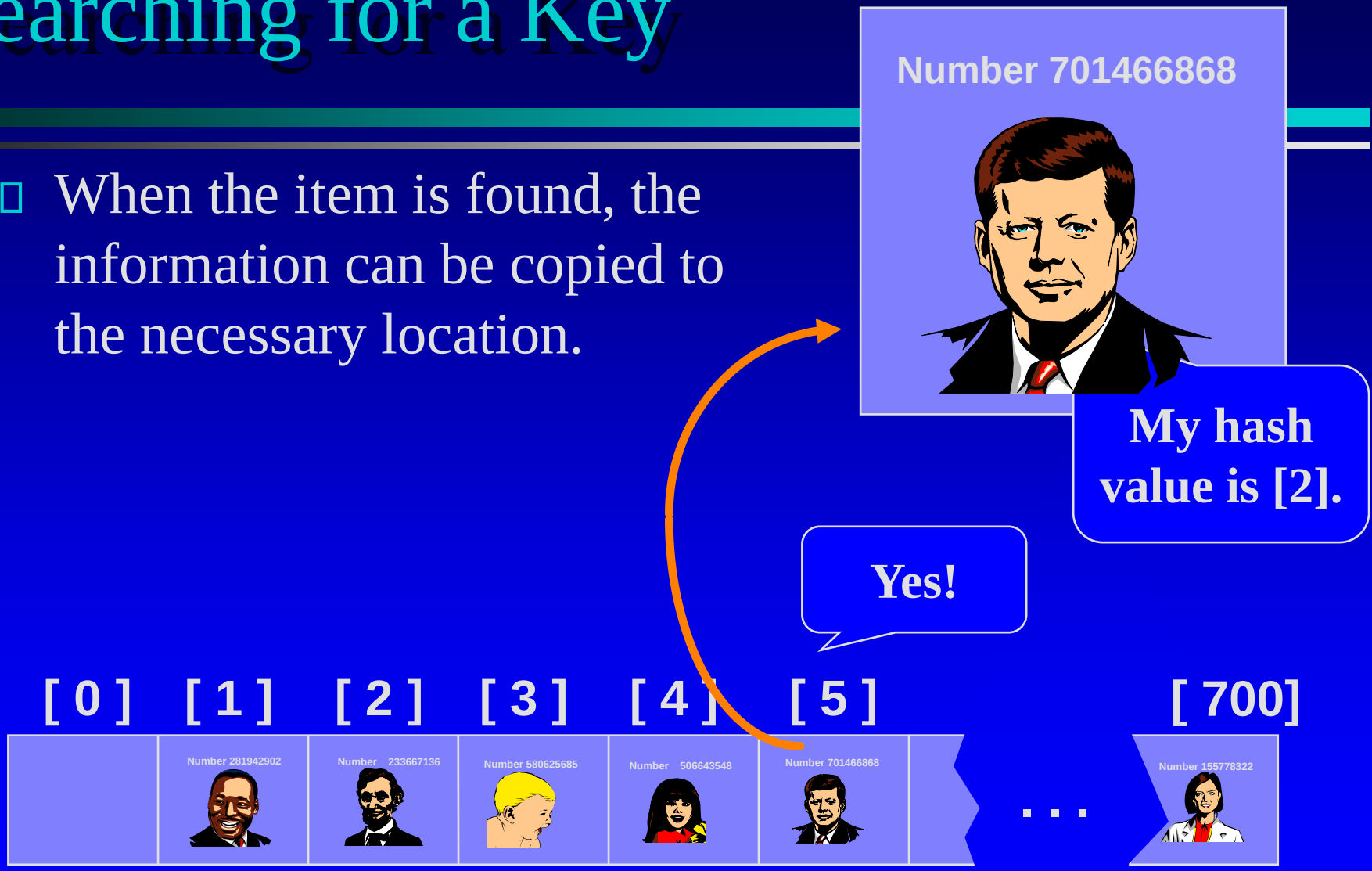
Searching for a Key

- Keep moving forward until you find the key, or you reach an empty spot.



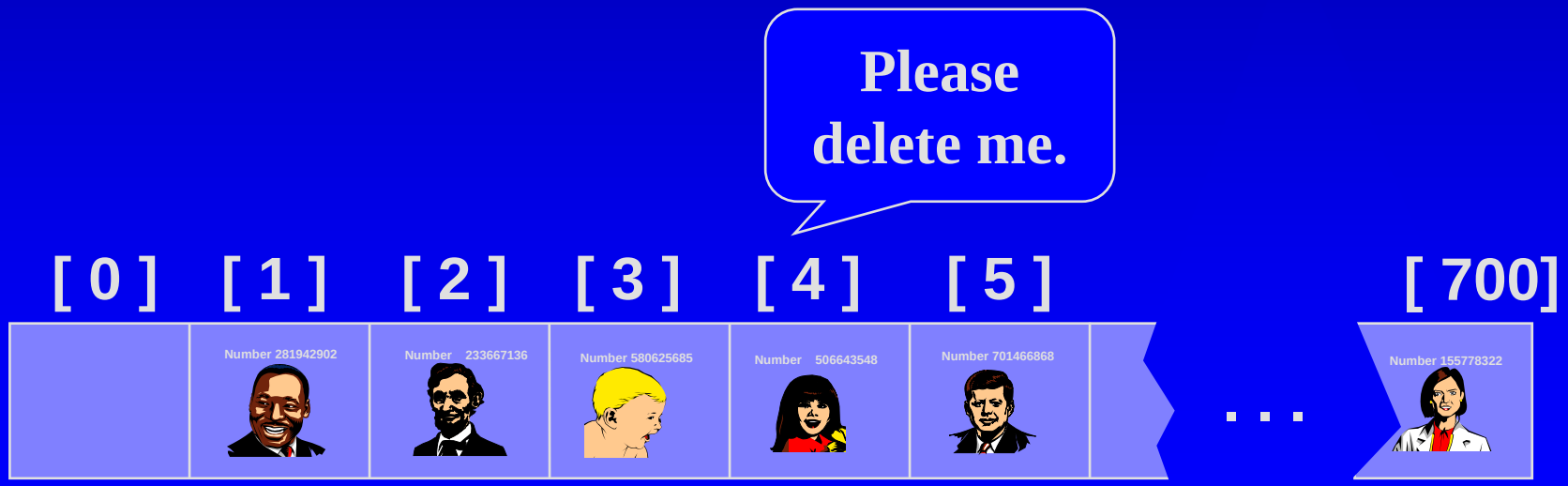
Searching for a Key

- When the item is found, the information can be copied to the necessary location.



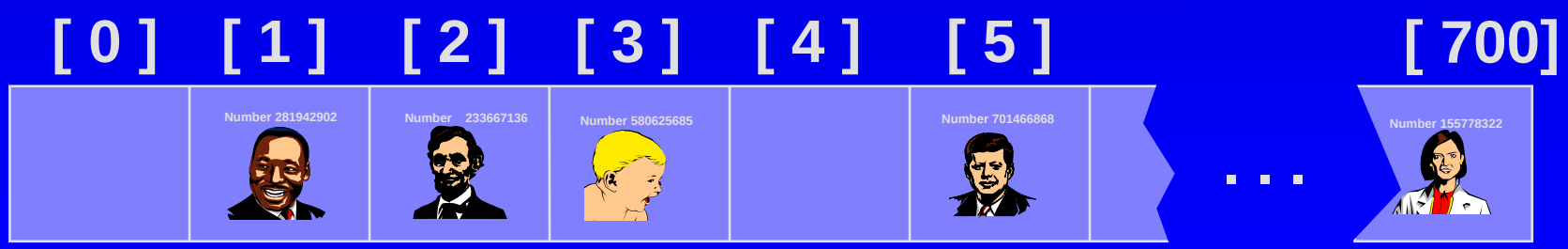
Deleting a Record

- Records may also be deleted from a hash table.



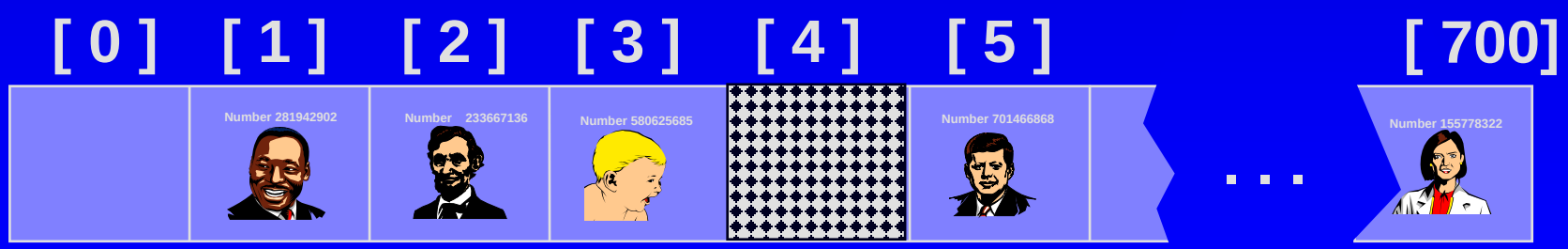
Deleting a Record

- Records may also be deleted from a hash table.
- But the location must not be left as an ordinary "empty spot" since that could interfere with searches.



Deleting a Record

- Records may also be deleted from a hash table.
- But the location must not be left as an ordinary "empty spot" since that could interfere with searches.
- The location must be marked in some special way so that a search can tell that the spot used to have something in it.



Time Analysis

- Without any collisions
  - constant
- With collisions
  - O(k) where k is the average collisions for items
  - k << n, size of the problem

Improving Hashing

- Size of the hashing table when using *division hash function*
  - prime number in the form of 4k+3
- Other hashing functions
  - mid-square, multiplicative
- Double hashing (instead of linear probing)
  - the 2nd hash function for stepping through the array
- Chained hashing
  - using a linked list for each component of the hash table

Summary

- Hash tables store a collection of records with keys.
- The location of a record depends on the hash value of the record's key.
- When a collision occurs, the next available location is used.
- Searching for a particular key is generally quick.
- When an item is deleted, the location must be marked in a special way, so that the searches know that the spot used to be used.

Hash Table Exercise

Five records of my past students

- Create a small hash table with size 5 (indexes 0 to 4).

- Insert the five items

- Remove Bill Clinton

- Do three searches (for Will Smith, Bill Clinton, and Elizabeth).

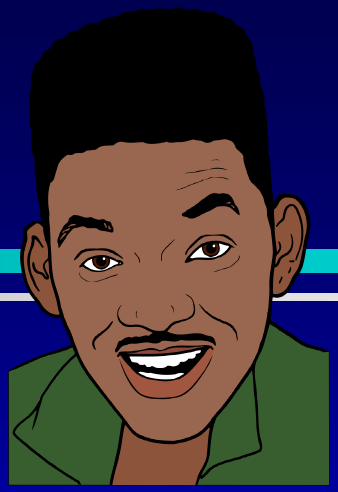
Kathy Martin
817339024

Took Data Structures in Fall 1993.
Grade A.

Hard worker. Always gets things done on time.

Currently working for ABC in New York City.

Will Smith
506643973



Took Data Structures in Fall 1995.
Grade A.

A bit of a goof-off, but he comes through in a pinch.

Currently saving the world from alien invasion.

William “Bill” Clinton
330220393

Took Data Structures in Fall 1995.
Grade B-.

Gets along with most
people well.

Been laid off even before the slowdown of the economy.

Elizabeth Windsor
092223340

Took Data Structures in Fall 1995.
Grade B-.

Prefers to be called “Elizabeth II” or “Her Majesty.” Has some family problems.

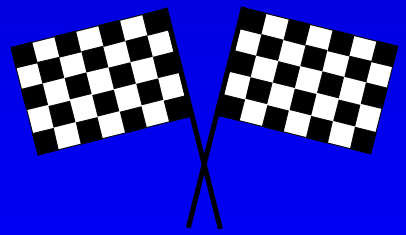
Currently working in public relations near London.

Al Einstein
699200102

Took CSCI 2270 in Fall 1995.
Grade F.

In spite of poor grade, I think there is good academic ability in Al.

Currently a well-known advocate for peace.



THE END