

# Lecture 11: Hashtables

## Wholeness of the Lesson

Hashtables provide even faster access to elements in a collection than a BST, but at the price of losing the sorted status of the elements. If maintaining order among the elements of a collection is not required, hashtables are the most efficient data structure for storing elements in memory, when insertion, deletion, and lookup operations are needed. Hashtables give concrete expression to the ability of pure intelligence to know any one thing instantaneously.

# Hashtable and HashMap ADT

• A Hashtable is an array in which any object can be used as the key instead of just using only integers. The key is used to look up the corresponding *values*.

Note: A typical example is to store records from a database in memory. A key field from the database is often used as a key in the hashtable, and the corresponding record is the value in the hashtable.

The Hashtable class serves the same purpose as the HashMap class with some slight differences. It has essentially the same interface.

Basic operations:

```
void put(Object key, Object value);
Object get(Object key)
remove(Object key)
```

## Example: Character Table

### User's view

Char key	String value		
'a'	"Adam"		
'b'	"Bob"		
'c'	"Charlie"		
'W'	"William"		

```
//insert into table
table.put('c', "Charlie");
//retrieve from table
table.get('c'); //returns "Charlie"
```

### *Implementation:*

#### put(c, s):

- c → (int) c = 99 [convert key to a number -- a hash code] → i = 99 % 4 = 3 [convert hash code to an array index -- a hash value]
- insert new Entry(c, s) into table[i]

### get(c):

- c → (int) c = 99 [convert key to a hash code] → i = 99 % 4 = 3 [convert hash code to hash value]
- Entry e = table[i]; return e.value

#### Pattern:

KEY (non-number) → HASHCODE (number) → HASH VALUE (array index)



### The Structure of Hashtable

The hashtable is like an array. It will have a fixed size when constructed.

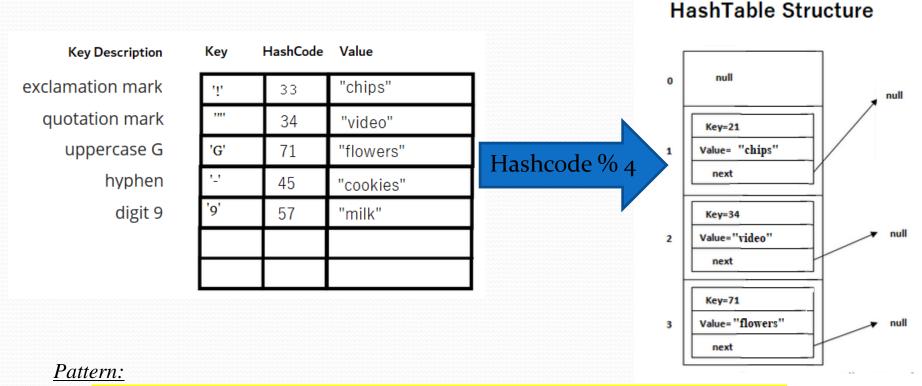
```
public class MyHashtable {
      private static final int INITIAL SIZE = 20;
      private int tableSize;
      private LinkedList[] table;
      public MyHashtable() {
             this (INITIAL SIZE);
      public MyHashtable(int tableSize) {
             this.tableSize = tableSize;
             table = new LinkedList[tableSize];
```

Converting the object into a HashCode integer usually results in a key that is *larger than the size of the HashTable array*. We need to convert that HashCode to an integer index that corresponds to an index in the underlying structure of our HashTable array.

% tablesize will always convert that number into an index that is within the range of the table.

## Second Example of Data into HashTable

The Key and Value goes into a HashTable. Hashcode is calculated from the Key. Same as previous example, the Hashcode is the decimal ASCII code for the Key. To get hashvalue, which is the array index in the table, use % tablesize on hashcode.



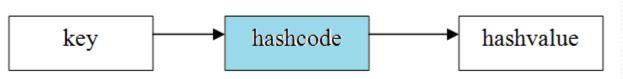
KEY (non-number) → HASHCODE (number) → HASH VALUE (array index)

# Summary: Convert Hashtable Key to Array Index

### • First Steps:

- a. Convert Keys to Integers (hash codes). Devise a way of converting keys to integers so that different keys are mapped to different integers. This is what Java's hashCode() function (in Object) does.
- b. Convert Hash Codes to Indices in an Array (hash values). Devise a way of converting hashcodes to smaller integers (called hash values) that will be the indices of a smaller array, called the table. To do this, you first decide the tableSize, which is the length of the array. A typical way (note: Java does it differently) of making hash value from hashcode is by the formula

hashvalue = Math.abs(hashcode % tableSize)



## Collision in HashTable

The Key of '!' has a Hashcode of 33, which converts to a hashvalue of 1. The Key of '-' has a Hashcode of 45, which also converts to a hashvalue of 1. But there is already an entry in slot 1 of the hashtable array. How is this handled?

The second entry is added onto the end of the linked list in the first slot or element of the hashtable array. The same process happens with the Entry that has

a key of '9'.

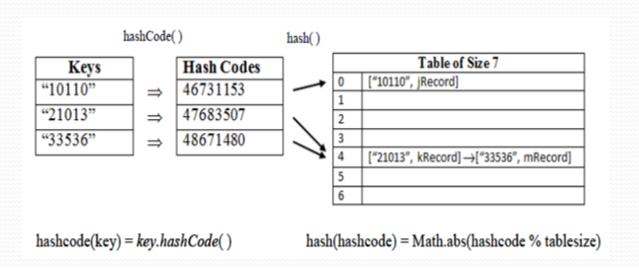
Key	HashCode	Value		ſ		
,i,	33	"chips"		0	null	Key: 45 Null
	34	"video"		,	Key=21 Value= "chips"	Value="cookies"  next  Value=:"milk"  next
'G'	71	"flowers"	Hashcode % 4	1	next next	
9	45	"cookies"		Ì	Key=34	,
'9'	57	"milk"		2	Value="video"	null
				ŀ	Key=71	
				3	Value="flowers"	null
				l	IICAL .	

HashTable Structure

Pattern:

## HashTable Keys

The keys in your hashtable can be of type String, as you see below, or Object. There is a hashCode function in the String class will convert the string key into a good hashcode, a large integer number.



### More on Collision in Hashtable

hashcode(key) = key.hashCode()

put operation: put(k, val)

```
Step 1: Create Entry object
Entry e = new Entry(k, val)
and compute index in table where data will be placed
```

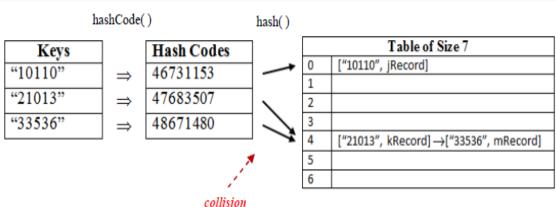
- a) hashcode = hashCode(k)

Step 2: Place the Entry object in the table in slot h.

- Would like to store e in table[h] but cannot because of potential collisions.
- Solution: Each table slot stores a LinkedList.

Store e by:

table[h].add(e)



put("10110", jRecord);

put ("21013", kRecord);

hash(hashcode) = Math.abs(hashcode % tablesize)

put ("33536," 'mRecord);

### get operation get(k)

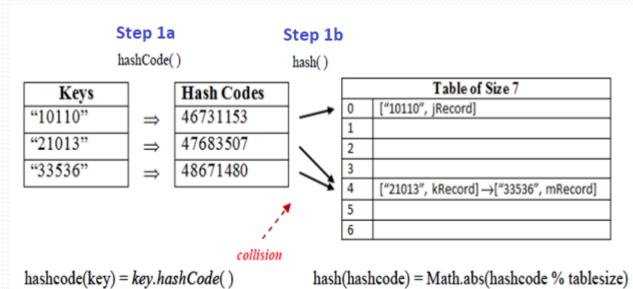
### Step 1: Compute index in table where data will be found

- a) hashcode = hashcode(k)

### Step 2: Read the Entry object from the table and return its value.

LinkedList l = table[h]
Entry e = l.find(k)

return e.val



```
// FIRST Version of Hashtable (SECOND TRY has improvement)
public void put(Object key, Object value) {
      //disallow null keys
       if(key==null) return;
      //convert key to an integer - assumes key not null
       int hashcode = key.hashCode();
      //convert hashcode to a hashvalue - an array index
       int h = hash(hashcode);
       //put the value and the key into an Entry object
       //which will be placed in the table in the correct
       //spot (h) - note: null value is allowed
       Entry e = new Entry(key, value);
      // now place it in the table
      if(table[h] == null) {
             table[h] = new LinkedList();
      table[h].add(e);
```

<u>Problem</u>: Handling Duplicate Keys. Our implementation does not handle duplicate keys in a reliable way.

Example: If the hash<u>value</u> for both key and key2 in the example below is 5, then both entries [key, "Bob"] and [key2, "Dave"] will be put in the list in slot 5:

```
put(key, "Bob")
put(key2, "Dave")
```

Why would this lead to unpredictable results when a get (key) operation is performed ??

### // SECOND\_VERSION

```
public void put(Object key, Object value) {
    //disallow null keys
    if(key==null) return;
    //convert key to an integer - assumes key not null
    int hashcode = key.hashCode();
    //convert hashcode to a hashvalue - an array index
    int h = hash(hashcode);
    //create the entry
    Entry e = new Entry(key, value);
    boolean keyAlreadyInUse = false;
    if(table[h] != null) {
      for(Object ob : table[h]) {
         Entry ent = (Entry) ob;
         if (ent.key.equals(key))
            kevAlreadyInUse = true;
            ent.value = value; //update value for this Entry
    //we handled case keyAlreadyInUse==true in loop
    if(!keyAlreadyInUse) {
        // now place it in the table
        if(table[h] == null){
            table[h] = new LinkedList();
        table[h].add(e);
```

### How are Duplicates Handled?

From the code in the previous slide, you can see that:

If the key of the Entry being put into the hashtable is the same as the key of an existing entry, then the value of the existing entry is updated.

The get operation also supports this.

```
public Object get(Object key) {
     //null key not allowed
     if (key==null) return null;
     //convert key to an integer
     int hashcode = key.hashCode();
     //convert hashcode to a hashvalue - an array index
     int hash = hash(hashcode);
     //if slot given by hash not yet in use, return null
     if(table[hash] == null) return null;
     //now look for the desired Entry
     Entry e = null;
     Iterator it = table[hash].iterator();
     while(it.hasNext()) {
        e = (Entry)it.next();
        if(e.key.equals(key)) {
          return e.value;
     return null;
```

# Overriding the hashCode() Method

- Any implementation of the Hashtable ADT in Java will make use of the hashCode() function as the first step in producing a hash value (or table index) for an object that is being used as a key.
- Default implementation of hashCode() provided in the Object class is not generally useful—gives a numeric representation of the memory location of an object. See package lesson11.defaulthashcode
- Example: We wish to use pairs (firstName, lastName) as keys for Person objects in a hashtable.
   (See package lesson11.needoverridehashcode for demo that has MyHashtable and HashMap)

Demo shows default hashCode method is not useful. If two Pair objects, created at different times, are equal (using the equals method), we would expect them to have the same hashCodes, so that after hashing they are sent to the same table slot. But default hashCode method does not take into account the fields used by equals method, so equal Pair objects are almost always assigned different slots in the table.

<u>First Rule About Creating Keys for a Hashtable</u>: To use an object as a key in hashtable, you must override equals() and hashCode()

# Creating Good Hash Codes When Overriding hashCode()

- There are two general rules for creating hash codes (i.e. for overriding hashCode()):
  - I. (Primary Hashing Rule) Equal keys must be given the same hash code (otherwise, the same key will occupy different slots in the table)

```
If k1.equals(k2) then k1.hashCode() == k2.hashCode()
```

II. (Secondary Hashing Guideline) Different keys should be given different hash codes (if not, then many different keys may be sent to the same slot in the table, thereby degrading hashtable performance).

Best Practice: The hash codes should be distributed as evenly as possible (this means that one integer occurs as a hash code approximately just as frequently as any other)

# Exercise 11.1 – What Happens When hashCode Is Not Overridden Carefully?

- In the InClassExercises project, a MyHashtable class is provided, and a class MyClass is also provided.
- Try overriding hashCode in MyClass as follows:

```
public int hashCode() {
   return 2;
}
```

- In the main method, create 10 instances of MyClass, passing into the constructor values 1, 2, ..., 10 in succession
- Put these 10 instances into a new instance of MyHashtable, using each MyClass instance as a key and its integer value as the value:

```
myHashtable.put(myClass1, 1), etc
```

- Use the printTable method provided in this version of MyHashtable to print the underlying table, including the indices.
- What does the printTable output tell you about this way of overriding hashCode in MyClass?

### Exercise 11.1 – Code

```
public static void main(String[] args) {
  MyClass[] cl = new MyClass[10];
  MyHashtable hashtable = new MyHashtable();
  for (int i = 0; i<10; i++) {
      cl[i] = new MyClass(i);
      hashtable.put(cl[i], i);
   }
                                                public int hashCode() {
  hashtable.printTable();
                                                       return 2;
  public void put(Object key, Object value){
     if(key==null) return;
     //Original: int hashcode = key.hashCode();
     //Exercise 11.1 modification, where hashcode is always 2:
     int hashcode = hashCode();
     int hash = hash(hashcode);
     Etc...
```

### Exercise 11.1 - Solution

### Result "original", when not overridden:

```
o -> [MyClass object #8:8]
1 -> [MyClass object #4:4, MyClass object #6:6]
2 -> null
3 -> null
4 -> [MyClass object #1:1, MyClass object #9:9]
5 -> null
6 -> null
7 -> null
8 -> null
9 -> null
10 -> [MyClass object #0:0, MyClass object #2:2]
11 -> null
12 -> null
13 -> null
14 -> null
15 -> [MyClass object #5:5]
16 -> null
17 -> null
18 -> [MyClass object #3:3]
19 -> [MyClass object #7:7]
```

### Result when **hashcode is always 2**:

```
1 -> null
2 -> [MyClass object #0:0, MyClass object #1:1,
MyClass object #2:2, MyClass object #3:3, MyClass object #4:4, MyClass object #5:5, MyClass object
#6:6, MyClass object #7:7, MyClass object #8:8,
MyClass object #9:9
3 -> null
4 -> null
5 -> null
6 -> null
7 -> null
8 -> null
9 -> null
10 -> null
11 -> null
12 -> null
13 -> null
14 -> null
15 -> null
16 -> null
```

17 -> null 18 -> null 19 -> null

o -> null

# Creating Hash Codes from Object Data (Legacy Approach)

You are trying to define hashcode() for your class and you want to build the hash code from the hash codes of each variable in the class. First step is to assign hash code values to each of these instance variables.

The first step of Legacy approach: Assign hash code values to each of these instance variables.

If f is an instance variable in your class, then:

- If f is boolean, compute (f ? 1 : 0)
- If f is a byte, char, short, or int, compute (int) f.
- If f is a long, compute (int) (f ^ (f >>> 32))
- If f is a float, compute Float.floatToIntBits(f)
- If f is a double, compute Double.doubleToLongBits(f) which produces a V f1, then return (int) (f1 ^ (f1 >>> 32))
- If f is an object, use f.hashCode() (best if implementation of f has overridden hashCode() already)

Note: This is a list of the ways that the hashcode can be calculated for instance variables.

# Combining HashCodes of Instance Variables to Produce a Final HashCode (Legacy Approach)

• Suppose your class has instance variables u, v, w and corresponding hashCodes hash\_u, hash\_v, hash\_w (obtained as in the previous slide). Then:

```
@Override
public int hashCode() {
  int result = 17;
  result += 31 * result + hash_u;
  result += 31 * result + hash_v;
  result += 31 * result + hash_w;
  return result;
}
```

# Combining HashCodes of Instance Variables to Produce a Final HashCode (Modern Approach)

• Use the following method in the Objects class to compute hash code.

```
public static int hash(Object... values)
```

• For example, if an object has three fields, x, y, and z, we could compute hash code in this way:

```
@Override
public int hashCode() {
  return Objects.hash(x, y, z);
}
```

# Overriding HashCode for Pair class

Example. Overriding hashCode in the Person-Pair example. We must take in account the same fields in computing hashCode as those used in overriding equals. The fields in Pair are Strings, and Java already provides hashCodes for Strings. So we make use of these and combine them to produce a complex hashCode for Pair. (See the solution in lesson11.needoverridehashcode)

```
public class Pair {
    String first;
    String second;
    public Pair(String f, String s) {
        first = f;
        second = s;
    }
}
```

```
legacy approach
 public int hashCode()
     ine result = 17
     int hashFirst = first.hashCode();
     int hashSccood = second.hashCode();
     result 4= 31 * result + hashFirst;
     result += 31 * result + hashSecond;
     eturn result:
//modern way
public int hashCode() {
    return Objects.hash(first, second);
```

# HashCode in Java's String Class

Reminder: To use an object as a key in hashtable, you must override equals() and hashCode(). The Java String class is the only exception to this.

The Java String class has a hashCode() method that overrides the Object hashCode(). There is no need to improve upon this hashcode method in the String class.

Any Java String is converted to an integer via hashCode () by this formula:

```
Given a String s of length k+1 s.hashCode()equals s.charAt(0) * 31^{k-0} + s.charAt(1) * 31^{k-1} + s.charAt(2) * 31^{k-2} + . . . + s.charAt(k) * 31^{k-k}
```

Since every character in the String is taken into account, equal Strings must have equal hashCodes. A prime number, 31, is used in this formula so that it is highly unlikely that two distinct Strings will be assigned the same hashCode (though it's possible).

# The Java library class HashTable Requirements

From the Java JDK 10:

The java library class HashTable maps keys to values. To successfully store and retrieve objects from a hashtable, the objects used as keys must implement the hashCode method and the equals method.

Note: The Java String class implements the hashCode method and the equals method. If your key is a String, then there is no need to add support for a hashCode or equals method.

# Java's Implementation of Hashtables

 HashMap and Hashtable (HashMap is preferred; Hashtable is "legacy")

Feature	java.util.HashMap	java.util.Hashtable	MyHashtable	
Allows null key	Yes	No	No	
Allows null values	Yes	No	Yes	
Allows duplicate keys	No	No	No (after correction)	
Synchronized (for safe multithreading)	No	Yes	No	

# Java's Implementation of Hashtables

• Pre-j2se5.0

```
HashMap map = new HashMap();
map.put("Bob", new Employee("Bob", 40000, 1996, 10, 2));
Employee emp = (Employee) map.get("Bob");
```

j2se5.o version is parametrized:

## **Main Point**

The most common implementation of hashtables uses separate chaining; each hashvalue is an index in an array of Lists; all objects with colliding hashvalue i are stored in the *i*th list in the array. The solution to the problem of avoiding hashvalue collisions (namely, by using a List in each of the array slots) illustrates the principle of the second element: Using a List in each array slot provides a way to "harmonize" objects that were apparently in conflict because of identical hashvalues.

# Hashtable Application #1: Removing Duplicates

- The most common use of hashtables is as in-memory look-up tables. For example, Employee records from a database could be stored by using Employee ID as key and the entire Employee record as value.
- Another application of hashtables is for "bookkeeping" purposes. A simple example is an efficient procedure for removing duplicates from a list.

A less efficient way to remove duplicates is to use nested loops: For each element e in the list, use an inner loop to look at all elements preceding e in the list to see if e has occurred before; if so, remove this second occurrence of e.

A more efficient approach is to do the following: Create an auxiliary hashtable h. For each e in the list, check to see if e is a key in h. If so, remove e from the list. If not, add the entry <e, e> to h.

• See package lesson11.removedups

## Exercise 11.2

Use a hashtable as a fast way to solve the following problem:

You are given two arrays arr1, arr2, of Strings Determine whether every String contained in arr1 is also contained in arr2. You may not use nested loops.

See the method isSubArray in the Subarray class, in the InClassExercises.

## Exercise 11.2

```
public class Subarray {
  public static void main(String[] args) {
         String[] arr1 = {"A", "C", "E"};
         String[] arr2 = {\text{"A", "B", "C", "D", "E"}};
         String[] arr<sub>3</sub> = {"A", "C", "F"};
         System.out.println("Is arrı a subarray of arr2? (Expected answer: yes) " +
                  isSubArray(arr1, arr2));
         System.out.println("Is arr3 a subarray of arr2? (Expected answer: no) " +
                  isSubArray(arr3, arr2));
  /** Returns true if every string in arr1 is in arr2. Assume neither array is null */
  public static boolean isSubArray(String[] arr1, String[] arr2) {
         //implement
         return false;
```

## Solution

```
public static boolean isSubArray(String[] arr1, String[] arr2) {
    var h = new HashMap<String, String>();
    for(String s : arr2) {
        h.put(s, s);
    }
    for(String s: arr1) {
        if(!h.containsKey(s)) return false;
    }
    return true;
}
```

# Hashtable Application #2: The Set ADT

Mathematically, a set is (roughly) a collection of objects.
 Two sets are said to be equal if they have the same elements.

For example:

$$\{1, 1, 3\} = \{1, 3\} = \{3, 1\}$$

because all have the same elements.

A set does not impose an ordering of elements (the set may contain elements that have their own natural order, like integers, but the set itself does not impose an order)

- 2. We can represent the mathematical notion of a set as an ADT called Set. In order to faithfully represent the properties of the mathematical idea, the Set ADT must have the following characteristics:
  - It does not allow duplicate elements
  - Iterating through a set does not guarantee any special order on its elements (in particular, there is no guarantee that the order in which an Iterator will provide elements corresponds to the order in which the elements were added in the first place)
  - Its overridden equals () method declares two Sets to be equal if and only if they have the same elements

Java provides a Set interface (and in j2se5.0, the parametrized version Set<E>) as part of the Collections API.

# The Set Interface for Strings

```
public interface Set<String> {
   //adds a String to the Set
   //do not allow null and prevent duplicates
   public boolean add(String ob);
   //removes a String
   public boolean remove(String ob);
   //returns true if ob is an element
   public boolean contains(String ob);
   //returns true if the set has no element
   public boolean isEmpty();
   //returns the number of elements in the set
   public int size();
```

### Two implementations of the Set interface in Java:

- HashSet Using a HashMap to implement Sets
  prevents duplicate elements from being added in the
  Set. Does not guarantee order in which elements are
  stored
- TreeSet uses a red-black tree (a very efficient BST) to store elements, so when they are output, they are in sorted order
- Demo: lesson11.sets

## Sample Implementation of Set

```
public class MySetImpl implements MySet {
     private MyHashtable table;
     public MySetImpl(){
           table = new MyHashtable();
     public boolean add(Object ob){
           if(ob == null | isUnmodifiable) {
                return false;
           if(table.containsKey(ob)){
                return false;
           table.put(ob,ob);
           return true;
     public boolean contains(Object ob){
           if(ob == null) return false;
           return table.containsKey(ob);
     public boolean remove (Object ob){
           if(ob == null || isUnmodifiable) {
                return false;
           if(table.containsKey(ob)){
                Object obj = table.remove(ob);
                return (obj != null);
           return false;
```

## Main Point

The Hashtable ADT is a generalization of the concept of an array. It supports (nearly) random access of table elements by looking up with a (possibly) non-integer key. In the usual implementations, objects used as keys in a hashtable are "hashed", producing hashcode (a numeric value) and hashvalue (numeric value reduced in size to be less than the table size). The hashvalue is an index in an array that can be used to locate or insert an object. Hashtables illustrate the principle of Do less and accomplish more - they provide an incredibly fast implementation of the main List operations.

# Guidelines for Use of Common Data Structures

### 1. Array List

- <u>Use When</u>: Main need for a list is random access reads, relatively infrequent adds (beyond initial capacity) and/or number of list elements is known in advance. Sorting routines run faster on an ArrayList than on a Linked List.
- Avoid When: Many inserts and removes will be needed and/or when many adds expected, but number of elements unpredictable. Also: Maintaining data in sorted order is very inefficient.

### 2. Linked List

- <u>Use When</u>: Insertions and deletions are frequent, and/or many elements need to be added, but total number is unknown in advance. There is no faster data structure for repeatedly adding new elements than a Linked List (since elements are always added to the front).
- <u>Avoid When</u>: There is a need for repeated access to ith element as in binary search random access is not supported.

### 3. Binary Search Tree

- <u>Use When</u>: Data needs to be maintained in sorted order. Faster than Linked Lists for insertions and deletions, but ordinary adds are slower. Provides very fast search for keys.
- Avoid When: The extra benefit of keeping data in sorted order is not needed and rapid read access is needed (Array List provides faster read access by index and hashtables provide faster read access by key)

### 4. Hashtable/Hashmap

- <u>Use When</u>: Random access to objects is needed but array indexing is not practical. Provides fastest possible insertion and deletion (faster than BST's).
- Avoid When: The order of data must be preserved (example: you want to find all employees whose salaries are in the range 60000..65000) or "find Max" or "find Min" operations are needed.

### 5. Sets

• <u>Use When</u>: Duplicates should be disallowed, and there is no need for rapid lookup of individual set elements. Example: keySet() in HashMap returns a Set. To order the elements, use TreeSet.

47

# Connecting the Parts of Knowledge With the Wholeness of Knowledge

Random access expanded from integer index to arbitrary index, from "point" to "infinity"

- 1. Arrays and ArrayLists provide highly efficient index-based access to a collection of elements.
- The Hashtable ADT generalizes the behavior of an array by allowing noninteger keys (in fact, any object type can be used for a key), while retaining nearly random access efficiency for insertions, deletions, and lookups.
- **Transcendental Consciousness:** TC is the home of all knowledge. The Upanishads declare "Know that by which all else is known" this is the field of pure consciousness.
- 4. Wholeness moving within itself: In Unity Consciousness, one sees that the "key" to accessing complete knowledge of any object is the infinite value of that object, pure consciousness, which is known in this state to be one's own Self. Knowing that level of the object, it then becomes possible to know any more relative level of the object as well.