Large Scale Data Processing
CSE3025

Dr. Ramesh Ragala

School of Computer Science and Engineering
VIT Chennai

February 12, 2021

- Hadoop framework is based on java $\rightarrow$ it is written in Java $\rightarrow$ Understanding of java is helpful
- The minimal java concepts for Hadoop
  - ▶ Object Oriented Programming Systems (OOPS) and interfaces
  - ▶ Arrays, Linear and Non-linear Data structures
  - ▶ Collections
  - ▶ Exception Handling

- **Arrays in Java**
  - ▶ Data Structure is a way to organize the data in a computer so that it can be used efficiently for storing, modifying, accessing and deleting data.
  - ▶ Data Structures are classified into two types linear Data Structures and non-linear Data Structures. Example: Arrays, List, Stacks
  - ▶ **Linear Data Structures:** Data elements are attached one after the other in a linear fashion and any retrieval and manipulation follows the same order.
  - ▶ **Non-linear Data Structures:** One data element is connected to another data element through a relationship and all elements in a non-linear data structure cannot be traversed in a single run. Example: Graph, Trees

- **Arrays in Java**
  - ▶ Arrays hold sequences of values of the same type
  - ▶ Generally, Array is a data structure that stores a collection of values of the same type.
  - ▶ The individual elements can be accessed using index.
  - ▶ Declaring an Arrays specifies the array type followed by [] and the array variable name

Declarion of Array

dataType[] arrayName;

- ▶ **Example on declaration of Array:** int[] arr;
- ▶ The declaration of the array not yet initialize the array with actual array values. It can be done with the help of **new** operator. This is called Array initilization.

Initialization of Array

ArrayVariableName = new DataType [ArraySize];

- ▶ **Example on Initilization of Array:** arr = new int[10]; $\rightarrow$ it declares and initilizes an array of 10 integers
- ▶ Once the array creation completes, then the length of the array can not be changed. $\rightarrow$ array lists data structure is introduced to eliminate this constraint.

- Java is an object-oriented programming (OOP) language
- The object-oriented programming is about creating objects that contain both data and methods.
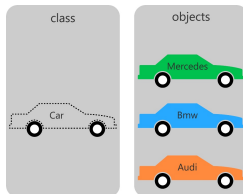- Classes and Objects are the two core aspects of OOP.

Figure: class and object exampel for car

Java Documentation

In the real world, you'll often find many individual objects all of the same kind. There may be thousands of other bicycles in existence, all of the same make and model. Each bicycle was built from the same set of blueprints and therefore contains the same components. In object-oriented terms, we say that your bicycle is an instance of the class of objects known as bicycles. A class is the blueprint from which individual objects are created

- Class is a blueprint that defines the behavior and property of an object $\rightarrow$ it is the template for an object.
- Every class has a state and behavior.
- Example: Bicycles have state like (current gear, current speed, current pedal cadence, etc.) and behavior like (changing pedal, changing gear, applying brakes, etc.)

- **Object:**
  - ▶ It is an instance or copy of the blueprint known as Class
  - ▶ For programmers to use the properties and the behavior of a particular class, it is necessary to create an object
  - ▶ An object of a class has all the attributes bundled in it
  - ▶ Identifying the state and behavior for an object is crucial when designing a class for a real world application

Syntax of Class

class ClassName {
fields
methods
}

- Here, fields (variables) and methods represent the state and behavior of the object respectively.
- fields are used to store data
- methods are used to perform some operations

Syntax for object creation

ClassName object = new ClassName()

- **new** keyword is used to create an object (sets the memory)
- **.** operator is used to access the members of a class through object name

- **Java Constructor**
  - ▶ Java Constructor is a special method that is used to initialize objects.
  - ▶ The constructor is called when an object of a class is created.
  - ▶ It can be used to set initial values for object attributes.
  - ▶ The memory for the object is allocated in the memory during the call of constructor
  - ▶ Every time an object is created using the new() keyword, at least one constructor is called.
  - ▶ Constructor name must be the same as its class name
  - ▶ A Constructor must have no explicit return type