

Scala Tutorial I

Jan Macháček

June 9, 2018

Abstract

Scala is a fusion language that combines functional and object-oriented programming paradigms in a syntax that is similar to most other *C-like* languages. The ...

1 Syntax crash-course

Scala's syntax follows the syntax of other C-like languages, though—like Pascal—the type specification follows an identifier. Scala's **class** behaves exactly like Java's **class**, and its syntax is not wildly different. (See Listing 1.)

Class declaration means the same thing as Java; constructor parameters are specified in the block immediately following the class name.

Note that the types follow the identifier; instead of `String constructorParam1` Scala uses `constructorParam1: String`

```
class MyClass(constructorParam1: String, constructorParam2: Int) {
```

Methods begin with the keyword **def**, followed by name and parameters. The return type follows similar pattern; `Unit` means `void`.
The body of the method follows the equals sign.

```
  def execute(methodParam1: List[Int]): Unit = {
```

```
  }
```

```
}
```

Constructing instances uses the typical **new** keyword...

```
new MyClass("foo", 42)
```

Method invocation is exactly like Java's

```
.execute(List(1, 2, 3))
```

Listing 1: Classes and methods

Most of this syntax is familiar and unsurprising; the only thing that might feel odd is the square bracket for “generics” in `List[Int]`: in Java, this would be written as `List<Int>`. This is part of Scala's legacy. A long, long time ago, XML was very exciting; and Scala allows XML literals. These XML literals use the angle brackets. This meant that a different symbol had to be used for type parameters. Because the square bracket is used for type parameters, array indexing is also done using regular parentheses¹.

Interfaces use the **trait** keyword in Scala; their usage and features are similar to **interface** in Java (particularly Java 8 which adds default implementations). It is possible to make anonymous implementations of a **trait**, as well as to implement it in ordinary **class**es. (Viz Listing 2.)

Apart from the **trait** keyword, the syntax is unremarkable

```
trait ReportGenerator {
```

Interface methods are public and abstract; they specify parameters and return type

```
  def generate(userId: Int): Array[Byte]
```

```
}
```

A **trait** can be implemented in a **class** using the keyword **extends**. Additional traits to be implemented use the **with** keyword.

```
class ReportGeneratorImpl extends ReportGenerator with Cloneable {
```

Instead of the `@Override` annotation Scala uses the **override** keyword.

```
  override def generate(userId: Int): Array[Byte] = ...
```

```
  override def clone(): AnyRef = ...
```

```
}
```

It is also possible to make an anonymous implementation of a trait using the **new** keyword.

```
new ReportGenerator {
```

```
  override def generate(userId: Int): Array[Byte] = ...
```

```
}
```

Listing 2: Traits

¹I know, it looks like VisualBasic or Fortran (formerly FORTRAN)!

It is worth noting that there is no special syntax for *array of X* in Scala. Instead, it uses `Array` with the specified type parameter. (So, Java's `byte[]` becomes `Array[Byte]`, `User[]` becomes `Array[User]`, and so on.) Also notice the `AnyRef` in the implementation of the `clone()` method—it is equivalent to `java.lang.Object`.

So far, there are no major surprises: classes and interfaces work just like most other languages, constructor, method, and parameter definition also looks fairly ordinary. The syntax for functions (finally!) uses parameters (each with its type following the `:` symbol, if needed) followed by fat arrow \Rightarrow , and the body of the function. Usage is the same as Java 8; and Scala's collections library contains concepts that are fairly similar.

Double every `Int` in the list

```
List(1, 2, 3, ...).map { (x: Int) => 2 * x }
```

Scala can infer the type of `x`, so there's no need to specify it

```
List(1, 2, 3, ...).map { x => 2 * x }
```

An alternative syntax uses parentheses in place of brackets

```
List(1, 2, 3, ...).map(x => 2 * x)
```

It is possible to avoid having to declare the parameter `x`, and use `_` instead. Scala compiler replaces the every occurrence of `_` with a fresh parameter declaration.

```
List(1, 2, 3, ...).map(2 * _)
```

Listing 3: Fields & variables

Finally, fields (and variables) use the keywords `val` and `var`. The first declares an immutable variable (and a getter if field), the second declares a mutable variable (and a getter and setter if field); see Listing 4.

Field definitions have to specify initial value; use `_` for default value.

```
class User {
  var id: Long = _
  var name: String = _
  var dob: Date = _
}
```

```
val user = new User
```

Listing 4: Fields & variables

This code is terrible! The default value for reference types (`AnyRefs`) is `null`, and variants of `zero` for primitive types. The `name` and `dob` fields in the `user` instance are `null`, and `id` is `0L`. What's worse, the class has setters for these fields, and it's possible to invoke them. The syntax is somewhat nicer—it looks like plain assignment using the `=` operator, though it is actually invoking a setter—but the semantics of the code in Listing 5 is terrible.

Field definitions have to specify initial value; use `_` for default value.

```
class User {
  var id: Long = _
  var name: String = _
  var dob: Date = _
}
```

```
val user = new User
user.id = 5
user.name = "Foo_Quux"
user.dob = ...
```

Listing 5: Fields & variables II

While syntactically valid code, it is very confusing. The `user` variable is declared immutable, yet it is possible to invoke its setters. The equivalent Java code would declare the `user` variable `final`, but then still use the setters to mutate it.

1.1 Killer features

The syntax (and its application) so far looks just like Java with less typing. There must be something else that makes it worth leaving the creature comforts of Java.

- everything is an expression
- case classes
- pattern matching
- for comprehensions
- implicits
- rich type system

Everything is an expression. In Java, C, and similar languages, there are statements and expressions. Statements do not have value, they are typically control-flow constructs. For example, in Java, `if (a == b) X else Y` cannot be assigned to a variable, because *if-then-else* is a statement. In Scala, everything yields a value that can be assigned to a variable.

Case classes. A case class is just like a class in that it is a container for data and methods, but the fields it contains only have getters. (Immutability only goes as far as immutability of the references. Even in Scala, immutability without any additional code is equivalent to using *final* in Java.) Nevertheless, case classes are fantastically convenient to define data structures. Consider the *Person* case class defined in Listing 6.

```
case class Person(firstName: String, lastName: String, age: Int)
```

Listing 6: Case class *Person*

This is all it takes to define an immutable structure (with the T&Cs from above) with the fields *firstName*, *lastName*, and *age*; but also with appropriate *toString*, *hashCode*, and *equals* implementations. These automatically generated implementations delegate to the *toString*, *hashCode*, and *equals* methods of all the fields, in the order in which they are specified.

To create an instance of a case class, do not use the *new* keyword; instead, write the parameters directly after the case class name, as shown in Listing 7.

Notice that there is no *new* keyword; the parameter values are applied directly after the case class name.

```
val fq = Person("Foo", "Quux", 42)
```

To access the fields, use the familiar *.* notation.

```
fq.firstName | "Foo"
```

Invoking the *toString* method prints a reasonable representation of the case class.

```
fq.toString | "Person(Foo,Quux,42)"
```

It is possible to vary the order of the parameters if the parameter names are also specified. This can help readability.

```
val fb = Person(lastName = "Baz", firstName = "Foo", age = 50)
```

```
fb.toString | "Person(Foo,Baz,50)"
```

Equality is implemented by delegation to the parameters' *hashCode* and *equals* methods.

```
Person("Foo", "Quux", 42) == fq | true, even though they are different instances.
```

```
fq == Person("Foo", "Quux", 42) | true, even though they are different instances.
```

Listing 7: Using case class *Person*

The conciseness of Scala's syntax is beginning to show. It would have been much more cumbersome to implement all this (including correct *hashCode*, *equals*, and reasonable *toString*) in Java. Even with correct implementations, it would not have been possible to use *==* to test for instance equality.

Exercise: Using case classes and functions

In *src/scala*, run *sbt console*; in the console, define the *Person* case class from Listing 6. Then use *List.fill(N)(Person(...))* and useful methods in *scala.util.Random* to generate *N* random *Person* instances. Assign those to a variable.

- find the oldest person in the list. (Hint: use *map*, *max*, and *find* functions on the *List[Person]*.)
- find the 10 youngest people in the list. (Hint: use *sortBy* or *sortWith*, followed by *take*.)

Case classes are also called *product* types. The word product refers to the possible number of values that a case class can hold. Take *case class ABC(a: Boolean, b: Boolean, c: Boolean)*: there are eight possible values: *ABC(false, false, false)*, ..., *ABC(true, true, true)*. The value 8 is the result of *multiplying* the possible values of all parameters. *Boolean*s have two values; three *Boolean* values yield $2 \times 2 \times 2$ possible values of the *ABC* type. Similarly, *case class IB(i: Int, b: Boolean)* has 4294967295×2 possible values.

Imagine, for a moment, that *Boolean* is not a built-in primitive. It would be defined as sum type: *sealed trait Boolean; case object True extends Boolean; case object False extends Boolean*. The case instances *True* and *False* have exactly one value; consequently, the sum type *Boolean* has $1 + 1$ possible values.

Sum types are a very useful way to express alternatives that might not have anything in common other than being of the given type, without clumsy *instanceof* checking with 1.1.

Pattern matching. A pattern match is a way to check that a value has the right “shape”, and to pull out some or all values from that shape. Think of the simplest pattern matching expression as Java's *switch* statement in Listing 8.

This is the equivalent of Java's `switch` statement, with multi-case and a default case.

```
Random.nextInt(10) match {
  Matches values 0..3
  case 0|1|2|3 =>
  Matches only value 6
  case 6 =>
  Matches all other values
  case _ =>
}
```

Listing 8: Pattern matching

It is possible to match on much more complicated structures; such as tuples. A tuple is a collection of a specific length with elements of arbitrary types. For example $(1, \text{"foo"})$ is a tuple of 2 elements of types `Int` and `String`; $(\text{"foo"}, \text{"bar"}, 1.3)$ is a tuple of 3 elements of types `String`, `String`, and `Double`. It is possible to pattern match on those, as shown in Listing 9

```
import scala.util.Random
(Random.nextString(Random.nextInt(16)), Random.nextInt(100)) match {
  Matches tuples with any first element, and second element == 10.
  case (s, 10) => The first element is accessible as s here.

  Matches only the ("Lucky?", 42) tuple.
  case ("Lucky?", 42) => None of the elements are accessible here.

  Matches any tuple as long as the first element is longer than 5 characters and second element is less than 10.
  case (x, y) if x.length > 5 && y < 10 => Both first and second elements are accessible as x and y.

  Matches any value
  case _ => None of the elements are accessible here.
}
```

Listing 9: Pattern matching II

As useful as tuples are for quick ad-hoc structures, the real deal in pattern matching are case classes. Just like matching on a tuple, it is possible to match on case classes and extract their fields as needed. In addition to using the `match` keyword followed by `case` s, pattern matching also applies to declarations of variables in Listing 10.

Pattern matches on the right-hand side, pulls out the first parameter as `val first: String`.

```
val Person(first, _, _) = Person("Foo", "Bar", 99)
```

Pattern matches on the right-hand side, comparing the value of the first parameter to be equal to in-scope variable `first`. (Notice the backticks.)

```
val Person(`first`, _, _) = Person("Foo", "Bar", 99)
```

Because there are no alternatives, if the `val` pattern match fails, it raises an exception.

```
val Person(`first`, _, _) = Person("Fooo", "Bar", 99)
```

Listing 10: Pattern matching III

The pattern match that raises an exception if it fails is often convenient to use in tests. Suppose there

Exercise: Using case classes and pattern matching

Build an numerical expression evaluator; the evaluator should support binary operators `+`, `-`, `*`, `/` and support arbitrary nesting. It should be able to evaluate, for example $4 + 3(5/(12 - 7))$. To make things simpler, it is not necessary to include error reporting (division by zero, for example), and it is not necessary to provide a way to turn a `String` into an expression.

The sum type `Expr`'s subtypes define the "operations" we support.

```
sealed trait Expr
```

The different cases need to support our operations and constants.

```
case class Plus(left: Expr, right: Expr) extends Expr
...
case class Const(const: Int) extends Expr
```

It will be convenient to package our evaluator in its own module.

```
object Evaluator {
```

Hint: you will need to pattern-match on the different `Expr`s.

```
def eval(expr: Expr): Int = expr match {
  ...
}
```

```
}
```

```
object EvaluatorMain extends App {
```

```
val e = Evaluator.eval(Plus(Plus(Const(4), Const(5)), Const(3)))
println(e) // 12
```

```
}
```

Listing 11: Expression evaluator

Exercise: boolean expression simplifier

Build a logical expression simplifier that reduces the number of logical operations to be performed, and can spot tautologies and contradictions. For example, $(a \wedge b) \vee (a \wedge b)$ should simplify to just $a \wedge b$; $(a \vee \neg a)$ should simplify to \top (tautology; always true); $(a \vee \neg a) \wedge (a \oplus a)$ should simplify to \perp (contradiction; always false).

The sum type `LExpr` defines the different logical expressions we support.

```
sealed trait LExpr
```

The different cases need to support our operations and constants.

```
case class LAnd(left: LExpr, right: LExpr) extends LExpr
case object Contradiction extends LExpr
```

```
...
```

```
case class Var(named: String) extends LExpr
```

Just like the evaluator, package the simplifier in its own module.

```
object Simplifier {
```

Hint: you will need to pattern-match on the different `LExpr`s.

```
def simplify(expr: LExpr): LExpr = expr match {
  ...
}
```

```
}
```

```
object SimplifierMain extends App {
```

```
  val se = Simplifier.simplify(
    LAnd(
      LOr(Var("a"), LNot(Var("a"))),
      LXor(Var("a"), Var("a"))
    )
  )
  println(se) // "Contradiction"
```

```
}
```

Listing 12: Logical expression simplifier

2 The standard library

The Scala standard library defines many standard structures in case classes. Picking the two most useful ones, *Option* and *Either* are defined as case classes. This means that it is possible to pattern match on their values.

These are the definition in the standard library; note that the actual definitions are more complex, but this code expresses the essence.

`Option[A]` has two alternatives: `Some[A]` and `None`. It is used to express one-or-missing values.

```
sealed trait Option[A]
case class Some[A](a: A) extends Option[A]
case object None extends Option[Nothing]
```

`Either[L, R]` has also two alternatives: `Left[L]` and `Right[R]`. It is used to express success or failure, with success by convention on the right and failure on the left.

```
sealed trait Either[L, R]
case class Right[R](r: R) extends Either[Nothing, R]
case class Left[L](l: L) extends Either[L, Nothing]
```

Listing 13: Pattern matching in stdlib

...

3 Spring Framework

The Spring Framework is a dependency injection framework; it encourages composition over inheritance, it encourages expressing dependencies as interfaces rather than concrete implementations. The framework takes care of instantiating the components in the correct order; most components (the ones that fall into the *@Component* stereotype) are *singletons*. This means that it is possible to treat the *@Component*-annotated components as namespaces rather than containers of

state². The reason why Spring Framework encourages programming to interfaces is to make the software easily testable: there can be separate implementations or mocks for unit and integration tests.

```
interface ReportGenerator {
    Generates the PDF report for the given user,
    returns the byte array representing the PDF contents
    byte[] generate(final String user);
}

@Component
public class ReportService {
    private final ReportGenerator reportGenerator;

    @Inject
    public ReportService(final ReportGenerator reportGenerator) {
        this.reportGenerator = reportGenerator;
    }

    public void reportAll() {
        for (final String user : Arrays.asList("a", "b", "c")) {
            final byte[] pdf = this.reportGenerator.generate(user);
            Now you're on your own...
        }
    }
}
```

Listing 14: Components

For a Spring Framework application to be able to construct the *ReportService*, it needs to be able to construct exactly one component that implements the *ReportGenerator* interface.

```
@Component
public class JasperReportsReportGenerator implements ReportGenerator {
    public byte[] generate(final String user) {
        ...
    }
}
```

Listing 15: Components

Without a DI framework, the work of constructing the dependencies would fall on the programmers, yielding code similar to Listing 16.

Typically in `public static void main(String[] args)` or in a test:

```
ReportGenerator rg = new JasperReportsReportGenerator();
ReportService rs = new ReportService(rg);
```

Listing 16: Manual DI

Constructing the instances of the *JasperReportsReportGenerator* and *ReportService* using their constructors isn't a problem per se, but with growing number of dependencies this grows to be tedious.

4 Zero to hundred

FizzBuzz is a typical program that follows *Hello, world*, adding iteration and conditions. The Scala version of FizzBuzz is shown in Listing 17—it shows the definition of a function *def*, followed by name and arguments, and its implementation that follows the *=* sign. The loop (*for*) and condition (*if*, *else*) keywords are the old friends from other languages.

```
def fizzBuzz = {
    for (i ← 1 to 100) {
        if (i % 15 == 0) println("FizzBuzz")
        else if (i % 3 == 0) println("Fizz")
        else if (i % 5 == 0) println("Buzz")
        else println(i)
    }
}
```

Listing 17: Fizz Buzz

The FizzBuzz from Listing 17 isn't particularly re-usable: it simply prints 100 elements to the standard output, nothing else and nothing more. There is no way, for example, to direct the output to a web socket, or to use it to determine how it maps of the value in the *integer* domain to the "FizzBuzz domain". Hmm!—*mapping* and *domain* sound like

²In fact if the methods in *@Component*-annotated classes mutates & accesses its fields, it will suffer from race conditions.

mathematics; and functional programming is supposed to be somehow more mathematical. And mathematics is jolly wonderful.

The first step in making the *fizzBuzz* more mathematical is to make it map an input to exactly one useful output. Right now, its return type now is *Unit*, which is a bit like *void* in Java and C; changing its definition to *def fizzBuzz2(max: Int): Unit* (and then using the *max* parameter in the loop) isn't particularly useful: it is a mapping from a number to *Unit*. And, if this were mathematics, there can be only one such mapping: *def fizzBuzz2(max: Int): Unit = ()*. Instead of printing the elements to the console, the implementation needs to return a value that can be printed. A simple *String* would do, but a *Seq* of *Strings* is better. The type becomes *Int* \Rightarrow *Seq[String]*, and the implementation is shown in Listing 18.

```
def fizzBuzz(max: Int): Seq[String] = {
  var result = List.empty[String]
  for (i ← 1 to max) {
    if (i % 15 == 0) result = result :+ "FizzBuzz"
    else if (i % 3 == 0) result = result :+ "Fizz"
    else if (i % 5 == 0) result = result :+ "Buzz"
    else result = result :+ i.toString
  }
  result
}
```

Listing 18: Fizz Buzz

This is a huge improvement! The *fizzBuzz* is now indeed a function: it maps input to output and its result depends only on the value of the parameter. It would even be possible to pre-compute the result for all possible values of the input and replace the function's body with a look-up in that table: the function would become just data!

Well, the outside looks great, but the implementation stinks! It uses mutation, and what about the strange *:+* operator in *result :+ "Fizz"*, never mind the *for (i ← 1 to max) {...}* nonsense!

```
def fizzBuzz(max: Int): Seq[String] = {
  def fb(i: Int): String =
    if (i % 15 == 0) "FizzBuzz"
    else if (i % 3 == 0) "Fizz"
    else if (i % 5 == 0) "Buzz"
    else i.toString
  (1 to max).map(fb)
}
```

Listing 19: Fizz Buzz

In Scala, every concrete type (except *Nothing*) can have a value: for example, the type *Boolean* is inhabited by values *true*, *false*; the type *Int* is inhabited by values such as *5*, *42*, *-100*, *0*, *...*; the type *String* is inhabited by values such as *"Hi"*, *":)"*, *""*; the type *Unit* is inhabited by the only value *()*. (No, really, it's perfectly good Scala syntax to write *()* as value. It's just not particularly useful.) The only type that does not have any inhabitants is *Nothing*: it represents expressions that *diverge*, for example throwing an exception.

Taking a more precise look at *def fizzBuzz* reveals its type to be *Unit*; it evaluates to only one value, namely *()*. If it were a function in the sense of strictly mapping input to output, it would be no different from any other *()* *constant*. But *fizzBuzz* does some additional work before returning *()*; this additional work is not represented by its type, even though it is its *raison d'être*.

In Java and C, there is no *value* of type *void*

As it stands, its type is *()* \Rightarrow *Unit*,

5 Pattern matching

sasd

