**第6章 类文件结构**

Java代码编译的结果：从本地机器码转变为字节码

各种运行在不同平台（不同操作系统和硬件体系结构）的虚拟机都可以载入和执行同一种与平台无关的字节码，从而实现了程序的

“一次编写，到处运行”—— Write Once, Run Anywhere.

各种不同平台的虚拟机与所有平台都统一使用的程序存储格式——**字节码（ByteCode）**是构成平台无关性、语言无关性的基石。

java程序 🡺 javac编译器 🡺 字节码 🡺 java虚拟机执行

\*.java \*.class

一、Class类文件的结构

任何一个Class文件都对应着唯一一个类或接口的定义信息；

但相反，类或接口并不一定得定义在文件里（比如类或接口也可以通过类加载器直接生成，即不一定以磁盘文件的形式存在：\*.class）

Class文件是一组以**8位字节为基础单位**的二进制流，各个数据项目严格按照顺序紧凑地排列在Class文件之中，中间没有添加任何分隔符。因此Class文件存储内容几乎全部是程序运行的必要数据，没有空隙存在。当遇到需要占用**8位字节以上空间的数据项**时，会按照**高位在前**（Big-Endian）的方式**分割成若干个8位字节进行存储**。

注：Big-Endian：用于SPARC、PowerPC等处理器，具体指最高位字节在地址最低位、最低位字节在地址最高位来存储数据。

Little-Endian：用于x86等处理器

Class文件格式采用一种类似于C语言结构体的伪结构来存储数据，这种伪结构中只有两种数据类型：**无符号数**和**表**，后面的解析都要以这两种数据类型为基础。

无符号数：基本的数据类型，以u1、u2、u4、u8来分别代表1个字节、2个字节、4个字节和8个字节的无符号数，无符号数可以用来描述数字、索引引用、数量值或者按照UTF-8编码构成字符串值。

表：由多个无符号数或者其他表作为数据项构成的复合数据类型，所有表都习惯性地以\_info结尾。表用于描述有层次关系的复合结构的数据，整个Class文件本质上就是一张表。

|  |  |  |
| --- | --- | --- |
| Class 文件格式 | | |
| 类型 | 名称 | 数量 |
| u4 | magic | 1 |
| u2 | minor\_version | 1 |
| u2 | major\_version | 1 |
| u2 | constant\_pool\_count | 1 |
| cp\_info | constant\_pool | constant\_pool\_count-1 |
| u2 | access\_flags | 1 |
| u2 | this\_class | 1 |
| u2 | super\_class | 1 |
| u2 | interfaces\_count | 1 |
| u2 | interfaces | interfaces\_count |
| u2 | fields\_count | 1 |
| field\_info | fields | fields\_count |
| u2 | methods\_count | 1 |
| method\_info | methods | methods\_count |
| u2 | attributes\_count | 1 |
| attribute\_info | attributes | attributes\_count |

无论是无符号数还是表，当需要描述同一类型但数量不定的多个数据时，经常会使用：**一个前置的容量计数器** + **若干个连续的数据项形式**，这时称这一系列连续的某一类型的数据为**某一类型的集合**。

TestClass.class文件

addr = 0x00000000 … const #1 const #2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **CA** | **FE** | **BA** | **BE** | **00** | **00** | **00** | **32** | **00** | **16** | **07** | **00** | **02** | **01** | **00** | **1D** |
| Magic Number 魔数  值 = 0xCAFEBABE | | | | Minor  Version  次版本号  0x0000 | | Major  Version  主版本号  0x0032 =  50,JDK1.6 | | constant\_  pool\_count  常量池容量计数值,计数  从1开始。  0x0016=22D  即有21项  索引值取值范围[1,21] | | tag=  0x07  即表示该常量为类或接口的符号引用 | name index  索引值，指向常量池中一个CONSTANT  \_Utf8\_info类型常量，此常量代表类/接口的全限定名  0x0002指向常量池第2项常量 | | tag=  0x01即表示该常量为UTF-  8编码的字符串 | length=  0x001D=29D  表示该utf-  8缩略编码  的字符串长  度大小是29个字节 | |

addr = 0x00000010 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | | **C** | **D** | **E** | **F** |
| **6F** | **72** | **67** | **2F** | **66** | **65** | **6E** | **69** | **78** | **73** | **6F** | **66** | | **74** | **2F** | **63** | **6C** |
| ‘o’ | ‘r’ | ‘g’ | ‘/’ | ‘f’ | ‘e’ | ‘n’ | ‘I’ | ‘x’ | ‘s’ | ‘o’ | ‘f’ | ‘t’ | | ‘/’ | ‘c’ | ‘l’ |

addr = 0x00000020 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | | **C** | **D** | **E** | **F** |
| **61** | **7A** | **7A** | **2F** | **54** | **65** | **73** | **74** | **43** | **6C** | **61** | **73** | | **74** | **07** | **00** | **04** |
| ‘a’ | ‘z’ | ‘z’ | ‘/’ | ‘T’ | ‘e’ | ‘s’ | ‘t’ | ‘C’ | ‘l’ | ‘a’ | ‘s’ | ‘s’ | |  |  |  |

addr = 0x00000030 …

…

**魔数：**

每个Class文件的头4个字节，用于确定该文件是否为一个能被虚拟机接受的Class文件。gif、jpeg文件头都存有魔数。使用魔数而不用扩展名识别主要基于安全方面的考虑，因为文件扩展名可以被随意改动。Class文件的魔数值为：0xCAFEBABE

**主版本号、次版本号：**

java版本号从45开始，JDK1.1之后，每个JDK大版本发布，主版本号向上 + 1。高版本的JDK能向下兼容以前版本的Class文件，但不能运行以后版本的Class文件，即使文件格式并未发生任何变化，虚拟机也必须拒绝执行超过其版本号的Class文件。0x0032 = 50D -> JDK1.6

eg：JDK1.1 能支持 45.0 ~ 45.65535 1111111111111111 = 2^17-1

JDK1.2 能支持 45.0 ~ 46.65535

JDK1.7 能支持 45.0 ~ 51.65535

JDK1.8 能支持 45.0 ~ 52.65535

**常量池**：

Class文件之中的资源仓库。是Class文件结构中与其他项目关联最多的数据类型，也是占用Class文件空间最大的数据项目之一，同时它还是在Class文件中第一个出现的表类型数据项目。常量池入口需放置一项u2类型的数据，代表常量池容量计数值(constant\_pool\_count)。注意这个容量计数是从1开始的。0x0016(22D)表示常量池中有21项常量，索引值范围为1~21。将第0项常量空出来，目的在于满足后面某些指向常量池的索引值的数据在特定情况下需要表达“不引用任何一个常量池项目”的含义。这种情况就可以把索引值置为0来表示。

注：**Class文件结构中只有常量池的容量计数是从1开始**，对于其他集合类型，包括接口索引集合、字段表集合、方法表集合等的容量计数都与一般习惯相同，是从0开始的。

常量池中主要存放两大类常量：**字面量**(Literal)和**符号引用**(Symbolic References)。

字面量：接近于Java语言层面的常量概念，如：

（1）文本字符串，

（2）声明为final的常量值等。

符号引用：属于编译原理方面的概念，包括以下三类常量：

（1）类和接口的全限定名

（2）字段的名称和描述符

（3）方法的名称和描述符

注：在Class文件中不会保存各个方法、字段的最终内存布局信息，因此这些字段、方法的符号引用不经过运行期转换的话无法得到真正的内存入口地址，也就无法直接被虚拟机使用。当虚拟机运行时，需要**从常量池获得对应的符号引用**，再在**类创建时或运行时解析、翻译到具体的内存地址**之中。

常量池中**每一项常量都是一个表**，JDK1.7共有14种结构各不相同的表结构数据。

|  |  |  |
| --- | --- | --- |
| 常量池的项目类型 | | |
| 类型 | 标志(tag) | 描述 |
| CONSTANT\_Utf8\_info | 1 | UTF-8编码的字符串 |
| CONSTANT\_Integer\_info | 3 | 整型字面量 |
| CONSTANT\_Float\_info | 4 | 浮点型字面量 |
| CONSTANT\_Long\_info | 5 | 长整型字面量 |
| CONSTANT\_Double\_info | 6 | 双精度浮点型字面量 |
| CONSTANT\_Class\_info | 7 | 类或接口的符号引用 |
| CONSTANT\_String\_info | 8 | 字符串类型字面量 |
| CONSTANT\_Fieldref\_info | 9 | 字段的符号引用 |
| CONSTANT\_Methodref\_info | 10 | 类中方法的符号引用 |
| CONSTANT\_InterfaceMethodref\_info | 11 | 接口中方法的符号引用 |
| CONSTANT\_NameAndType\_info | 12 | 字段或方法的部分符号引用 |
| CONSTANT\_MethodHandle\_info | 15 | 表示方法句柄 |
| CONSTANT\_MethodType\_info | 16 | 标识方法类型 |
| CONSTANT\_InvokeDynamic\_info | 18 | 表示一个动态方法调用点 |

表开始第一个字节是一个u1类型的标志位(tag)，代表当前这个常量属于哪种常量类型。

UTF-8缩略编码与普通UTF-8编码的区别是：

（1）从 ‘\u0001’ 到 ‘\u007f’ 之间的字符（相当于1~127的ASCII码）的缩略编码使用一个字节表示；

（2）从 ‘\u0080’ 到 ‘\u07ff’ 之间的所有字符的缩略编码用两个字符表示；

（3）从 ‘\u0800’ 到 ‘\uffff’ 之间的所有字符的缩略编码就按照普通UTF-8编码规则使用三个字节表示

|  |  |  |  |
| --- | --- | --- | --- |
| 常量池中的14种常量项的结构总表 | | | |
| 常量 | 常量表项 | 类型 | 取值及描述 |
| CONSTANT\_Utf8\_info | tag | u1 | 值为1,表示该常量为utf-8编码的字符串 |
| length | u2 | 字符串占用的字节数 |
| bytes | u1 | 长度为length的utf-8编码的字符串 |
| CONSTANT\_Integer\_info | tag | u1 | 值为3,表示该常量为整型字面量 |
| bytes | u4 | 按照高位在前存储的int值 |
| CONSTANT\_Float\_info | tag | u1 | 值为4,表示该常量为浮点型字面量 |
| bytes | u4 | 按照高位在前存储的float值 |
| CONSTANT\_Long\_info | tag | u1 | 值为5,表示该常量为长整型字面量 |
| bytes | u4 | 按照高位在前存储的long值 |
| CONSTANT\_Double\_info | tag | u1 | 值为6,表示该常量为双精度浮点型字面量 |
| bytes | u8 | 按照高位在前存储的Double值 |
| CONSTANT\_Class\_info | tag | u1 | 值为7, 表示该常量为类或接口的符号引用 |
| name\_index | u2 | 是一个指向全限定名常量项CONSTANT\_Utf8\_  info的索引 |
| CONSTANT\_String\_info | tag | u1 | 值为8, 表示该常量为字符串类型字面量 |
| index | u2 | 是一个指向字符串字面量的索引 |
| CONSTANT\_Fieldref\_info | tag | u1 | 值为9, 表示该常量为字段的符号引用 |
| index | u2 | 是一个指向声明字段的类或者接口描述符CONSTANT\_Class\_info的索引项 |
| index | u2 | 是一个指向字段描述符CONSTANT\_NameAnd  Type的索引项 |
| CONSTANT\_Methodref\_info | tag | u1 | 值为10, 表示该常量为类中方法的符号引用 |
| index | u2 | 是一个指向声明方法的类描述符常量项CONST  ANT\_Class\_info的索引 |
| index | u2 | 是一个指向名称及类型描述符常量项CONSTANT\_NameAndType的索引 |
| CONSTANT\_InterfaceMetho  dref\_info | tag | u1 | 值为11, 表示该常量为接口中方法的符号引用 |
| index | u2 | 是一个指向声明方法的接口描述符常量项CONSTANT\_Class\_info的索引 |
| index | u2 | 是一个指向名称及类型描述符常量项CONSTANT\_NameAndType的索引 |
| CONSTANT\_NameAndType  \_info | tag | u1 | 值为12, 表示该常量为字段或方法的部分符号引用 |
| index | u2 | 是一个指向该字段或方法名称常量项的索引 |
| index | u2 | 是一个指向该字段或方法描述符常量项的索引 |
| **CONSTANT\_MethodHandle**  **\_info** | tag | u1 | 值为15, 表示该常量为方法句柄 |
| reference\_  kind | u1 | 值的取值必须在[1,9], 它决定了方法句柄的类型。方法句柄类型的值表示方法句柄的字节码行为 |
| reference\_  index | u2 | 值必须是对常量池的有效索引 |
| **CONSTANT\_MethodType\_**  **info** | tag | u1 | 值为16, 表示该常量为标识方法类型 |
| descriptor\_  index | u2 | 值必须是对常量池的有效索引, 常量池在该索引处的项必须是CONSTANT\_Utf8\_info结构, 表示方法的描述符 |
| **CONSTANT\_InvokeDynamic\_**  **info** | tag | u1 | 值为18, 表示该常量为一个动态方法调用点 |
| bootstrap\_  method\_attr  \_index | u2 | 值必须是对当前Class文件中引导方法表的  bootstrap\_methods[] 数组的有效索引。 |
| name\_and\_typ  e\_index | u2 | 值必须是对当前常量池的有效索引, 常量池在该索引处的项必须是CONSTANT\_NameAndType  \_info结构, 表示方法名和方法描述符 |

蓝色部分为JDK1.7新增

**访问标志**

在**常量池结束之后，紧接着两个字节代表访问标志（access\_flags）**，这个标志用于标识一些**类或者接口层次**的访问信息，包括：这个Class是类还是接口；是否定义为public类型；是否定义为abstract类型；如果是类的话，是否被声明为final等。

|  |  |  |
| --- | --- | --- |
| 访问标志 | | |
| 标志名称 | 标志值 | 含义 |
| ACC\_PUBLIC | 0x0001 | 是否为public类型 |
| ACC\_FINAL | 0x0010 | 是否被声明为final，只有类可设置 |
| ACC\_SUPER | 0x0020 | 是否允许使用invokespecial字节码指令的新语意。标志必须为真 |
| ACC\_INTERFACE | 0x0200 | 标识这是一个接口 |
| ACC\_ABSTRACT | 0x0400 | 是否为abstract类型，对于接口或者抽象类来说，此标志值为真，其他类值为假 |
| ACC\_SYNTHETIC | 0x1000 | 标识这个类并非由用户代码产生的 |
| ACC\_ANNOTATION | 0x2000 | 标识这是一个注解 |
| ACC\_ENUM | 0x4000 | 标识这是一个枚举 |

eg: public class TestClass {}

ACC\_PUBLIC、ACC\_SUPER 为真

ACC\_FINAL、ACC\_INTERFACE、ACC\_ABSTRACT、ACC\_SYNTHETIC、ACC\_ANNOTATION、ACC\_ENUM 为假

∴ access\_flags = 0x0001|0x0020 = 0x0021

**类索引、父类索引与接口索引集合**

类索引、父类索引、接口索引集合按顺序排列在访问标志之后；

类索引（this\_class）和父类索引（super\_class）都是一个u2类型的数据，它们各自指向一个类型为CONSTANT\_Class\_info的类描述符常量，通过CONSTANT\_Class\_info类型的常量中的索引值（name\_index），可以查找到定义在CONSTANT\_Utf8\_info类型的常量中的全限定名字符串。

类索引查找全限定名过程

**this\_class** **CONSTANT\_Class\_info**  **CONSTANT\_Utf8\_info**

{u2 **index=0x0001** { u1 tag=0x07 { u1 tag=0x01

} u2 **name\_index=0x0002** u2 length=0x001D

} u1 bytes=“org/fenixsoft/clazz/

TestClass”

}

**常量池中**

而接口索引集合（interfaces）是一组u2类型的数据的集合，入口的第一项为u2类型的接口计数器（interfaces\_count），表示索引表的容量。如果该类没有实现任何接口，则该计数器值为0，后面接口的索引表不再占用任何字节。

Class文件中由这三项数据来确定这个类的继承关系：

类索引用于确定这个类的全限定名；

父类索引用于确定这个类的父类的全限定名。由于Java只允许单继承，因此父类索引只有一个。除java.lang.Object外，所有Java类的父类索引都不为0；

接口索引集合用于描述这个类实现了哪些接口。这些被实现的接口将按implements语句（如果这个类本身是一个接口，则应当是extends语句，因为接口可以继承多个接口）后的接口顺序从左到右排列在接口索引集合中。

addr = 0x00000000 … const #1 const #2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **CA** | **FE** | **BA** | **BE** | **00** | **00** | **00** | **32** | **00** | **16** | **07** | **00** | **02** | **01** | **00** | **1D** |
| Magic Number | | | | Minor  Version | | Major  Version | | constant\_  pool\_count | | tag=  0x07 | name\_index  =0x0002 | | tag=  0x01 | length=  0x001D=29D | |

addr = 0x000000**1**0 …

addr = 0x000000**2**0 …

… **常量池{}**

addr = 0x000000**E**0 … const #21

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | | **C** | **D** | **E** | **F** |
| **0E** | **54** | **65** | **73** | **74** | **43** | **6C** | **61** | **73** | **73** | **2E** | **6A** | | **61** | **76** | **61** | **00** |
|  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  | 访问标志 |

addr = 0x000000**F**0 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | 7 | 8 | 9 | A | B | | C | D | E | F |
| **21** | **00** | **01** | **00** | **03** | **00** | **00** | 00 | 01 | 00 | 02 | 00 | | 05 | 00 | 06 | 00 |
| 0x0  021 | 类索引=0x0001 | | 父类索引=0x0003 | | 接口计数器=0x0000 | |  |  |  |  |  |  | |  |  |  |

**字段表集合**

字段表（field\_info）用于描述接口或者类中声明的变量；

字段（field）包括类级变量以及实例级变量，但不包括在方法内部声明的局部变量。

字段信息：

（1）字段的作用域：public、private、protected修饰符

（2）实例变量还是类变量：是否有static修饰符

（3）可变性：final修饰符

（4）并发可见性：volatile修饰符，是否强制从主内存读写

（5）可否被序列化：transient修饰符

（6）字段数据类型：基本类型、对象、数组

（7）字段名称

上述信息中，修饰符类的字段信息都是布尔值，即要么有该修饰符，要么没有。因此适合用标志位表示。

而字段的名字，字段的数据类型都是无法固定的，必须引用常量池中的常量来描述。

|  |  |  |  |
| --- | --- | --- | --- |
| 字段表结构 | | | |
| 类型 | 名称 | 数量 | 作用 |
| u2 | access\_flags | 1 | 字段标志，访问修饰符 |
| u2 | name\_index | 1 | 对常量池的引用，字段的简单名称索引 |
| u2 | descriptor\_index | 1 | 对常量池的引用，字段的描述符索引 |
| u2 | attributes\_count | 1 |  |
| attribute\_info | attributes | attributes\_count | 属性表集合  {attributes\_name\_index; …; …} |

|  |  |  |
| --- | --- | --- |
| 字段访问标志  （access\_flags，字段访问修饰符） | | |
| 标志名称 | 标志值 | 含义 |
| ACC\_PUBLIC | 0x0001 | 字段是否被public修饰 |
| ACC\_PRIVATE | 0x0002 | 字段是否被private修饰 |
| ACC\_PROTECTED | 0x0004 | 字段是否被protected修饰 |
| ACC\_STATIC | 0x0008 | 字段是否被static修饰 |
| ACC\_FINAL | 0x0010 | 字段是否被final修饰 |
| ACC\_VOLATILE | 0x0040 | 字段是否被volatile修饰 |
| ACC\_TRANSIENT | 0x0080 | 字段是否被transient修饰 |
| ACC\_SYNTHETIC | 0x1000 | 字段是否由编译器自动产生 |
| ACC\_ENUM | 0x4000 | 字段是否被enum修饰 |

注：根据java语言规则：

（1）ACC\_PUBLIC、ACC\_PRIVATE、ACC\_PROTECTED 最多选其一

（2）ACC\_FINAL、ACC\_VOLATILE 不能同时选择

（3）接口（interface）之中的字段必须有ACC\_PUBLIC、ACC\_STATIC、ACC\_FINAL标志（如果我们没

有在编程中指定，编译时会默认产生并加入到字段表集合中）

public interface test {

int a = 10; //编译后：public static final int a = 10;

}

**“全限定名”、“简单名称”、“描述符”的概念和区别**

全限定名：”org/fenixsoft/clazz/TestClass;”

简单名称：没有类型和参数修饰的方法或者字段名称，如类中的inc()方法和m字段的简单名称为 ”inc” 和 ”m”

描述符：用来描述字段的数据类型，或方法的参数列表（包括数量、类型、顺序）和返回值。

根据描述符规则，基本数据类型（byte、char、double、float、int、long、short、boolean）以及代表无返回值的void类型都用一个大写字母表示，而对象类型则用字母L加对象的类全限定名来表示。

|  |  |
| --- | --- |
| 描述符标识字符含义 | |
| 标识字符 | 含义 |
| B | 基本类型byte |
| C | 基本类型char |
| D | 基本类型double |
| F | 基本类型float |
| I | 基本类型int |
| J | 基本类型long |
| S | 基本类型short |
| Z | 基本类型boolean |
| V | 基本类型void |
| L | 对象类型，如Ljava/lang/Object |

对于数组类型，每一维度使用一个前置的 “[” 字符来描述。例如：

（1）一个定义为 “java.lang.String[][]” 类型的二维数组，将被记录为“[[Ljava/lang/String”;

（2）一个整型数组“int[]”将被记录为 “[I”;

对于方法，按照先参数列表，后返回值类型的顺序描述，参数列表按照参数的严格顺序放在一组小括号“()”之内。例如：

（1）方法 void inc() “()V”

（2）方法 java.lang.String.toString() “()Ljava/lang/String;”

（3）方法 int indexOf(char[] source, int sourceOffset, int sourceCount, char[] target, int targetOffset, int targetCount, int fromIndex) “([CII[CIII)I”

“org/fenixsoft/clazz/TestClass”

addr = 0x00000000 … const #1 const #2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **CA** | **FE** | **BA** | **BE** | **00** | **00** | **00** | **32** | **00** | **16** | **07** | **00** | **02** | **01** | **00** | **1D** |
| Magic Number | | | | Minor  Version | | Major  Version | | constant\_  pool\_count | | tag=  0x07 | name\_index  =0x0002 | | tag=  0x01 | length=  0x001D=29D | |

addr = 0x00000010 …

… const #3->const #4 “java/lang/Object” **常量池{}**

addr = 0x000000**E**0 … const #21

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | | **C** | **D** | **E** | **F** |
| **0E** | **54** | **65** | **73** | **74** | **43** | **6C** | **61** | **73** | **73** | **2E** | **6A** | | **61** | **76** | **61** | **00** |
|  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  | 访问标志 |

**字段表集合（private int m）**

const #5 “m” const #6 “I”

addr = 0x000000F0 …I

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **21** | **00** | **01** | **00** | **03** | **00** | **00** | **00** | **01** | **00** | **02** | **00** | **05** | **00** | **06** | **00** |
| 0x0  021 | 类索引  this\_class  =**0x0001**  指向第1项常量 | | 父类索引  super\_class  =**0x0003**  指向第3项常量 | | 接口计数器  interfaces\_c  ount  =0x0000 | | 容量计数器fields\_count  = 0x0001,表明该类只有  1个字段表 | | 访问标志  access\_flags  = 0x0002,仅  ACC\_PRIVATE为真，余为假  private | | 字段简单名称索引，对常量池的引用  name\_index  =**0x0005**, 即指向第5项常量  “m” | | 字段描述符索引，对常  量池的引用descriptor  \_index =**0x0006,**  即指向第6项常量项  “I” | | 属性表计数器  attri  bute  s\_co  unt =0x0  000 |

**方法表集合（public void <init>()、public void inc()）**

const #7 “<init>” const #8 “()V” const #9 “Code”

addr = 0x00000100 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **00** | **00** | **02** | **00** | **01** | **00** | **07** | **00** | **08** | **00** | **01** | **00** | **09** | **00** | **00** | **00** |
| 即没有额外描述的信息 | 容量计数器  method\_cou  nt  = 0x0002  即有2个方  法 | | access\_flags  =0x0001,  public | | name\_index  =0x0007  “<init>” | | descriptor  \_index  =0x0008  “()V” | | attributes\_co  unt = 0x0001  即该方法的属性表集合有一项属性 | | attributes\_na  me\_index =  0x0009  “code”说明此属性是方法的字节码描述 | |  |  |  |

**方法表集合**

对方法的描述与对字段的描述在结构上几乎完全一致。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法表结构 | | | |
| 类型 | 名称 | 数量 | 作用 |
| u2 | access\_flags | 1 | 方法访问标志，修饰符 |
| u2 | name\_index | 1 | 对常量池的引用，方法的简单名称索引 |
| u2 | descriptor\_index | 1 | 对常量池的引用，方法的描述符索引 |
| u2 | attributes\_count | 1 |  |
| attribute\_info | attributes | attributes\_count | 属性表集合 |

|  |  |  |
| --- | --- | --- |
| 方法访问标志 | | |
| 标志名称 | 标志值 | 含义 |
| ACC\_PUBLIC | 0x0001 | 方法是否被public修饰 |
| ACC\_PRIVATE | 0x0002 | 方法是否被private修饰 |
| ACC\_PROTECTED | 0x0004 | 方法是否被protected修饰 |
| ACC\_STATIC | 0x0008 | 方法是否被static修饰 |
| ACC\_FINAL | 0x0010 | 方法是否被final修饰 |
| **ACC\_SYNCHRONIZED** | 0x0020 | 方法是否被synchronized修饰 |
| ACC\_BRIDGE | 0x0040 | 方法是否是由编译器产生的桥接方法 |
| ACC\_VARARGS | 0x0080 | 方法是否接受不定参数 |
| **ACC\_NATIVE** | 0x0100 | 方法是否被native修饰 |
| **ACC\_ABSTRACT** | 0x0400 | 方法是否被abstract修饰 |
| **ACC\_STRICTFP** | 0x0800 | 方法是否被strictfp修饰 |
| ACC\_SYNTHETIC | 0x1000 | 方法是否是由编译器自动产生的 |

方法里的Java代码，经过编译器编译成字节码指令后，存放在方法属性表集合中一个名为“Code”的属性里面。

与字段表集合相对应，如果父类方法在子类中没有被重写，方法表集合中就不会出现来自父类的方法信息。但同样的，有可能会出现由编译器自动添加的方法，最典型的便是类构造器“<clinit>”方法和实例构造器“<init>”方法。

在Java语言中，要**重载**（Overload）一个方法，除了要与原方法具有相同的简单名称之外，还要求必须拥有一个与原方法不同的特征签名，**特征签名**就是一个方法中各个参数在常量池中的字段符号引用的集合，即方法返回值是不包括在特征签名之内的。因此Java语言层面无法仅仅依靠返回值的不同来对一个已有方法进行重载。而在字节码层面（class文件）中，如果两个方法有相同的名称和特征签名，但返回值不同，那么也是可以合法共存在同一个Class文件中的。

**属性表集合**

在每个字段表、每个方法表都可以携带自己的属性表集合，以用于描述某些场景专有的信息。

属性表集合不要求各个属性表和其他数据项目一样具有严格的顺序、长度和内容。只要不和已有属性名重复，任何人实现的编译器都可以向属性表中写入自己定义的属性信息，Java虚拟机运行时会忽略掉它不认识的属性。为了能正确解析Class文件。《Java虚拟机规范（Java SE 7）》版中，预定义属性有21项，具体如下：

|  |  |  |
| --- | --- | --- |
| 虚拟机规范预定义的所有属性 | | |
| 属性名称 | 使用位置 | 含义 |
| **Code** | 方法表 | java代码编译成的字节码指令 |
| **ConstantValue** | 字段表 | final关键字定义的常量值 |
| **Deprecated** | 类、方法表、字段表 | 被声明为deprecated的方法和字段 |
| **Exceptions** | 方法表 | 方法抛出的异常 |
| EnclosingMethod | 类文件 | 仅当一个类为局部类或者匿名类时才能拥有这个属性，这个属性用于标识这个类所在的外围方法 |
| **InnerClasses** | 类文件 | 内部类列表 |
| **LineNumberTable** | Code属性 | Java源码的行号与字节码指令的对应关系 |
| **LocalVariableTable** | Code属性 | 方法的局部变量描述 |
| **StackMapTable** | Code属性 | 供类型检查验证器（Type Checker）检查和处理目标方法的局部变量和操作数栈所需要的类型是否匹配 |
| **Signature** | 类、方法表、字段表 | 用于支持泛型情况下的方法签名，在Java语言中，任何类、接口、初始化方法或成员的泛型签名如果包含了类型变量（Type Variables）或参数化类型（Parameterized Types），则Signature属性会为它记录泛型签名信息。由于Java的泛型采用擦除法实现，在为了避免类型信息被擦除后导致签名混乱，需要这个属性记录泛型中的相关信息 |
| **SourceFile** | 类文件 | 记录源文件名称 |
| SourceDebugExtension | 类文件 | 用于存储额外的调试信息 |
| **Synthetic** | 类、方法表、字段表 | 标识方法或字段为编译器自动生成的 |
| LocalVariableTypeTable | 类 | 使用特征签名代替描述符，是为了引入泛型语法之后能描述泛型参数化类型而添加 |
| RuntimeVisibleAnnotations | 类、方法表、字段表 | 为动态注解提供支持。用于指明哪些注解是运行时（实际上运行时就是进行反射调用）可见的。 |
| RuntimeInvisibleAnnotations | 类、方法表、字段表 | 指明哪些注解是运行时不可见的。 |
| RuntimeVisibleParameterAnnotations | 方法表 | 指明哪些方法参数注解是运行时可见的。 |
| RuntimeInvisibleParameterAnnotations | 方法表 | 指明哪些方法参数注解是运行时不可见的。 |
| AnnotationDefault | 方法表 | 用于记录注解类元素的默认值 |
| **BootstrapMethods** | 类文件 | 用于保存invokedynamic指令引用的引导方法限定符 |

**对每个属性，它的名称都需要从常量池中引用一个CONSTANT\_Utf8\_info类型的常量来表示**，而属性值的结构则是完全自定义的，只需要通过一个u4的长度属性去说明属性值所占用的位数即可。

一个符合规则的属性表应该满足如下定义的结构：

|  |  |  |
| --- | --- | --- |
| 属性表结构 | | |
| 类型 | 名称 | 数量 |
| u2 | attribute\_name\_index | 1 |
| u4 | attribute\_length | 1 |
| u1 | info | attribute\_length |

**Code属性的属性表**（Class文件中最重要的属性）

|  |  |  |  |
| --- | --- | --- | --- |
| Code属性表的结构 | | | |
| 类型 | 名称 | 数量 | 作用 |
| u2 | attribute\_name\_index | 1 | 是一个指向CONSTANT  \_Utf8\_info型常量的索引，常量值固定为“Code” |
| u4 | attribute\_length | 1 | 属性值长度，固定为：  属性表长度 – 6 bytes |
| u2 | max\_stack | 1 | 操作数栈深度的最大值 |
| u2 | max\_locals | 1 | 表示局部变量表所需的存储空间。单位为Slot。 |
| u4 | code\_length | 1 | java源程序编译后的字节码长度 |
| u1 | code | code\_length | 字节码指令的一系列字节流 |
| u2  **非**  **必**  **须** | exception\_table\_length | 1 |  |
| exception\_info | exception\_table | exception\_table\_length |  |
| u2 | attributes\_count | 1 |  |
| attribute\_info | attributes | attributes\_count |  |

注：（1）max\_stack，即操作数栈深度的最大值。在方法执行的任意时刻，操作数栈都不会超过这个深度。虚拟机运行时根据这个值来分配**栈帧**（Stack Frame）**中的操作栈**深度。

（2）max\_locals，表示局部变量表所需的存储空间。单位为Slot。**Slot是虚拟机为局部变量分配内存所使用的最小单位**，对于byte、char、float、int、short、boolean和returnAddress等长度不超过32位的数据类型，每个局部变量占用1个Slot，而double和long这两种64位的数据类型则需要2个Slot来存放。方法参数（包括实例方法中的**隐藏参数（“this”）**、显式异常处理器的参数（Exception Handler Parameter，就是try-catch语句中**catch块所定义的异常**）、方法体中定义的**局部变量**都需要使用局部变量表来存放。另外局部变量表的Slot可以重用，当代码执行超出一个局部变量的作用域时，这个局部变量所占的Slot可以被其它局部变量所使用。javac编译器会根据变量的作用域来分配Slot给各个变量使用，然后计算出max\_locals的大小。

（3）code，表示字节码指令的一系列字节流。称为字节码指令，就是指每条指令就是一个u1类型的单字节。虚拟机毎读取code中的一个字节码，就可以对应找出这个字节码代表的是什么指令，并且可以知道这条指令后面是否需要跟随参数，以及参数应当如何理解。u1数据类型的取值范围为[0x00，0xFF]，即[0,255]，所以可以表示256条指令。当前JVM规范定义了约200条编码值对应的指令含义。值得注意的是u4类型的code\_length实际只能使用u2的长度，因为JVM规范明确限制一个方法不允许超过65535条字节码指令。如果超过这个限制，javac编译器也会拒绝编译。

Code属性是Class文件中最重要的一个属性，如果把一个Java程序中的信息分为代码（Code，方法体里面的Java代码）和元数据（Metadata，包括类、字段、方法定义及其他信息）两部分，那么在整个Class文件中，Code属性用于描述代码，所有的其他数据项目用于描述元数据。

**方法表集合**

**<init>()方法表** **init()方法表的属性表集合**

const #7 “<init>” const #8 “()V” const #9 “**Code**”

addr = 0x00000100 … **Code属性表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **00** | **00** | **02** | **00** | **01** | **00** | **07** | **00** | **08** | **00** | **01** | **00** | **09** | **00** | **00** | **00** |
|  | 容量计数器  method\_cou  nt  = 0x0002  即有2个方  法 | | access\_flags  =0x0001,  public | | name\_index  =0x0007  “<init>” | | descriptor  \_index  =0x0008  “()V” | | attributes\_co  unt = 0x0001  即该方法的属性表集合有一项属性 | | attributes\_na  me\_index =  0x0009  “**Code**”说明此属性是方法的字节码描述 | | attribute\_length = 0x0000002F  = 47 bytes  属性值的长度为47 bytes | | |

**字节码指令**

addr = 0x00000110 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **2F** | **00** | **01** | **00** | **01** | **00** | **00** | **00** | **05** | **2A** | **B7** | **00** | **0A** | **B1** | **00** | **00** |
|  | max\_stack  = 0x0001 | | max\_locals  = 0x0001 | | code\_length = 0x00000005  字节码区域长度为5，虚拟机按照顺序依次读入5个字节 | | | | code  =2A B7 00 0A B1 | | | | |  |  |

0x2A: aload\_0 指令 0xB7: invokespecial指令 u2参数：0x000A: 0xB1:return void

const #10: CONSTANT\_Methodref\_info常量

“<init>”

addr = 0x00000120 …

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **A** | **B** | **C** | **D** | **E** | **F** |
| **00** | **02** | **00** | **0C** | **00** | **00** | **00** | **0E** | **00** | **01** | **00** | **00** | **0C** | **03** | **00** | **0D** |
|  |  | |  | |  | | | |  | | | | |  |  |

可以发现字节码执行过程中的数据交换、方法调用等操作都是基于栈（操作栈）的。所以，Java虚拟机执行字节码是基于栈的体系结构

检验：

javac:编译源程序生成字节码

javap:反编译工具，根据class字节码文件，反解析出当前类对前类对应的常量池、code区属性表（汇编指令）、本地变量表、异常表和代码行偏移量映射表等等。

源代码: javac编译下述代码：

package org.fenixsoft.clazz;

public class TestClass {

private int m;

public int inc(){

return m + 1;

}

}

命令:javap –verbose TestClass

//… 在这里省略常量池的输出

{

public org.fenixsoft.clazz.TestClass(); this关键字

Code:

Stack=1, Locals=1, Args\_size=1

0: aload\_0

1: invokespecial #10; //Method java/lang/Object."<init>":()V

4: return

LineNumberTable:

line 3: 0

LocalVariableTable:

Start Length Slot Name Signature

0 5 0 this Lorg/fenixsoft/clazz/TestClass;

**表示int()方法参数个数。无static修饰方法，则有this参数传入，**

**值至少为1；如有用static修饰方法，则无Args\_size值至少为0**

public int inc();

Code:

Stack=2, **Locals**=**1**, **Args\_size**=**1**

0: aload\_0

1: getfield #18; //Field m:I

4: iconst\_1

5: iadd

6: ireturn

**表示int()方法内部的局部变量个数。由于javac编译时会把this关键字的访问转变为对一个普通方法参数的访问，然后在虚拟机调用实例方法时自动传入此参数。因此在实例方法的局部变量表中至少会存在一个指向当前对象实例的局部变量，局部变量表也会预留出第一个Slot位来存放对象实例的引用，方法参数值从1开始计算。**

LineNumberTable:

line 8: 0

LocalVariableTable:

Start Length Slot Name Signature

0 7 0 this Lorg/fenixsoft/clazz/TestClass;

}

在字节码指令之后的是这个方法的显式异常处理表（下文简称异常表）集合。异常表对于Code属性来说并不是必须存在的。

属性表格式如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 属性表结构 | | | |
| 类型 | 名称 | 数量 | 作用 |
| u2 | start\_pc | 1 | 当字节码在第start\_pc行到第end\_pc之间（不含第end\_pc行）出现了类型为catch\_type（catch\_type为指向一个CONSTANT\_Class\_info型常量的索引）或者其子类的异常，则转到第handler\_pc行继续处理。当catch\_type的值为0时，代表任意异常情况都需要转向到handler\_pc处进行处理。 |
| u2 | end\_pc | 1 |
| u2 | handler\_pc | 1 |
| u2 | catch\_type | 1 |

虽然异常也是java代码的一部分，但是如今javac编译器使用异常表而不是简单的跳转命令来实现Java异常及finally处理机制。

检验：

package org.fenixsoft.clazz;

public class TestException {

public int inc() {

int x;

try {

x = 1;

return x;

} catch (Exception e) {

x = 2;

return x;

} finally {

x = 3;

}

}

}

注：局部变量表存放如下：

数组元素[0]对应 ：变量实例对象this

数组元素[1]对应 ：变量x

数组元素[2]对应 ：returnValue、e

数组元素[3]对应 ：returnValue2

数组元素[4]对应 ：不属于java.lang.Excption及其子类的异常的引用

//编译后的code字节码以及异常表

public int inc();

Code:

Stack=1, Locals=5, Args\_size=1

0: iconst\_1 //将int型常量1压入操作数栈（try块中的 x=1）

1: istore\_1 //将操作数栈顶int型元素出栈，并保存到局部变量表的第2个位置

2: iload\_1 //将局部变量表第2个位置的int型元素入栈

3: istore\_2 //将操作数栈顶int型元素出栈，并保存到局部变量表第3个位置 returnValue

**出现其他**

**异常情况**

**任一位置出现指定异常**

**没有异常**

4: iconst\_3 //将int型常量3压入操作数栈（finally块中的 x = 3）

**执行finally并return**

5: istore\_1 //将操作数栈顶int型元素出栈，并保存到局部变量表的第2个位置

6: iload\_2 //将局部变量表第3个位置的int型元素入栈（将returnValue中的值放到栈顶，准备

给ireturn返回）

注：这就是为什么try和catch语句块中的返回值会被保留下来，即使finally语句块修改了返回值的大小。

7: ireturn //返回栈顶int元素（从当前方法返回int，执行完finally后才真正返回）

8: astore\_2 //将栈顶引用型数值出栈，并保存到局部变量表的第3个位置（给catch中定义的Exception e赋值，存储在Slot 2 中）

9: iconst\_2 //将int型常量2压入操作数栈（catch块中的 x = 2）

10: istore\_1 //将操作数栈顶int型元素出栈，并保存到局部变量表的第2个位置

11: iload\_1 //将局部变量表第2个位置的int型元素入栈

12: istore\_3 //将操作数栈顶int型元素出栈，并保存到局部变量表的第4个位置 returnValue2

13: iconst\_3 //将int型常量3压入操作数栈（finally块中的 x = 3）

**执行finally并return**

14: istore\_1 //将操作数栈顶int型元素出栈，并保存到局部变量表的第2个位置

15: iload\_3 //将局部变量表第4个位置的int型元素入栈（returnValue中的值放到栈顶，准备给

ireturn返回）

16: ireturn //返回栈顶int元素（从当前方法返回int，执行完finally后才真正返回）

17: astore 4 //将栈顶引用型数值出栈，并保存到局部变量表的第5个位置（如果出现了不属于java.lang.Exception及其子类的异常才会走到这里）

19: iconst\_3 //将int型常量3压入操作数栈（finally块中的 x = 3）

**执行finally并抛出异常**

20: istore\_1 //将操作数栈顶int型元素出栈，并保存到局部变量表的第2个位置

21: aload 4 //将局部变量表第5个位置的引用类型（异常）入栈

23: athrow //抛出异常

Exception table: //异常表记录

from to target type

0 4 8 Class java/lang/Exception

//执行路径1：如果try语句块中出现Exception或其子类的异常，则转到catch语句块处理

0 4 17 any

//执行路径2：如果try语句块中出现不属于Exception或其子类的异常，则转到finally语句块处理

8 13 17 any

//执行路径3：如果catch语句块中出现任何异常，则转到finally语句块处理