**第7章 虚拟机类加载机制**

虚拟机的类加载机制：

虚拟机把描述类的数据从Class文件加载到内存，并对数据进行校验、转换解析和初始化，最终形成可以被虚拟机直接使用的Java类型，这就是虚拟机的类加载机制。在Java语言里，类型的加载、连接和初始化过程都是在程序运行期间完成的。Java里天生可以动态扩展的语言特性就是依赖运行期动态加载和动态连接这个特点实现的。

eg：

（1）编写一个面向接口的应用程序，可以等到运行时再指定其实际的实现类；

（2）用户可以通过Java预定义的和自定义类加载器，让一个本地的应用程序可以在运行时从网络或其他地方加载一个二进制流作为程序代码的一部分，这种组装应用程序的方式目前已广泛应用于Java程序之中，如JSP、OSGi技术等，都使用了Java语言运行期类加载的特性。

类加载的时机

类字节码从被加载到虚拟机内存中开始，到卸载出内存为止，它的整个生命周期包括：加载（Loading）、验证（Verification）、准备（Preparation）、解析（Resolution）、初始化（Initialization）、使用（Using）和卸载（Unloading）7个阶段。其中验证、准备、解析3个部分统称为连接（Linking）。

类生命周期图

连接（Linking）

卸载

Unloading

初始化

Initialization

准备

Preparation

加载

Loading

验证

Verification

使用

Using

解析

Resolution

注：加载、验证、准备、初始化和卸载这5个阶段的顺序是确定的，类的加载过程必须按照这种顺序按部就班地“开始”——强调这点是因为这些阶段通常都是**互相交叉地混合式**进行的，通常会在一个阶段执行的过程中调用、激活另外一个阶段。

而解析阶段则不一定：它在某些情况下可以在初始化阶段之后再开始，这是为了支持Java语言的运行时绑定（称为动态绑定/晚期绑定）。

JVM没有说明什么时候可以开始类加载的第一个阶段——加载，因为Java虚拟机规范中将这点交给虚拟机的具体实现来自由把握。

但是虚拟机规范严格规定了**有且只有5种情况必须立即对类进行初始化**（而加载、验证、准备自然需要在此之前开始）：

（1）遇到new、getstatic、putstatic、invokestatic 这4条字节码指令时，如果类没有进行过初始化，则需要先触发初始化。

new指令对应的Java代码场景：new关键字实例化对象

getstatic指令对应的Java代码场景：读取一个类的静态字段（被final修饰、已在编译器把结果放入常量池的静态字段除外）

putstatic指令对应的Java代码场景：设置一个类的静态字段（被final修饰、已在编译器把结果放入常量池的静态字段除外）

invokestatic指令对应的Java代码场景：调用一个类的静态方法时

（2）使用java.lang.reflect包的方法对类进行反射调用，如果类没有进行过初始化，则需要触发初始化。

（3）当初始化一个类的时候，如果发现其父类还没有进行过初始化，则需要先触发其父类的初始化。

（4）当虚拟机启动时，用户需要指定一个要执行的主类（包含main()方法的那个类），虚拟机会初始化这个主类。

（5）当使用JDK1.7的动态语言支持时，如果一个java.lang.invoke.MethodHand

le实例最后的解析结果REF\_getStatic、REF\_putStatic、REF\_invokeStatic的方法句柄，并且这个方法句柄所对应的类没有进行过初始化，则需要先触发其初始化。

这五种场景的行为称为对一个类进行主动引用，除此之外，其它引用类的方式都不会触发初始化，称为被动引用。

接口与类初始化过程稍有不同，主要体现在第（3）种行为：即一个接口在初始化时，不要求其父接口全部都完成了初始化，只有在该接口使用到父接口的时候（如引用接口中定义的常量）才会初始化父接口。

加载

加载阶段JVM需要完成以下3件事情：

（1）通过一个类的全限定名来获取此类的二进制字节流

（2）将这个字节流所代表的静态存储结构转化为内存方法区的运行时数据结构

（3）在内存中生成一个代表这个类的java.lang.Class对象，作为方法区这个类的各种数据的访问入口。

非数组类加载阶段：

开发人员可控性最强，因为加载阶段可以使用系统提供的引导类加载器来完成，也可以由用户自定义的类加载器去完成，开发人员可以通过定义自己的类加载器去控制字节流的获取方式（即**重写一个类加载器的loadClass()方法**）

数组类加载阶段：

数组类本身不通过类加载器创建，它是由Java虚拟机直接创建的。但数组类与类加载器仍然有很密切的关系，因为数组类的元素类型（Element Type，指数组去掉所有维度的类型）最终是要靠类加载器去创建，一个数组类C创建过程需要遵循以下规则：

①如果数组的组件类型（Component Type，指数组去掉一个维度的类型）是引用类型，则递归采用JVM的加载过程去加载这个组件类型，数组C将在加载该组件类型的类加载器的类名称空间上被标识（一个类必须与类加载器一起确定唯一性）。

②如果数组的组件类型不是引用类型（如int[]数组），Java虚拟机将会把数组C标记为与引导类加载器关联。

③数组类的可见性与它的组件类型的可见性一致，如果组件类型不是引用类型，则数组类的可见性将默认为public。

加载阶段完成后，虚拟机外部的二进制字节流就按照虚拟机所需格式存储在方法区之中。然后在内存中实例化一个java.lang.Class类的对象（没有明确必须是在Java堆，因为对于HotSpot虚拟机而言，Class对象比较特殊，它虽然是对象，但是存放在方法区里），这个对象将作为程序访问方法区中的这些类型数据的外部接口。加载阶段与连接阶段的部分内容（如一部分字节码文件格式验证动作）是交叉进行的，加载阶段尚未完成，连接阶段可能已经开始。

验证

验证是连接阶段的第一步，这一阶段的目的是为了确保Class文件的字节流中包含的信息符合当前虚拟机的要求，并且不会危害虚拟机自身的安全。

验证阶段大致上会完成下面4个阶段的检验动作：

（1）文件格式验证

检验字节流是否符合Class文件格式的规范，并且能被当前版本的虚拟机处理。这阶段的验证是基于二进制字节流进行的，只有通过这个阶段的验证后，字节流才会进入内存的方法区中进行存储。所以后面的3个验证阶段全部是基于方法区的存储结构进行的，不会再直接操作字节流。

（2）元数据验证

对字节码描述的信息（类的元数据信息）进行语义分析，以保证其描述的信息符合Java语言规范要求。

（3）字节码验证

最复杂的阶段。主要目的是通过数据流和控制流分析，确定程序语义是合法的、符合逻辑的。

（4）符号引用验证

发生在虚拟机将符号引用转化为直接引用的时候，这个转化动作将在连接阶段的第三阶段——解析阶段中发何时能。符号引用验证可以认为是对类自身以外（常量池中的各种符号引用）的信息进行匹配性校验。

准备

准备阶段是正式为类变量分配内存并设置类变量初始值的阶段，这些变量所使用的内存都将在方法区中进行分配。

应该注意的是，这时候进行内存分配的仅包括类变量（被static修饰的变量），而不包括实例变量，实例变量将会在对象实例化时随着对象一起分配在Java堆中。其次，这里所说的初始值“通常情况”下是数据类型的零值。假设一个类变量的定义为：

public static int value = 123;

则变量value在准备阶段过后的初始值为0而不是123，因为这时候尚未开始执行任何Java方法，而把value赋值为123的putstatic指令是程序被编译后，存放于类构造器<clinit>()方法之中，所以把value赋值为123的动作将在初始化阶段后才会执行。

|  |  |
| --- | --- |
| 基本数据类型的零值 | |
| 数据类型 | 零值 |
| int | 0 |
| long | 0L |
| short | (short)0 |
| char | ‘\u0000’ |
| byte | (byte)0 |
| boolean | false |
| float | 0.0f |
| double | 0.0d |
| reference | null |

如果类字段的字段属性表存在ConstantValue属性，则在准备阶段变量value就会被初始化为ConstantValue属性所指定的值，如：

public static final int value = 123;

编译时Javac将会为value生成ConstantValue属性，在准备阶段虚拟机就会根据ConstantValue的设置将value赋值为123。

解析

解析阶段是虚拟机将常量池内的符号引用替换为直接引用的过程，符号引用在Class文件中以CONSTANT\_Class\_info、CONSTANT\_Fieldref\_info、CONSTANT\_Methodref\_info等类型的常量出现。

符号引用（Symbolic References）：

符号引用以一组符号来描述所引用的目标，符号可以是任何形式的字面量，只要使用时能无歧义地定位到目标即可。符号引用与虚拟机实现的内存布局无关，引用的目标并不一定已经加载到内存中。各种虚拟机实现的内存布局可以各不相同，但是它们能接受的符号引用必须都是一致的。

直接引用（Direct References）：

直接引用可以是直接指向目标的指针、相对偏移量或是一个能间接定位到目标的句柄。直接引用是和虚拟机实现的内存布局相关的，同一个符号引用在不同虚拟机实例上翻译出来的直接引用一般不会相同。如果有了直接引用，那引用的目标必定已经在内存中存在。

用于操作符号引用的字节码指令

|  |  |  |  |
| --- | --- | --- | --- |
| anewarray | checkcast | getfield | getstatic |
| instanceof | invokedynamic | invokeinterface | invokespecial |
| invokestatic | invokevirtual | ldc | ldc\_w |
| multianewarray | new | putfield | putstatic |

除invokedynamic指令外，虚拟机实现可以对第一次解析的结果进行缓存（在运行时常量池中记录直接引用，并把常量标识为已解析状态）从而避免解析动作重复进行。无论是否真正执行了多次解析动作，虚拟机需要保证的是在同一个实体中，如果一个符号引用之前已经被成功解析过，那么后续的引用解析请求就应当一致成功；同样如果第一次解析失败，那么其他指令对这个符号的解请求也应当收到相同的异常。

invokedynamic指令，上述规则不成立。当碰到某个前面已经由