目 录

[1. 概述 1](#_Toc2137929493)

[1.1 术语 1](#_Toc1718793721)

[2. 整个llvm相关的部分 1](#_Toc210082896)

[2.1 数据结构 1](#_Toc1706334525)

[2.1.1 PassAnalysisSupport相关数据结构 1](#_Toc1506659269)

[2.1.1.1 AnalysisUsage 1](#_Toc1951610518)

[3. llvm command line解析llvm::cl workspace 2](#_Toc184393696)

[3.1 相关的接口函数 2](#_Toc698150981)

[3.2 相关全局变量 2](#_Toc2069649864)

[3.2.1 GlobalParser 2](#_Toc1709610010)

[3.2.2 GeneralCategory 2](#_Toc1054148973)

[3.2.3 TopLevelSubCommand 2](#_Toc805453909)

[3.2.4 AllSubCommands 2](#_Toc1347144797)

[3.3 相关数据结构 2](#_Toc1454474772)

[3.3.1 ExpansionContext 2](#_Toc417539702)

[3.3.1.1 构造函数 2](#_Toc2146441699)

[3.3.2 CommandLineParser 3](#_Toc782178270)

[3.3.2.1 registerSubCommand 注册subcommand 3](#_Toc2009850058)

[3.4 回调函数 3](#_Toc2019985964)

[3.4.1 TokenizeGNUCommandLine 3](#_Toc1113294483)

[4. ADT代码解析 3](#_Toc815464694)

[4.1 数据结构 3](#_Toc295036081)

[4.1.1 int数值相关 3](#_Toc1293446048)

[4.1.2 浮点数值相关 4](#_Toc1353680210)

[4.1.2.1 IEEEFloat 4](#_Toc1576276713)

[4.1.2.2 APFloat 5](#_Toc300251706)

[4.2 small vector代码解析 5](#_Toc1764996676)

[4.3 Twine代码解析 5](#_Toc356966466)

[4.3.1 概述 5](#_Toc1543394751)

[4.3.2 构造函数 6](#_Toc1840664841)

[5. IR代码解析 6](#_Toc111116236)

[5.1 数据结构 6](#_Toc1533840596)

[5.1.1 Type 6](#_Toc1411974914)

[5.1.2 FunctionType 6](#_Toc321199132)

[5.1.3 Use 7](#_Toc1092691473)

[5.1.4 DataLayout相关数据结构 7](#_Toc771150535)

[5.1.4.1 DataLayout 7](#_Toc125326002)

[5.1.5 Value相关 7](#_Toc1277085169)

[5.1.5.1 Value 7](#_Toc1469301517)

[5.1.5.2 GlobalValue 7](#_Toc47492218)

[5.1.6 User 8](#_Toc839211532)

[5.1.7 OperandTraits相关类 8](#_Toc375966842)

[5.1.7.1 FixedNumOperandTraits 8](#_Toc852946127)

[5.1.7.2 OptionalOperandTraits 8](#_Toc38872681)

[5.1.7.3 VariadicOperandTraits 8](#_Toc1830441614)

[5.1.7.4 HungoffOperandTraits 8](#_Toc1270485830)

[5.1.7.5 DECLARE\_TRANSPARENT\_OPERAND\_ACCESSORS 宏定义 9](#_Toc37830732)

[5.1.7.6 DEFINE\_TRANSPARENT\_OPERAND\_ACCESSORS 宏定义 9](#_Toc465136236)

[5.1.8 Constant相关数据结构 9](#_Toc1132852240)

[5.1.8.1 Constant 9](#_Toc2057816696)

[5.1.8.2 ConstantData 9](#_Toc1578430720)

[5.1.8.3 ConstantFP 10](#_Toc1948316934)

[5.1.8.4 ConstantExpr 10](#_Toc205369130)

[5.1.9 Comdat 10](#_Toc724393120)

[5.1.10 GlobalObject 10](#_Toc1154513497)

[5.1.11 BasicBlock 10](#_Toc1781645843)

[5.1.12 指令类相关 11](#_Toc1024644826)

[5.1.12.1 Instruction 11](#_Toc772026525)

[5.1.12.2 ReturnInst 11](#_Toc2138612310)

[5.1.12.3 基于ReturnInst显示特化OperandTraits 11](#_Toc420555929)

[5.1.12.4 InstVisitor 类 11](#_Toc465207718)

[5.1.13 Function相关 12](#_Toc102244898)

[5.1.13.1 Argment 12](#_Toc1954396525)

[5.1.13.2 Function 12](#_Toc1877182632)

[5.1.14 LLVMContext 13](#_Toc423444030)

[5.1.15 Module 13](#_Toc899604351)

[5.1.16 IRBuilder相关 14](#_Toc500849520)

[5.1.16.1 IRBuilderFolder 14](#_Toc548770032)

[5.1.16.2 ConstantFolder 14](#_Toc29205872)

[5.1.16.3 IRBuilderBase 14](#_Toc1970151037)

[5.1.16.4 IRBuilder 14](#_Toc596262250)

[5.1.17 DominatorTree相关类 15](#_Toc868417404)

[5.1.18 Pass相关的类 15](#_Toc198634231)

[5.1.18.1 PassInfoMixin 15](#_Toc1449208378)

[5.1.18.2 AnalysisInfoMixin 16](#_Toc907290086)

[5.1.19 LegacyPassManager相关类 16](#_Toc2029075845)

[5.1.19.1 legacy::PassManagerBase 16](#_Toc572210560)

[5.1.20 Verifier相关类 16](#_Toc945120818)

[5.1.20.1 DominatorTree 16](#_Toc346728433)

[5.1.21 AsmWriter 相关类 16](#_Toc1705062800)

[5.1.21.1 ModuleSlotTracker 17](#_Toc855453867)

[5.1.21.2 TypePrinting 17](#_Toc1925159153)

[5.1.21.3 SlotTracker 17](#_Toc1505896087)

[5.2 接口 17](#_Toc1060822997)

[5.2.1 ConstantFold相关接口 17](#_Toc502068625)

[5.2.1.1 ConstantFoldBinaryInstruction 17](#_Toc512925936)

[5.2.2 Verifier相关接口 17](#_Toc694985192)

[5.2.2.1 verifyFunction 17](#_Toc1526713451)

[6. support 代码解析 17](#_Toc1284952461)

[6.1 数据结构 17](#_Toc686113854)

[6.1.1 TrailingObjects相关数据结构 17](#_Toc1947269380)

[6.2 接口 18](#_Toc1750160180)

[6.2.1 TargetSelect相关接口 18](#_Toc788358752)

[6.2.1.1 InitializeNativeTarget 18](#_Toc1754182258)

[6.2.1.2 InitializeNativeTargetAsmPrinter 18](#_Toc1479859164)

[6.2.1.3 InitializeNativeTargetAsmParser 18](#_Toc1211802782)

[7. Transforms相关代码解析 18](#_Toc506302961)

[7.1 数据结构 18](#_Toc1980708684)

[7.1.1 Utils 相关 18](#_Toc1760572814)

[7.1.1.1 InstructionWorklist 18](#_Toc535508833)

[7.1.2 InstCombine相关 19](#_Toc1803376073)

[7.1.2.1 InstructionCombiningPass 19](#_Toc209351417)

[7.1.3 Scalar 相关 19](#_Toc1403926238)

[7.2 接口 19](#_Toc2002010304)

[7.2.1 InstCombine相关接口 19](#_Toc1658559795)

[7.2.1.1 createInstructionCombiningPass 19](#_Toc163732676)

[7.2.2 Scalar 相关接口 19](#_Toc1883602501)

[7.2.2.1 createReassociatePass 19](#_Toc83286707)

[8. ExecutionEngine 相关代码解析 19](#_Toc1108853494)

[8.1 数据结构 19](#_Toc82847287)

[8.1.1 orc ExecutorProcessControl相关结构体 19](#_Toc1788349507)

[8.1.1.1 ExecutorProcessControl 19](#_Toc1964307361)

[8.1.2 Orc core相关结构体 19](#_Toc2008006440)

[8.1.2.1 ExecutionSession 20](#_Toc1146761946)

[8.2 接口 20](#_Toc877646710)

[9. Target相关代码解析 20](#_Toc362591418)

[9.1 数据结构 20](#_Toc1659687882)

[9.2 接口 20](#_Toc1572631903)

[9.2.1 X86相关 20](#_Toc1889304869)

[9.2.1.1 LLVMInitializeX86Target 20](#_Toc797156696)

[9.2.1.2 LLVMInitializeX86AsmPrinter 20](#_Toc111262109)

[10. 参考文档 20](#_Toc1689090602)

# 概述

## 术语

|  |  |
| --- | --- |
| 术语 | 解释 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| POD | POD 是 Plain Old Data 的缩写，是 C++ 定义的一类数据结构概念，比如 int、float 等都是 POD 类型的。Plain 代表它是一个普通类型，Old 代表它是旧的，与几十年前的 C 语言兼容，那么就意味着可以使用 memcpy() 这种最原始的函数进行操作。两个系统进行交换数据，如果没有办法对数据进行语义检查和解释，那就只能以非常底层的数据形式进行交互，而拥有 POD 特征的类或者结构体通过二进制拷贝后依然能保持数据结构不变。也就是说，能用 C 的 memcpy() 等函数进行操作的类、结构体就是 POD 类型的数据。 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# 整个llvm相关的部分

## 数据结构

### PassAnalysisSupport相关数据结构

#### AnalysisUsage

### PassSupport相关数据结构

# llvm command line解析llvm::cl workspace

## 相关的接口函数

|  |  |
| --- | --- |
| 函数名 | 解释 |
| ParseCommandLineOptions | 解析命令行，并保存信息到全局变量中。 |
| getGeneralCategory | general option category |
|  |  |
| getFileOrSTDIN | 读取输入文件或stdio，返回读取内容到MemoryBuffer。 |
|  |  |

## 相关全局变量

### GlobalParser

static ManagedStatic<CommandLineParser> GlobalParser;

### GeneralCategory

static OptionCategory GeneralCategory{"General options"};

### TopLevelSubCommand

ManagedStatic<SubCommand> llvm::cl::TopLevelSubCommand

### AllSubCommands

ManagedStatic<SubCommand> llvm::cl::AllSubCommands

## 相关数据结构

### ExpansionContext

#### 构造函数

基于BumpPtrAllocator和TokenizerCallback。

### CommandLineParser

#### registerSubCommand 注册subcommand

初始化全局变量TopLevelSubCommand和AllSubCommands。

## 回调函数

### TokenizeGNUCommandLine

# ADT代码解析

## 数据结构

### int数值相关

文件：APInt.h

说明：任意精度整数的类。

APInt 是常见大小写无符号整数类型（如“unsigned”、“unsigned long”或“uint64\_t”）的功能替代，但也允许non-byte-width的整数大小和超大的整数值类型，例如 3-bits、15-bits或超过 64 位的精度。 APInt 提供了多种算术运算符和方法来操作任何位宽的整数值。它同时支持典型的整数算术和比较运算以及按位操作。

该类有几个值得注意的不变量：

* 所有位、字节和字位置都是从零开始的。
* 一旦位宽设置后，它不会更改，除非通过 Truncate、SignExtend 或 ZeroExtend 操作。
* 所有二元运算符必须在具有相同位宽的 APInt 实例上。尝试在实例上使用不同的位宽的这些运算符将产生断言。
* 该值被规范地存储为无符号值。对于操作有所不同，操作有有符号和无符号变体。例如，sdiv 和 udiv。但是，由于位宽必须相同，因此无论值是否被解释为有符号，诸如 Mul 和 Add 之类的操作都会产生相同的结果。
* 一般来说，该类尝试遵循 LLVM 在其 IR 中使用的计算风格。这简化了它在 LLVM 中的使用。
* APInt 支持零位宽值，但需要位的操作未在其上定义（例如，你不能要求零位整数的符号）。这意味着定义了零扩展和逻辑移位等操作，但没有定义符号扩展和 ashr。零位值比较和散列相等自己，countLeadingZeros 返回 0。

类操作：

* static WordType tcAdd(WordType \*, const WordType \*, WordType carry, unsigned)

DST += RHS + CARRY where CARRY is zero or one. Returns the carry flag.

### 浮点数值相关

文件：APFloat.h

说明：独立于主机和目标的任意精度浮点软件实现。

APFloat 使用 APInt 类中的静态函数提供的 bignum 整数算术。lib使用 bignum 整数，其部分是至少 16 位宽的任何无符号类型，但建议使用 64 位。

为清晰而不是速度而写，特别是为了交叉编译器在前端使用，以便目标算法可以在主机上正确执行。尽管如此，性能应该是合理的，特别是对于其预期用途。它在开发更快的目标特定库期间，作为一个运行时库的基础实现可能很有用的。

IEEE-754R 草案中的所有 5 种舍入模式都针对所有已实现的操作进行了正确处理。当前实现的操作是加法， 减法、乘法、除法、融合乘法加法、转换为浮点数、转换为整数和从整数转换。可以使用三或四行代码添加新的舍入模式（例如，远离零）。

内置四种格式：IEEE 单精度、双精度、四倍精度和 x87 80 位扩展双精度（以完全扩展精度运行时）。添加符合 IEEE 语义的新格式只需要添加两行代码：格式的声明和定义。

所有操作都以exception bit-mask的形式返回该操作的状态，因此可以连续执行多个操作与他们的结果或运算在一起。返回的状态可用于编译器诊断；例如，不精确、下溢和溢出可以很容易地在常量折叠上诊断出来，编译器优化器可以确定折叠操作会引发哪些异常并相应地优化或不优化。

目前，下溢微小是在四舍五入后检测的；应该直接添加支持舍入前的情况。

该库按照 C99 读取十六进制浮点数，并在必要时根据指定的舍入模式正确舍入。语法必须被调用者验证。它还根据 C99 %a 和 %A 转换将浮点数转换为十六进制文本。 输出精度（或者自然最小精度）可以被指定；如果请求的精度小于自然精度，则针对指定的舍入模式正确舍入输出。

它还读取十进制浮点数并根据指定的舍入模式正确舍入。

当前未实现到十进制文本的转换。

非零有限数在内部表示为一个符号位、一个 16 位有符号指数和一个有效数字整数部分数组。数精度 P 归一化后，指数在格式范围内，如果数不是非正规数，则尾数的第 P 位设置为显式整数。对于非正规数，最高有效位向右移动，以便指数保持在格式的最小值，因此最小的非正规数只有有效数的最低有效位。零和无穷大的符号很重要；此类数字的指数和尾数未存储，但具有已知的隐式（确定性）值：有效数为 0，零指数为 0，所有指数位为1 用于无穷大指数。对于 NaN，符号和有效数是确定性的，尽管没有真正意义，并且在非转换操作中保留。指数是隐含的全 1 位。

除了默认异常处理之外，APFloat不提供任何异常处理。We represent Signaling NaNs via IEEE-754R 2008 6.2.1 should clause by encoding Signaling NaNs with the first bit of its trailing significand as 0.

#### IEEEFloat

#### APFloat

操作：

* opStatus add(const APFloat &RHS, roundingMode RM)

根据语义，调用IEEE.add或Double.add操作。

## small vector代码解析

## Twine代码解析

### 概述

Twine-一种轻量级数据结构，用于有效地将临时值串接为字符串。

Twine是一种绳子，它使用二叉树表示连接的字符串，其中字符串是节点的前序。由于使用结果时可以将Twine有效地渲染到缓冲区中，因此避免了为中间字符串结果生成临时值的成本——特别是在从不需要Twine结果的情况下。通过显式跟踪叶节点的类型，我们还可以避免为转换操作创建临时字符串（例如，将整数附加到字符串）。  
Twine不打算直接使用，也不应该存储，它的实现依赖于存储指向临时堆栈对象的指针的能力，这些对象会在在语句末尾取消分配。当API希望接受可能连接的字符串时，只能将Twins用作参数中的常量引用。

Twine支持特殊的“null”值，它总是可以连接到表单本身，并呈现为空字符串。这This can be returned from APIs to effectively nullify any concatenations performed on the result。

**实现**

由于Twine的性质，Twine的串联方法不可能构造内部节点；结果必须在返回值内表示。因此，一个Twine对象实际上包含两个值，即串联的左侧和右侧。我们还有空的Twine对象，它们实际上是表示空字符串的 sentinel values 。

因此，Twine可以有效地有零个、一个或两个孩子。\see isNullary（）、\see isUnary（）和\see isBinary（）谓词用于测试孩子的数量。

我们在Twine对象上维护了许多不变量（FIXME:为什么）：

-Nullary twines总是在左侧表示其种类，而在右侧表示空的种类。

-Unary twines线始终用左侧的值表示，右边是空的。

-如果一个Twine有另一个Twine作为child，则该child应始终是二进制的（否则它可能已折叠到parent中）。

这些不变量请参见isValid（）。

**效率考虑**

twines的设计目的是为常见情况生成高效且小的代码。因此，concat（）方法被内联，以便连接叶节点的操作可以被优化为直接存储到单个堆栈分配的对象中。

实际上，并非所有的编译器都可以完全优化concat（），因此我们提供了两种额外的方法：（以及伴随的运算符+重载），以确保特别重要的情况（cstring加StringRef）按需生成代码。

### 构造函数

|  |  |
| --- | --- |
| 函数声明 | 解释 |
| Twine() | Construct from an empty string |
| Twine(const Twine &) | 默认复制构造函数 |
| Twine(const char \*Str) | 通过输入字符串构造Twine;如果字符串是’\0’,LHSKind是EmptyKind;否则，LHSKind是CStringKind，且LHS.cString为Str。 |
| Twine(std::nullptr\_t) | 禁用 |
| Twine(const std::string &Str) | 通过std::string构造Twine;LHSKind是StdStringKind，LHS.stdString是指向Str的指针。 |
| Twine(const std::string\_view &Str) | 通过std::string\_view构造Twine，并且把它转换成一个指针和一个length。LHS.ptrAndLength.ptr指向std::string\_view的data()，LHS.ptrAndLength.length等于std::string\_view的length() |
| Twine(const StringRef &Str) | 通过StringRef构造Twine。LHS.ptrAndLength.ptr指向StringRef.的data()，LHS.ptrAndLength.length等于StringRef.的size() |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# IR代码解析

## 数据结构

### Type

文件：Type.h

说明：Type 类的实例是不可变的：它们一旦被创建，就永远不会改变。另请注意，只会创建特定类型的一个实例。因此，查看两种类型是否相等是进行简单的指针比较的问题。为了强制不创建两个相等的实例，只能通过类 Type 和派生类中的静态工厂方法创建 Type 实例。一旦分配，类型永远不会被释放。

### FunctionType

文件：DerivedTypes.h

说明：标识funtion type，继承于Type。

### Use

文件：Use.h

说明：Use 表示值定义与其用户之间的一条边。

这在名义上是一个二维链表。它支持遍历特定值定义的所有用途。它也支持，当我们从用户的操作数到达时，端口直接跳转到已使用的值;当我们从值的使用中到达时，端口直接跳转到用户。

### DataLayout相关数据结构

文件：DataLayout.h

说明：此文件定义与数据类型大小/偏移量/对齐方式信息相关的布局属性。 它使用惰性批注来缓存有关如何布局和使用结构类型的信息。

此结构应该创建一次，如果默认值不正确，则填写，然后由 const&传递。 任何成员函数都不需要修改对象。

#### DataLayout

说明：目标数据布局字符串的解析版本以及用于查询它的方法。 目标数据布局字符串由target\*指定。 需要生成 LLVM IR 的前端才能生成被codegen的目标的正确的目标数据。

### Value相关

文件：Value.h，GlobalValue.h，ValueHandle.h

#### Value

文件：Value.h

说明：LLVM 值表示。这是一个非常重要的 LLVM 类。它是程序计算的所有值的基类，可用作其他值的操作数。Value是其他重要类的超类，例如，Instruction和Function。所有值都有一个类型。类型不是Value的子类。某些值可以具有名称，并且它们属于某个Module。 在值上设置名称会自动更新模块的符号表。

每个值都有一个“使用列表”，用于跟踪哪些其他值正在使用此值。 一个值还可以具有任意数量的 ValueHandle 对象，这些对象watch并侦听 RAUW 和销毁事件。 有关详细信息，请参阅 llvm/IR/ValueHandle.h。

#### GlobalValue

文件：GlobalValue.h

说明：该类是所有全局可定义对象的公共基类。例如，它由 GlobalVariable、GlobalAlias 和 Function 子类化。

#### ValueHandleBase

### User

文件：User.h

说明：此类定义使用 Value 的人必须实现的接口。Value 类的实例跟踪用户对其处理的内容。

Instructions 是最大的Users class。Constants可能是其他Constants的Users（想想数组之类的）

### OperandTraits相关类

文件：User.h，OperandTraits.h

说明：该文件定义了特征类，这些特征类有助于强制执行各种用户子类的正确布局。它还提供了以最有效的方式访问操作数的方法。

特别说明：OperandTraits 主模板类是在User.h上声明的(template <class> struct OperandTraits;)。主模板类只声明没有定义，如果用到主模板类，会是不完整类型，在编译时会报“xxx has incomplete type and cannot be defined”，见https://en.cppreference.com/w/cpp/language/type#Incomplete\_type。

#### FixedNumOperandTraits

声明：template <typename SubClass, unsigned ARITY> struct FixedNumOperandTraits；

说明：FixedNumOperandTraits - 当 Use 数组是 User 对象的前缀时，确定 Use 数组的分配机制，并且 Use 对象的数量在编译时已知。

#### OptionalOperandTraits

声明：template <typename SubClass, unsigned ARITY> struct FixedNumOperandTraits；

说明：OptionalOperandTraits - 当操作数的数量可能在运行时发生变化时。本质上，它可能只会减少，因为分配可能不会改变。

#### VariadicOperandTraits

声明：template <typename SubClass, unsigned MINARITY = 0> struct VariadicOperandTraits；

说明：VariadicOperandTraits - 当 Use 数组是 User 对象的前缀时，确定 Use 数组的分配机制，并且 Use 对象的数量仅在分配时已知。

#### HungoffOperandTraits

声明：template <unsigned MINARITY = 1> struct HungoffOperandTraits；

说明：HungoffOperandTraits - 当 Use 数组不是 User 对象的前缀，但分配在不相关的堆地址时，确定 Use 数组的分配机制。

这是 Use 数组必须可调整大小时所需的特征类。

#### DECLARE\_TRANSPARENT\_OPERAND\_ACCESSORS 宏定义

说明：在特定class内声明OperandTraits相关的操作；

#### DEFINE\_TRANSPARENT\_OPERAND\_ACCESSORS 宏定义

说明：在特定class内实现DECLARE\_TRANSPARENT\_OPERAND\_ACCESSORS 宏定义声明的OperandTraits相关的操作。

### Constant相关数据结构

文件：Constant.h Constants.h

#### Constant

说明：这是 LLVM 中的一个重要基类。它提供了 LLVM 程序中所有常量值的通用设施。常量是在运行时不可变的值。函数是常量，因为 EIR 地址是不可变的。与全局变量相同。

所有常量共享此类中提供的功能。所有常量都可以具有空值。他们可以有一个操作数列表。常量可以是简单的（整数和浮点值）、复杂的（数组和结构）或基于表达式（仅由某些运算符和其他常量值组成的常量值的计算产生）。

说明：常量是不可变的（一旦创建，它们永远不会改变），并且完全由结构等价共享。 这意味着两个结构上等效的常量将始终具有相同的地址。常量是根据需要按需创建的，永远不会被删除：因此客户端不必担心对象的生存期。

类操作：

* static Constant \*getBinOpIdentity(unsigned Opcode, Type \*Ty,bool AllowRHSConstant = false,bool NSZ = false)

返回二进制操作码的identity常量。当二元运算可交换时， 对于每个 X，identity常数 C 定义为 X op C = X 和 C op X = X。如果 binop 不可交换，调用者通过将 AllowRHSConstant 设置为 true 获取操作数 1 identity常量。例如，对于操作数 1，任何移位都有一个零identity常量：X 移位 0 = X。

如果这是一个 fadd/fsub 操作，我们不关心带符号的零，然后将 NSZ 设置为 true 返回标识 +0.0 而不是 -0.0。如果运算符没有identity常量，则返回 nullptr。

#### ConstantData

#### ConstantFP

说明：浮点值，float或double。

类操作：

* static ConstantFP \*get(LLVMContext &Context, const APFloat &V)

从Context.pImpl->FPConstants中获取指定的ConstantFP 对象，如果没有，则创建一个。

* static Constant \*get(Type \*Ty, double V)

对于指定类型的指定值，这将返回一个 ConstantFP 或包含一个 ConstantFP的splat 的向量。这应该只用于已知合法的简单常量值，如 2.0/1.0 等，既作为主机 double 又作为目标格式。

* static Constant \*get(Type \*Ty, const APFloat &V)

如果 Ty 是向量类型，则返回具有给定值的 splat 的常量。否则返回给定值的 ConstantFP对象。

#### ConstantExpr

说明：使用其他常量值的表达式初始化的常量值。

此类使用标准指令操作码来定义各种常量表达式。 ConstantExpr 类的 Opcode 字段在 Value::SubclassData 字段中维护。

类操作：

* static Constant \*get(unsigned Opcode, Constant \*C1, Constant \*C2, unsigned Flags = 0, Type \*OnlyIfReducedTy = nullptr);

### Comdat

略，https://maskray.me/blog/2021-07-25-comdat-and-section-group。

What is a Comdat object?

A Comdat section is a section in the object file, in which objects are placed, which can be duplicated in other object files. Each object has information for the linker, indicating what it must do when duplicates are detected. The options can be: Any — do anything, ExactMatch — duplicates must completely match, otherwise an error occurs, Largest — take the object with the largest value, NoDublicates — there should not be a duplicate, SameSize — duplicates must have the same size, otherwise an error occurs.

COMDAT段（Section）是对8086目标文件原始集合的扩展。 它是在Microsoft C 7.0添加的。

### GlobalObject

文件：GlobalObject.h

说明：它标识一个独立的对象。它可以是一个function或一个全局变量，但是，不能是一个别名。

### BasicBlock

文件：BasicBlock.h

说明：该类表示 LLVM 中的单个基本块。基本块只是顺序执行的指令的容器。基本块是Values，因为它们被指令引用，例如， branches和switchtables。 BasicBlock 的类型是“Type::LabelTy”，因为基本块表示分支可以跳转到的标签。

well-formed 的基本块由non-terminating指令后跟一个terminator 指令的列表组成。terminator 指令不能出现在基本块的中间，必须终止块。 BasicBlock 类允许出现格式错误的基本块，因为它可能在构建或修改程序的中间阶段很有用。然而，验证者将确保基本块是“well formed”。

### 指令类相关

文件：Instruction.h， Instructions.h

#### Instruction

说明：该类是所有LLVM指令的基类。

#### ReturnInst

说明： Return a value (possibly void), from a function. Execution does not continue in this function any longer.

#### 基于ReturnInst显示特化OperandTraits

#### InstVisitor 类

文件：InstVisitor.h

说明：指令访问者的基类。

当您想要对不同类型的指令执行不同的操作而不必使用大量强制转换和大的 switch 语句（在你的代码里，就是这样）时，可以使用指令访问者。要定义您自己的访问者，请继承此类，为“SubClass”模板参数指定您的新类型，并“override”您类中的 visitXXX 函数。我说“override”是因为使用此类是根据静态解析重载而不是虚函数定义的。

例如，这里是一个计算处理的 malloc 指令数的访问者：

|  |
| --- |
| /// Declare the class. Note that we derive from InstVisitor instantiated  /// with \_our new subclasses\_ type.  ///  struct CountAllocaVisitor : public InstVisitor<CountAllocaVisitor> {  unsigned Count;  CountAllocaVisitor() : Count(0) {}  void visitAllocaInst(AllocaInst &AI) { ++Count; }  }; |

该类可以按如下方式

|  |
| --- |
| CountAllocaVisitor CAV;  CAV.visit(function);  NumAllocas = CAV.Count; |

定义的类具有对指令的“visit”方法，也适用于 BasicBlock、Function 和 Module，它们递归地处理所有包含的指令。

请注意，如果您不为某些指令类型实现 visitXXX，指令 super 的 visitXXX 方法

类将被调用。因此，如果将来添加指令，如果您处理其中一个超类，它们将自动得到支持。

可选的第二个模板参数指定类型，构造访问函数应该返回该该类型。如果你指定这个类型，你\*必须\*提供 visitInstruction 的实现！。

注意这个类是专门设计成模板来避免虚函数调用开销。定义和使用 InstVisitor 与在指令操作码上使用自己的 switch 语句一样高效。

### Function相关

#### Argment

文件：Argment.h

说明：此类表示函数的传入formal 参数。formal参数，因为它是“formal”，不包含实际值，而是表示特定函数参数的类型、参数编号和属性。当在所述函数的主体中使用时，参数当然代表调用函数的实际参数的值。

#### Function

文件：Function.h

说明：表示 LLVM 中的单个函数/过程。

一个函数基本上由基本块列表、参数列表和符号表组成。

属性及解释：

|  |  |  |
| --- | --- | --- |
| 属性 | 类型 | 解释 |
| BasicBlocks | SymbolTableList<BasicBlock> | 基本块列表 |
| Arguments | mutable Argument \* | 参数 |
| NumArgs | size\_t | 参数个数 |
| SymTab | std::unique\_ptr<ValueSymbolTable> | 参数/指令的symbol 表 |
| AttributeSets | AttributeList | 参数属性。 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

类操作及解释：

* static Function \*Create(FunctionType \*Ty, LinkageTypes Linkage,unsigned AddrSpace, const Twine &N = "", Module \*M = nullptr)

static Function \*Create(FunctionType \*Ty, LinkageTypes Linkage,const Twine &N, Module &M)

新建一个function，把function放置在module的数据layout指定的程序地址空间中。

对象操作：

* void print(raw\_ostream &OS, AssemblyAnnotationWriter \*AAW,bool ShouldPreserveUseListOrder = false,bool IsForDebug = false) const;

文件：Function.h，AsmWriter.cpp

说明：Utility functions for printing and dumping Module objects

### LLVMContext

文件：LLVMContext.h

说明：这是在线程上下文中使用 LLVM 的重要类。 它（不透明地）拥有并管理LLVM核心基础设施的核心“global”数据，包括类型和常量单表。

LLVMContext 本身不提供锁定保证，因此您应该注意每个线程都有一个上下文。

### Module

文件: Module.h

说明：模块实例用于存储与 LLVM 模块相关的所有信息。模块是所有其他 LLVM 中间表示 （IR） 对象的顶级容器。每个模块直接包含全局变量列表、函数列表、此模块所依赖的库（或其他模块）列表、符号表以及有关目标特征的各种数据。

模块维护一个 GlobalList 对象，该对象用于保存对模块中全局变量的所有常量引用。 销毁全局变量时，全局列表中不应包含任何条目。

LLVM 中间表示形式的主容器类。

属性及解释：

|  |  |  |
| --- | --- | --- |
| 属性 | 类型 | 解释 |
|  |  |  |
|  |  |  |
|  |  |  |
| ValSymTab | std::unique\_ptr<ValueSymbolTable> | 全局变量和函数标识的symbol 表 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

操作及解释：

* const ValueSymbolTable &getValueSymbolTable() const { return \*ValSymTab; }

获取全局变量和函数标识的symbol 表。

* GlobalValue \*Module::getNamedValue(StringRef Name)

在model symbol表中返回指定名称的第一个全局值，可以是任何的类型。如果该名称没有找到对应的全局值，则返回null。

* Function \*Module::getFunction(StringRef Name)

在module symbol 表中查找特定的函数，如果不存在，则返回null。

### IRBuilder相关

文件：IRBuilderFolder.h， IRBuilder.h

#### IRBuilderFolder

说明：IRBuilder 中常量折叠的接口。

#### ConstantFolder

说明：ConstantFolder - 创建具有最小、目标独立、折叠的常量。

操作：

* Value \*FoldBinOp(Instruction::BinaryOps Opc, Value \*LHS,

Value \*RHS) const override

* Value \*FoldBinOpFMF(Instruction::BinaryOps Opc, Value \*LHS, Value \*RHS,

FastMathFlags FMF) const override

#### IRBuilderBase

说明：各种IRBuilders的通用基类。

操作：

* Value \*CreateFAdd(Value \*L, Value \*R, const Twine &Name = "",

MDNode \*FPMD = nullptr)

* CallInst \*CreateConstrainedFPBinOp(

Intrinsic::ID ID, Value \*L, Value \*R, Instruction \*FMFSource = nullptr,

const Twine &Name = "", MDNode \*FPMathTag = nullptr,

std::optional<RoundingMode> Rounding = std::nullopt,

std::optional<fp::ExceptionBehavior> Except = std::nullopt)

#### IRBuilder

文件：IRBuilder.h

说明：这提供了一个统一的 API，用于创建指令并将其插入基本块：在 BasicBlock 的末尾，或在块中的特定迭代器位置。

请注意，builder不会公开 LLVM 指令的完整通用性。 要访问额外的指令属性，请在创建指令后在指令上使用mutators（例如 setVolatile）。存在方便状态来指定fast-math标志和 fp-math 标记。

第一个模板参数指定用于创建常量的类。这默认为创建最小折叠的常量。 第二个模板参数允许客户端指定在每次新创建的插入时调用的自定义插入挂钩。

操作：

* Value \*CreateFAdd(Value \*L, Value \*R, const Twine &Name = "",MDNode \*FPMD = nullptr)
* ReturnInst \*CreateRet(Value \*V)

创建一个'ret <val>' 指令。

### DominatorTree相关类

文件：Support/GenericDomTree.h，Dominators.h

说明：该文件定义了一组模板，可以在通用图上有效地计算支配树。这通常在 LLVM 中用于对 CFG 进行快速支配查询，但它是完全通用的 w.r.t. 底层graph类型。

与 ADT/\* 图算法不同，通用支配树对图的 NodeRef 有更多的要求。 NodeRef 应该是一个指针，并且 NodeRef->getParent() 必须返回父节点，父节点也是一个指针或 DomTreeNodeTraits 需要专门化。

FIXME：也许 GenericDomTree 需要一个 TreeTraits，而不是 GraphTraits。

### Pass相关的类

文件：PassManager.h

说明：

* PassManager.h，此文件定义了用于 LLVM 中pass management的各种接口。LLVM本身没有“pass”接口。相反，任何类的实例支持在 IR 单元上“run”它的一个方法都可以用作一个pass。pass management通常是一种工具，用于收集在特定 IR 构造上运行的传递序列(a sequence of passes)，并在包含IR构造的每个此类构造上按顺序运行每个传递。由于模块没有包含 IR 构造，因此用于模块传递的管理器形成了基本情况，该基本情况在提供的单个模块上按顺序运行其管理的传递。

core IR lib提供了用于运行模块和函数传递的管理器。

* \* FunctionPassManager可以在一个模块上运行，在函数上运行每个pass。
* \* ModulePassManager必须直接运行，运行模块上的每个pass。

请注意，pass management的实现使用基于概念的多态性，如“Value Semantics and Concept-based”讨论中所述（or its abbreviated sibling "Inheritance Is The Base Class of Evil"）：

* <http://github.com/sean-parent/sean-parent.github.com/wiki/Papers-and-Presentations>
* http://www.youtube.com/watch?v=\_BpMYeUFXv8
* http://channel9.msdn.com/Events/GoingNative/2013/Inheritance-Is-The-Base-Class-of-Evil

#### PassInfoMixin

文件：PassManager.h

#### AnalysisInfoMixin

文件：PassManager.h

### LegacyPassManager相关类

文件：LegacyPassManager.h

说明：该文件定义了Legacy PassManager class。此类用于保存、维护和优化passes的执行。 PassManager class确保分析结果在一个pass运行之前可用，并且当 PassManager 被销毁时，Pass 也被销毁。

#### legacy::PassManagerBase

操作：

* virtual void add(Pass \*P) = 0

将pass添加到要运行的pass队列中。这会将 Pass 的所有权传递给 PassManager。当PassManager被销毁时，pass也会被销毁，所以不需要删除pass。如果发现pass是多余的，甚至可能会立即销毁该pass。这意味着所有pass都必须通过”new”来分配。

### Verifier相关类

文件：

#### DominatorTree

说明：用于计算normal dominator tree的 DominatorTreeBase 的具体子类。

定义：如果存在从函数入口到一个block的路径，则称该块是前向静态可达的。在优化过程中，静态可访问块可能会变得静态不可访问。

前向不可达块可能出现在支配树中，也可能不出现。如果是这样，支配查询将返回

结果好像所有可达的块都支配它。当请求与潜在无法访问的块对应的节点时，调用代码必须处理块无法访问的情况以及 getNode()的结果是 nullptr 的情况。、

通常，在构建支配树时，已知不可达的块不会在树中。在最初构造支配树后变得无法访问的块可能会一直存在于树中，即使树被正确更新。调用代码不应依赖于前面的语句；这只是为了帮助人类理解。

具体解释见《编译原理》中的机器无关代码优化部分。

### AsmWriter 相关类

文件：ModuleSlotTracker.h, AsmWriter.cpp

#### ModuleSlotTracker

文件：ModuleSlotTracker.h

说明：slot tracker 存储的抽象接口。

#### TypePrinting

#### SlotTracker

## 接口

### ConstantFold相关接口

#### ConstantFoldBinaryInstruction

Constant \*llvm::ConstantFoldBinaryInstruction(unsigned Opcode, Constant \*C1,Constant \*C2)

### Verifier相关接口

#### verifyFunction

# support 代码解析

## 数据结构

### TrailingObjects相关数据结构

说明：此文件定义对实现类的支持，这些类附加了一些trailing object（或对象数组）。主要目的是明确这个idiom的使用位置，并用法更惯用，更难出错。

trailing object模板抽象出用于分配和访问的reinterpret\_cast、指针算法和大小计算的附加对象数组，并注意它们都按所需的对齐方式分配。此外，它还确保基类型是final - 派生自一个类，该类期望在通常不安全之后立即追加数据。

用户应从此模板派生，并为除最后一个类型之外的每种trailing 类型提供 numTrailingObjects 实现，例如，像这个实例：

|  |
| --- |
| /// \code  class VarLengthObj : private TrailingObjects<VarLengthObj, int, double> {  friend TrailingObjects;  unsigned NumInts, NumDoubles;  size\_t numTrailingObjects(OverloadToken<int>) const { return NumInts; }  };  /// \endcode |

您可以通过“getTrailingObjects”访问附加的数组，并通过“additionalSizeToAlloc”和“totalSizeToAlloc”确定分配所需的大小。

此类实现的所有方法都旨在供类的实现使用，而不是作为其接口的一部分（因此，建议私有继承）。

## 接口

### TargetSelect相关接口

#### InitializeNativeTarget

声明：bool InitializeNativeTarget()

说明：InitializeNativeTarget - 主程序应调用此函数来初始化与主机对应的本机目标。 这对于 JIT 应用程序很有用，可确保目标以正确的方式链接。

客户端多次调用此函数是合法的。

#### InitializeNativeTargetAsmPrinter

声明：bool InitializeNativeTargetAsmPrinter()

说明：主程序应调用此函数来初始化本机目标 asm printer。

#### InitializeNativeTargetAsmParser

声明：bool InitializeNativeTargetAsmParser()

说明：主程序应调用此函数来初始化本机目标 asm parser。

# Transforms相关代码解析

## 数据结构

### Utils 相关

#### InstructionWorklist

### InstCombine相关

#### InstructionCombiningPass

### Scalar 相关

#### ReassociateLegacyPass

文件：lib/Transforms/Scalar/Reassociate.cpp

#### GVNLegacyPass

文件：lib/Transforms/Scalar/GVN.cpp

## 接口

### InstCombine相关接口

#### createInstructionCombiningPass

原型：FunctionPass \*llvm::createInstructionCombiningPass()

FunctionPass \*llvm::createInstructionCombiningPass(unsigned MaxIterations)

说明：创建一个InstructionCombiningPass实例。

### Scalar 相关接口

#### createReassociatePass

原型：FunctionPass \*llvm::createReassociatePass()

说明：创建一个ReassociateLegacyPass实例。

#### createGVNPass

原型：FunctionPass \*llvm::createGVNPass(bool NoMemDepAnalysis)

说明：创建一个GVNLegacyPass实例。

# ExecutionEngine 相关代码解析

## 数据结构

### orc ExecutorProcessControl相关结构体

文件：ExecutionEngine/Orc/ExecutorProcessControl.h

说明：支持与一个JIT 目标进程之间进行交互。

#### ExecutorProcessControl

说明：支持与一个JIT 目标进程之间进行交互。

### Orc core相关结构体

文件：llvm/include/llvm/ExecutionEngine/Orc/core.h

#### ExecutionSession

说明：一个ExecutionSession标识一个running JIT program。

## 接口

# Target相关代码解析

## 数据结构

## 接口

### X86相关

#### LLVMInitializeX86Target

#### LLVMInitializeX86AsmPrinter

# 参考文档