[1 基本概念 2](#_Toc497467738)

[1.1 注释 2](#_Toc497467739)

[**1.1.1 语法** 2](#_Toc497467740)

[**1.1.2 C风格** 3](#_Toc497467741)

[**1.1.3 C++风格** 3](#_Toc497467742)

[**1.1.4 注意** 3](#_Toc497467743)

[1.2 ASCII表 3](#_Toc497467744)

[1.3 名称和标识符 3](#_Toc497467745)

[**1.3.1 标识符** 3](#_Toc497467746)

[**1.3.2 Names** 5](#_Toc497467747)

[1.4 Types - Fundamental types 5](#_Toc497467748)

[**1.4.1 Type classification** 5](#_Toc497467749)

[**1.4.2 Type naming** 6](#_Toc497467750)

[**1.4.3 Elaborated type specifier** 8](#_Toc497467751)

[**1.4.4 Static type** 8](#_Toc497467752)

[1.4.6 Incomplete type 8](#_Toc497467753)

[1.5 Object - Scope – Lifetime 9](#_Toc497467754)

[1.6 Definitions and ODR 9](#_Toc497467755)

[1.10 转译阶段 10](#_Toc497467756)

[**1.10.1 阶段1** 10](#_Toc497467757)

[**1.10.2 阶段2** 10](#_Toc497467758)

[**1.10.3 阶段3** 11](#_Toc497467759)

[**1.10.4 阶段4** 11](#_Toc497467760)

[**1.10.5 阶段5** 11](#_Toc497467761)

[**1.10.6 阶段6** 12](#_Toc497467762)

[**1.10.7 阶段7** 12](#_Toc497467763)

[**1.10.8 阶段8** 12](#_Toc497467764)

[**1.10.9 阶段9** 12](#_Toc497467765)

[**1.10.10 注意** 12](#_Toc497467766)

[**1.10.11 参考** 12](#_Toc497467767)

[1.11 The main() function 12](#_Toc497467768)

[10 Templates 12](#_Toc497467769)

[10.9 Parameter pack 13](#_Toc497467770)

# **1 基本概念**

This section provides definitions for the specific terminology and the concepts used when describing the C++ programming language.

A C++ program is a sequence of text files (typically header and source files) that contain declarations. They undergo translation to become an executable program, which is executed when the OS calls its main function.

Certain words in a C++ program have special meaning, and these are known as keywords. Others can be used as identifiers. Comments are ignored during translation. Certain characters in the program have to be represented with escape sequences.

The entities of a C++ program are values, objects, references, functions, enumerators, types, class members, templates, template specializations, namespaces, parameter packs, and the “this” pointer. Preprocessor macros are not C++ entities.

Entities are introduced by declarations, which associate them with names and define their properties. The declarations that define all properties required to use an entity are definitions. A program must contain only one definition of any non-inline function or variable that is odr-used.

Definitions of functions include sequences of statements, some of which include expressions, which specify the computations to be performed by the program.

Names encountered in a program are associated with the declarations that introduced them using name lookup. Each name is only valid within a part of the program called its scope. Some names have linkage which makes them refer to the same entities when they appear in different scopes or translation units.

Each object, reference, function, expression in C++ is associated with a type, which may be fundamental, compound, or user-defined, complete or incomplete, etc.

Named objects and named references to objects are known as variables.

## **1.1 注释**

Comments serve as a sort of in-code documentation. When inserted into a program, they are effectively ignored by the compiler; they are solely intended to be used as notes by the humans that read source code. Although specific documentation is not part of the C++ standard, several utilities exist that parse comments with different documentation formats.

### **1.1.1 语法**

/\* 注释内容 \*/ (1)

// 注释内容 \n (2)

* Often known as "C-style" or "multi-line" comments.
* Often known as "C++-style" or "single-line" comments.

All comments are removed from the program at translation phase 3 by replacing each comment with a single whitespace character.

### **1.1.2 C风格**

C-style comments are usually used to comment large blocks of text, however, they can be used to comment single lines. To insert a C-style comment, simply surround text with /\* and \*/; this will cause the contents of the comment to be ignored by the compiler. Although it is not part of the C++ standard, /\*\* and \*/ are often used to indicate documentation blocks; this is legal because the second asterisk is simply treated as part of the comment. C-style comments cannot be nested.

### **1.1.3 C++风格**

C++-style comments are usually used to comment single lines, however, multiple C++-style comments can be placed together to form multi-line comments. C++-style comments tell the compiler to ignore all content between // and a new line.

### **1.1.4 注意**

Because comments are removed before the preprocessor stage, a macro cannot be used to form a comment and an unterminated C-style comment doesn't spill over from an #include'd file.

Besides commenting out, other mechanisms used for source code exclusion are

#if 0

std::cout << "this will not be executed or even compiled\n";

#endif

and

if(false) {

std::cout << "this will not be executed\n"

}

## **1.2 ASCII表**

The following chart contains all 128 ASCII decimal (dec), octal (oct), hexadecimal (hex) and character (ch) codes. The default is none.

## **1.3 名称和标识符**

### **1.3.1 标识符**

标识符由数字，下划线，小写字母和大写字母，和大部分的Unicode字符。合法的标识符不能以数字开头 ，而是以拉丁字母，下划线，或者Unicode编码的非数字字母开始。标识符是大小写敏感的，小写标识符和大写标识符是不同的。

Note: C++ grammar formally requires Unicode characters to be escaped with \u or \U, but due to translation phase 1, that is exactly how raw unicode characters from the source code are presented to the compiler. Also note that support of this feature may be limited, e.g. gcc.

**1.3.1.1 声明**

An identifier can be used to name objects, references, functions, enumerators, types, class members, namespaces, templates, template specializations, parameter packs, goto labels, and other entities, with the following exceptions:

* the identifiers that are keywords cannot be used for other purposes;
* the identifiers with a double underscore anywhere are reserved;
* the identifiers that begin with an underscore followed by an uppercase letter are reserved;
* the identifiers that begin with an underscore are reserved in the global namespace.

"Reserved" here means that the standard library headers #define or declare such identifiers for their internal needs, the compiler may predefine non-standard identifiers of that kind, and that name mangling algorithm may assume that some of these identifiers are not in use. If the programmer uses such identifiers, the behavior is undefined.

In addition, it's undefined behavior to #define or #undef names identical to keywords. If at least one standard library header is included, it's undefined behavior to #define or #undef identifiers identical to names declared in any standard library header.

*This section is incomplete*

*Reason: +other contents of 17.6.4.3 [reserved.names]*

**1.3.1.2 表达式**

An identifier that names a variable, a function, specialization of a concept, (since C++20) or an enumerator can be used as an expression. The expression consisting of just the identifier returns the entity named by the identifier. The value category of the expression is lvalue if the identifier names a function, a variable, or a data member, and prvalue otherwise (e.g. an enumerator is a prvalue expression, a specialization of a concept is a bool prvalue (since C++20)).

Within the body of a non-static member function, each identifier that names a non-static member is implicitly transformed to a class member access expression this->member.

**1.3.1.3 Unqualified identifiers**

Besides suitably declared identifiers, the following can be used in expressions in the same role:

* 函数表达式里的重载运算符，如***operator*+** 或 ***operator* new**；
* a user-defined conversion function name, such as ***operator bool***;
* a user-defined literal operator name, such as ***operator "" \_km***;
* a template name followed by its argument list, such as ***MyTemplate<int>***;
* the character ~ followed by a class name, such as ***~MyClass***;
* the character ~ followed by a decltype specifier, such as ***~decltype(str).***

Together with identifiers they are known as unqualified id-expressions.

**1.3.1.4 Qualified identifiers**

A qualified id-expression is an unqualified id-expression prepended by a scope resolution operator ::, and optionally, a sequence of enumeration, (since C++11)class or namespace names or decltype expressions (since C++11) separated by scope resolution operators. For example, the expression std::string::npos is an expression that names the static member npos in the class string in namespace std. The expression ::tolower names the function tolower in the global namespace. The expression ::std::cout names the global variable cout in namespace std, which is a top-level namespace. The expression boost::signals2::connection names the type connection declared in namespace signals2, which is declared in namespace boost.

The keyword template may appear in qualified identifiers as necessary to disambiguate dependent template names.

See qualified lookup for the details of the name lookup for qualified identifiers.

### **1.3.2 Names**

A name is the use of one of the following to refer to an entity or to a label:

* 标识符；
* 函数表达式里重载运算符的名称 (operator+, operator new);
* 用户自定义的类型转换函数 (operator bool);
* a user-defined literal operator name (operator "" \_km);
* a template name followed by its argument list (MyTemplate<int>).

Every name that denotes an entity is introduced into the program by a declaration. Every name that denotes a label is introduced into the program either by a goto statement or by a labeled statement. A name used in more than one translation unit may refer to the same or different entities, depending on linkage.

When the compiler encounters an unknown name in a program, it associates it with the declaration that introduced the name by means of name lookup, except for the dependent names in template declarations and definitions (for those names, the compiler determines whether they name a type, a template, or some other entity, which may require explicit disambiguation).

## **1.4 Types - Fundamental types**

Objects, references, functions including function template specializations, and expressions have a property called type, which both restricts the operations that are permitted for those entities and provides semantic meaning to the otherwise generic sequences of bits.

### **1.4.1 Type classification**

The C++ type system consists of the following types:

* fundamental types (see also std::is\_fundamental):
* the type void (see also std::is\_void);
* the type std::nullptr\_t (since C++11) (see also std::is\_null\_pointer);
* arithmetic types (see also std::is\_arithmetic):
* floating-point types (float, double, long double) (see also std::is\_floating\_point);
* integral types (see also std::is\_integral):
* the type bool;
* character types:
* narrow character types (char, signed char, unsigned char);
* wide character types (char16\_t, char32\_t, wchar\_t);
* signed integer types (short int, int, long int, long long int);
* unsigned integer types (unsigned short int, unsigned int, unsigned long int, unsigned long long int);
* compound types (see also std::is\_compound):
* reference types (see also std::is\_reference):
  + lvalue reference types (see also std::is\_lvalue\_reference):
    - lvalue reference to object types;
    - lvalue reference to function types;
* rvalue reference types (see also std::is\_rvalue\_reference):
* rvalue reference to object types;
* rvalue reference to function types;
* pointer types (see also std::is\_pointer):
* pointer to object types;
* pointer to function types;
* pointer to member types (see also std::is\_member\_pointer):
* pointer to data member types (see also std::is\_member\_object\_pointer);
* pointer to member function types (see also std::is\_member\_function\_pointer);
* array types (see also std::is\_array);
* function types (see also std::is\_function);
* enumeration types (see also std::is\_enum);
* class types:
* non-union types (see also std::is\_class);
* union types (see also std::is\_union).

For every type other than reference and function, the type system supports three additional cv-qualified versions of that type (const, volatile, and const volatile).

Types are grouped in various categories based on their properties:

* object type is a (possibly cv-qualified) type that is not a function type, not a reference type, and not void type (see also ***std::is\_object***);
* scalar types are (possibly cv-qualified) arithmetic, pointer, pointer to member, enumeration, and std::nullptr\_t types (see also std::is\_scalar);
* trivial types (see also std::is\_trivial), POD types (see also std::is\_pod), literal types (see also std::is\_literal\_type), and other categories listed in the the type traits library or as named type requirements.

### **1.4.2 Type naming**

A name can be declared to refer to a type by means of:

* class declaration;
* enum declaration;
* typedef declaration;
* type alias declaration.

Types that do not have names often need to be referred to in C++ programs; the syntax for that is known as type-id. The syntax of the type-id that names type T is exactly the syntax of a declaration of a variable or function of type T, with the identifier omitted, except that decl-specifier-seq of the declaration grammar is constrained to type-specifier-seq, and that new types may be defined only if the type-id appears on the right-hand side of a non-template type alias declaration.

int\* p; // 指向整数的指针声明

static\_cast<int\*>(p); // type-id是"int\*"

int a[3]; // 三个int型元素的数组

new int[3]; // type-id是"int[3]" (called new-type-id)

int (\*(\*x[2])())[3]; // 2个指向函数的指针元素的数组

// 返回指向3个int元素的数组

new (int (\*(\*[2])())[3]); // type-id是"int (\*(\*[2])())[3]"

void f(int); // 函数的声明，形参为int，返回值类型为void

std::function<void(int)> x = f; // 类型模板参数是type-id "void(int)"

std::function<auto(int) -> void> y = f; // same

std::vector<int> v; // declaration of a vector of int

sizeof(std::vector<int>); // type-id is "std::vector<int>"

struct { int x; } b; // creates a new type and declares an object b of that type

sizeof(struct{ int x; }); // error: cannot define new types in a sizeof expression

using t = struct { int x; }; // creates a new type and declares t as an alias of that type

sizeof(static int); // error: storage class specifiers not part of type-specifier-seq

std::function<inline void(int)> f; // error: neither are function specifiers

The declarator part of the declaration grammar with the name removed is referred to as abstract-declarator.

Type-id may be used in the following situations:

* to specify the target type in cast expressions;
* as arguments to sizeof, alignof, alignas, new, and typeid;
* on the right-hand side of a type alias declaration;
* as the trailing return type of a function declaration;
* as the default argument of a template type parameter;
* as the template argument for a template type parameter;
* in dynamic exception specification.

Type-id can be used with some modifications in the following situations:

* in the parameter list of a function (when the parameter name is omitted), type-id uses decl-specifier-seq instead of type-specifier-seq (in particular, some storage class specifiers are allowed);
* in the name of a user-defined conversion function, the abstract declarator cannot include function or array operators.

This section is incomplete

Reason: 8.2[dcl.ambig.res] if it can be compactly summarized

This section is incomplete

Reason: mention and link to decltype and auto

### **1.4.3 Elaborated type specifier**

Elaborated type specifiers may be used to refer to a previously-declared class name (class, struct, or union) or to a previously-declared enum name even if the name was hidden by a non-type declaration. They may also be used to declare new class names.

See elaborated type specifier for details.

### **1.4.4 Static type**

The type of an expression that results from the compile-time analysis of the program is known as the static type of the expression. The static type does not change while the program is executing.

### **1.4.5 Dynamic type**

If some glvalue expression refers to a polymorphic object, the type of its most derived object is known as the dynamic type.

// given

struct B { virtual ~B() {} }; // 多态类型

struct D: B {}; // 多态类型

D d; // 最底层的对象（most-derived object）

B\* ptr = &d;

// the static type of (\*ptr) is B

// the dynamic type of (\*ptr) is D

For prvalue expressions, the dynamic type is always the same as the static type.

### **1.4.6 Incomplete type**

The following types are incomplete types:

* the type void (possibly cv-qualified);
* class type that has been declared (e.g. by forward declaration) but not defined;
* array of unknown bound;
* array of elements of incomplete type;
* enumeration type from the point of declaration until its underlying type is determined.

Any of the following contexts requires class T to be complete:

* definition or function call to a function with return type T or argument type T;
* definition of an object of type T;
* declaration of a non-static class data member of type T;
* new-expression for an object of type T or an array whose element type is T;
* lvalue-to-rvalue conversion applied to a glvalue of type T;
* an implicit or explicit conversion to type T;
* a standard conversion, dynamic\_cast, or static\_cast to type T\* or T&, except when converting from the null pointer constant or from a pointer to void;
* class member access operator applied to an expression of type T;
* typeid, sizeof, or alignof operator applied to type T;
* arithmetic operator applied to a pointer to T;
* definition of a class with base class T;
* assignment to an lvalue of type T;
* a catch-clause for an exception of type T, T&, or T\*.

(In general, when the size and layout of T must be known.)

If any of these situations occur in a translation unit, the definition of the type must appear in the same translation unit. Otherwise, it is not required.

This section is incomplete

Reason: rules for completing the incomplete types from §3.9[basic.types]/6

1.4.7 Fundamental types

(See also type for type system overview and **the list of type-related utilities** that are provided by the C++ library)

1.4.7.1 void type

void - type with an empty set of values. It is an incomplete type that cannot be completed (consequently, objects of type void are disallowed). There are no arrays of void, nor references to void. However, pointers to void and functions returning type void (procedures in other languages) are permitted.

std::nullptr\_t

Defined in header <cstddef>

typedef decltype(nullptr) nullptr\_t; (since C++11)

std::nullptr\_t is the type of the null pointer literal, nullptr. It is a distinct type that is not itself a pointer type or a pointer to member type.

1.4.7.2 Boolean type

bool - type, capable of holding one of the two values: true or false. The value of sizeof(bool) is implementation defined and might differ from 1.

1.4.7.3 Character types

signed char - type for signed character representation.

unsigned char - type for unsigned character representation. Also used to inspect object representations (raw memory).

char - type for character representation which can be most efficiently processed on the target system (has the same representation and alignment as either signed char or unsigned char, but is always a distinct type). Multibyte characters strings use this type to represent code units. The character types are large enough to represent any UTF-8 code unit (since C++14). The signedness of char depends on the compiler and the target platform: the defaults for ARM and PowerPC are typically unsigned, the defaults for x86 and x64 are typically signed.

wchar\_t - type for wide character representation (see wide strings). Required to be large enough to represent any supported character code point (32 bits on systems that support Unicode. A notable exception is Windows, where wchar\_t is 16 bits and holds UTF-16 code units) It has the same size, signedness, and alignment as one of the integer types, but is a distinct type.

char16\_t - type for UTF-16 character representation, required to be large enough to represent any UTF-16 code unit (16 bits). It has the same size, signedness, and alignment as std::uint\_least16\_t, but is a distinct type. (since C++11)

char32\_t - type for UTF-32 character representation, required to be large enough to represent any UTF-32 code unit (32 bits). It has the same size, signedness, and alignment as std::uint\_least32\_t, but is a distinct type. (since C++11)

1.4.7.3 Integer types

int - basic integer type. The keyword int may be omitted if any of the modifiers listed below are used. If no length modifiers are present, it's guaranteed to have a width of at least 16 bits. However, on 32/64 bit systems it is almost exclusively guaranteed to have width of at least 32 bits (see below).

1.4.7.4 Modifiers

Modifies the integer type. Can be mixed in any order. Only one of each group can be present in type name.

1.4.7.5 Signedness

signed - target type will have signed representation (this is the default if omitted)

unsigned - target type will have unsigned representation

1.4.7.6 Size

short - target type will be optimized for space and will have width of at least 16 bits.

long - target type will have width of at least 32 bits.

long long - target type will have width of at least 64 bits.(since C++11)

Note: as with all type specifiers, any order is permitted: unsigned long long int and long int unsigned long name the same type.

Properties

The following table summarizes all available integer types and their properties:

## 1.5 Object - Scope – Lifetime

## **1.6 Definitions and ODR**

1.7 Name lookup

1.8 Qualified name lookup (qualified - unqualified)

A qualified name is a name that appears on the right hand side of the scope resolution operator :: (see also qualified identifiers). A qualified name may refer to a

* class member (including static and non-static functions, types, templates, etc)
* namespace member (including another namespace)
* enumerator

If there is nothing on the left hand side of the ::, the lookup considers only declarations made in the global namespace scope (or introduced into the global namespace by a using declaration). This makes it possible to refer to such names even if they were hidden by a local declaration:

#include <iostream>

int main() {

struct std{};

std::cout << "fail\n"; // Error: unqualified lookup for 'std' finds the struct

::std::cout << "ok\n"; // OK: ::std finds the namespace std

}

Before name lookup can be performed for the name on the right hand side of ::, lookup must be completed for the name on its left hand side (unless a decltype expression is used, or there is nothing on the left). This lookup, which may be qualified or unqualified, depending on whether there's another :: to the left of that name, considers only namespaces, class types, enumerations, and templates whose specializations are types:

struct A {

static int n;

};

int main() {

int A;

A::n = 42; // OK: unqualified lookup of A to the left of :: ignores the variable

A b; // Error: unqualified lookup of A finds the variable A

}

1.9 Memory model and data races

## **1.10 转译阶段**

编译器处理C++源文件的过程可以看做下面几个阶段：

### **1.10.1 阶段1**

1）源文件的每一个字节都会被映射为基本字符集中的字符。尤其是，依赖于操作系统的行结束符都会被换行符取代。基本字符集包含96个字符：

* 5 种空白字符 (空格， 水平制表符，垂直制表符，换页符，换行符)；
* 10 个数字[0-9]；
* 26 个英文字母的大小写；
* 29 标点符号：\_ { } [ ] # ( ) < > % : ; . ? \* + - / ^ & | ~ ! = , \ " '

2）源文件中，任何不能由基本字符集映射的字符，使用通用字符取代（即使用\转义符号进行转义）或者由其编译器作相应的处理。

3）三字符序列被相应的单字符表达方式替代。（until C++17）

### **1.10.2 阶段2**

1）一旦在某一行的结尾发现反斜杠“\”（后面紧跟换行符），符号“\”和换行符被删掉，把源文件的两行连接成一行。这是一次单程操作；如果一行结束时，后面紧跟两个反斜杠“\”字符和空行，那么是不会把这三行组成一个新行的。如果在这个阶段，出现通用字符（\uXXX），这种行为未被定义。

如果这个阶段之后，非空的源文件没有以换行符结束（不论是本来就没有换行符，或它以反斜杠符号结束），C++11之前没有定义这种行为，C++11之后会添加一个换行符。

### **1.10.3 阶段3**

1） 源文件被解析成注释，空白字符序列（空格，水平制表符，换行符，垂直制表符，和换页符），预处理符号，如下所示：

* 头文件名称，例如 <iostream> 或者 "myfile.h"
* 标识符
* 预处理数字
* 字符或字符串，包括用户自定义的 (C++11之后添加)
* 操作符和标点符号(包括可替换的符号), 如+，<<=，new，<%，##，and（&&）
* 不属于其它种类的单个非空白字符

2） 在阶段1和阶段2里，任何实施过转换的由双引号“”包含的原始字符串都会被恢复。（C++11之后）

3） 注释被一个空格替代

换行符被保留，没有明确说明，非换行符的空白字符序列是否会被整合成一个空格字符。

### **1.10.4 阶段4**

1）预处理程序被执行；

2）递归遍历阶段1到阶段4，用#include指令引入每一个文件；

3）在这个阶段结束时，所有源文件的预处理指令被移除。

### **1.10.5 阶段5**

1）由源文件转换而来的字符文字和字符串文字里的所有字符都被设为可执行字符集（有可能是如UTF-8一样的多字节字符集）

2） 字符文字和非原始字符串文字里的转义序列和通用字符被展开并转换为可执行字符集。如果由通用字符指定的字符不是可执行字符集里的成员，结果是编译器指定，但是保证不是一个null字符（广义上）

注意：这个阶段的转换执行，在某些编译器的实现里，可以由命令行选项进行控制：gcc和clang使用选项-finput-charset 指定源字符集的编码格式；-fexec-charset和-fwide-exec-charset，指定字符串和字符文字的可执行字符集的编码格式（没有编码前缀）。（C++11之后）

### **1.10.6 阶段6**

相邻字符串文字链接为新的字符串文字。

### **1.10.7 阶段7**

编译阶段：每个预处理符号被转换为一个符号。这些符号被从语法上和语义上分析，然后转换为一个翻译单元。

### **1.10.8 阶段8**

每一个翻译单元被检查，找出要求模板实例化的列表，包括哪些被要求显式实例化的模板。找到那些被要求实例化的模板，然后执行，产生实例化单元。

### **1.10.9 阶段9**

翻译单元，实例化单元，库组件满足外部引用的被集成都一个程序镜像，它包括在执行环境里执行时需要的信息。

### **1.10.10 注意**

一些编译器不实现实例化单元（也被称为模板仓库或模板注册表）且在阶段7编译每一个模板实例，然后存储代码到对应的目标文件中，然后链接器在阶段9把这些编译后的目标文件集合为一个可执行文件。

### **1.10.11 参考**

* C++11 standard (ISO/IEC 14882:2011):
* 2.2 Phases of translation [lex.phases]
* C++98 standard (ISO/IEC 14882:1998):
* 2.1 Phases of translation [lex.phases]

## 1.11 The main() function

# **10 Templates**

A template is a C++ entity that defines one of the following:

* a family of classes (class template), which may be nested classes
* a family of functions (function template), which may be member functions
* an alias to a family of types (alias template) (since C++11)
* a family of variables (variable template) (since C++14)
* a concept (constraints and concepts) (since C++20)

## **10.9 Parameter pack**

A template parameter pack is a template parameter that accepts zero or more template arguments (non-types, types, or templates). A function parameter pack is a function parameter that accepts zero or more function arguments.

A template with at least one parameter pack is called a variadic template.