Inline Functions

- When the program executes the function call instructs the CPU to stores the memory address of the instruction following the function call, copies the arguments of the function on the stack and finally transfers control to the specified function. The CPU then executes the function code, stores the function return value in a predefined memory location/register and returns control to the calling function. This can become overhead if the execution time of function is less than the switching time from the caller function to called function (callee). For functions that are large and/or perform complex tasks, the overhead of the function call is usually insignificant compared to the amount of time the function takes to run. However, for small, commonly used functions, the time needed to make the function call is often a lot more than the time needed to actually execute the function's code. This overhead occurs for small functions because execution time of small function is less than the switching time.

- C++ provides an inline functions to reduce the function call overhead. Inline function is a function that is expanded in line when it is called. When the inline function is called whole code of the inline function gets inserted or substituted at the point of inline function call. This substitution is performed by the C++ compiler at compile time. Inline function may increase efficiency if it is small.

- **Syntax:**

```
inline return-type function-name(parameters)
{
   // function code
}
```

- Remember, inlining is only a request to the compiler, not a command. Compiler can ignore the request for inlining. Compiler may not perform inlining in such circumstances like:
  - If a function contains a loop. (for, while, do-while)
  - If a function contains static variables.
  - If a function is recursive.
  - If a function return type is other than void, and the return statement doesn't exist in function body.
  - If a function contains switch or goto statement.

- **Advantages:**
  - Function call overhead doesn't occur.
  - It also saves the overhead of push/pop variables on the stack when function is called.
  - It also saves overhead of a return call from a function.
  - When we inline a function, we may enable compiler to perform context specific optimization on the body of function. Such optimizations are not possible for normal function calls. Other optimizations can be obtained by considering the flows of calling context and the called context.

- Inline function may be useful (if it is small) for embedded systems because inline can yield less code than the function call preamble and return.

- **Disadvantages:**
  - The added variables from the inlined function consumes additional registers, After in-lining function if variables number which are going to use register increases than they may create overhead on register variable resource utilization. This means that when inline function body is substituted at the point of function call, total number of variables used by the function also gets inserted. So the number of register going to be used for the variables will also get increased. So if after function inlining variable numbers increase drastically then it would surely cause an overhead on register utilization.
  - If you use too many inline functions then the size of the binary executable file will be large, because of the duplication of same code.
  - Too much inlining can also reduce your instruction cache hit rate, thus reducing the speed of instruction fetch from that of cache memory to that of primary memory.
  - Inline function may increase compile time overhead if someone changes the code inside the inline function then all the calling location has to be recompiled because compiler would require to replace all the code once again to reflect the changes, otherwise it will continue with old functionality.
  - Inline functions may not be useful for many embedded systems. Because in embedded systems code size is more important than speed.
  - Inline functions might cause thrashing because inlining might increase size of the binary executable file. Thrashing in memory causes performance of computer to degrade.

- **Inline Function and Classes:**
  - It is also possible to define the inline function inside the class. In fact, all the functions defined inside the class are implicitly inline. Thus, all the restrictions of inline functions are also applied here. If we need to explicitly declare inline function in the class then just declare the function inside the class and define it outside the class using inline keyword.
  - **Example:**

```
class S
{
   public:
      inline int square(int s) // redundant use of inline
      {
         // this function is automatically inline
         // function body
      }
};
```

  - The above style is considered as a bad programming style. The best programming style is to just write the prototype of function inside the class and specify it as an inline in the function definition.

- **Example:**

```
class S
{
   public:
      int square(int s); // declare the function
};

inline int S::square(int s) // use inline prefix
{
    // function body
}
```

- **What is wrong with macro?**
  - The preprocessor replace all macro calls directly within the macro code. It is recommended to always use inline function instead of macro. According to Dr. Bjarne Stroustrup the creator of C++ that macros are almost never necessary in C++ and they are error prone. There are some problems with the use of macros in C++. Macro cannot access private members of class. Macros looks like function call but they are actually not.
  - **Example:**

```
#include <iostream>
using namespace std;
class S
{
     int m;
   public:
      #define MAC(S::m)   // error
};
```

  - C++ compiler checks the argument types of inline functions and necessary conversions are performed correctly. Preprocessor macro is not capable for doing this. One other thing is that the macros are managed by preprocessor and inline functions are managed by C++ compiler.

- **Remember:** It is true that all the functions defined inside the class are implicitly inline and C++ compiler will perform inline call of these functions, but C++ compiler cannot perform inlining if the function is virtual. The reason is call to a virtual function is resolved at runtime instead of compile time. Virtual means wait until runtime and inline means during compilation, if the compiler doesn't know which function will be called, how it can perform inlining?

- One other thing to remember is that it is only useful to make the function inline if the time spent during a function call is more compared to the function body execution time.

- An example where inline function has no effect at all:
  inline void show()

```
{
   cout << "value of S = " << S << endl;
}
```

- The above function relatively takes a long time to execute. In general function which performs input output (I/O) operation shouldn't be defined as inline because it spends a considerable amount of time. Technically inlining of show() function is of limited value because the amount of time the I/O statement will take far exceeds the overhead of a function call.

- Last thing to keep in mind that inline functions are the valuable feature of C++. An appropriate use of inline function can provide performance enhancement but if inline functions are used arbitrarily then they can't provide better result. In other words don't expect better performance of program. Don't make every function inline. It is better to keep inline functions as small as possible.