

# Программирование с блокировками

Виды мьютексов

Обертки для мьютексов

# Классы - «обертки» для мьютекса

Специфика: конструкторы классов «оберток» могут принимать параметр, определяющий политику блокировки:

- **defer\_lock\_t**: не захватывать мьютекс (отложить захват), а только «обернуть», чтобы в случае дальнейшего захвата мьютекса в деструкторе был вызван `unlock()`
- **try\_to\_lock\_t**: попытаться захватить мьютекс без блокировки
- **adopt\_lock\_t**: предполагается, что вызывающий поток уже захватил указанный мьютекс => требуется только в деструкторе вызвать `unlock()`

Типы и соответствующие объекты,  
предоставляемые стандартной библиотекой:

```
struct defer_lock_t { };
```

```
struct try_to_lock_t { };
```

```
struct adopt_lock_t { };
```

```
constexpr std::defer_lock_t defer_lock {};
```

```
constexpr std::try_to_lock_t try_to_lock {};
```

```
constexpr std::adopt_lock_t adopt_lock {};
```

# std::lock\_guard – самая простая обертка для мьютекса C++11

```
template< class Mutex > class lock_guard;
```

идиома **RAII** – «захват ресурса есть» инициализация

- в конструкторах:

`lock_guard(mutex_type& m);` – `m.lock()` => для  
нерекурсивного мьютекса???

`lock_guard( mutex_type& m, std::adopt_lock_t t );` -  
“оборачивает” мьютекс без вызова `m.lock()`

- в деструкторе – **unlock()**

# Специфика:

- `lock_guard( const lock_guard& ) = delete;`
- `lock_guard&  
operator=(const lock_guard& ) = delete;`

# Важно!

**//Глобальный мьютекс**  
**std::mutex m;**

```
void threadFunc()
{
    m.lock();
    //...
    //здесь может быть
    сгенерировано исключение!
    m.unlock(); //если
    сгенерировано исключение,
    мьютекс?
}
```

```
void threadFunc()
{
    std::lock_guard<std::mutex>
        l(m);
    //...
    //здесь может быть
    сгенерировано исключение!
    // мьютекс?
}
```

???

```
{  
    std::mutex m;  
    m.lock();  
    std::lock_guard<std::mutex> l(m);  
}  
  
{  
    std::mutex m;  
    std::lock_guard<std::mutex> l(m);  
    m.unlock();  
}
```

# Другие возможности lock\_guard

конструктор

```
lock_guard( mutex_type& m, std::adopt_lock_t t );
```

Специфика:

на момент создания объекта lock\_guard  
вызывающий поток уже **должен** владеть  
мьютексом, так как в деструкторе - ???



# Пример std::adopt\_lock

```
void f(std::mutex& m)
{
    while(m.try_lock() == false)
    {
        //делаем что-то другое (безопасное)
    }
    //сюда попадаем с захваченным мьютексом
    std::lock_guard(m, std::adopt_lock);
    //работаем под защитой
} // ???
```

# Модифицируем A::swap() с помощью std::lock\_guard

```
void A::swap(A& other){  
    if(this==&other) return;  
    std::lock(this->m, other.m);  
  
    std::lock_guard<std::mutex>  
        lock_this(this->m, std::adopt_lock );  
    std::lock_guard<std::mutex>  
        lock_other(other.m, std::adopt_lock );  
    //обмен  
} ???
```

???

```
{  
    std::mutex m;  
    m.lock();  
    std::lock_guard<std::mutex> l(m, std::adopt_lock );  
}  
  
{  
    std::mutex m;  
    std::lock_guard<std::mutex> l(m , std::adopt_lock );  
    m.unlock();  
}
```

# `std::unique_lock`

- гарантирует: в каждый момент времени мьютексом может владеть только один объект **`std::unique_lock`** !
- **=> movable**

Замечание: если на момент вызова `unique_lock& operator=( unique_lock&& other );` у обертки уже был залоченный мьютекс, то он освобождается – `unlock()`

# **std::unique\_lock** – обертка для мьютекса с бОльшей функциональностью по сравнению с **lock\_guard**

отличия от **std::lock\_guard**:

- **std::lock\_guard** заблокирован на всем протяжении своего существования  
**std::unique\_lock** – добавляет **признак** + возможность узнать свободен/захвачен мьютекс - **owns\_lock()**
- **std::lock\_guard** - мьютекс блокируется один раз в конструкторе, освобождается тоже только один раз – в деструкторе  
**std::unique\_lock** – можно блокировать и освобождать посредством **lock()** и **unlock()**, а также **try\_lock()**, **try\_lock\_for()**, **try\_lock\_until**
- **std::lock\_guard** – политика захвата мьютекса может быть только **std::adopt\_lock**  
**std::unique\_lock** – может быть **adopt\_lock**, **defer\_lock** или **try\_to\_lock**

# По сравнению с `lock_guard` **`std::unique_lock` :**

- Занимает больше памяти (для хранения флага)
- Чуть медленнее, чем `lock_guard` (за счет выполнения дополнительных проверок флага)

=> менее эффективен

# Функциональность **std::unique\_lock**:

- `explicit unique_lock( mutex_type& m );` - `m.lock()`
- `unique_lock( unique_lock&& other );` - можно
- `unique_lock( mutex_type& m, std::adopt_lock_t t );` - аналогично `lock_guard`
- `unique_lock( mutex_type& m, std::defer_lock_t t );` - отложить захват мьютекса (если до вызова деструктора мьютекс не будет занят, то `unlock()` не вызывается). Захватить `mutex` можно позже посредством метода `lock()`
- `unique_lock( mutex_type& m, std::try_to_lock_t t );` - неблокирующая попытка захвата мьютекса + установка признака

# Специфика `unique_lock`:

- `mutex_type* release();`
  - не выполняются ни `lock()`, ни `unlock()`
  - возвращается указатель на ассоциированный мьютекс или `nullptr`
  - обертка больше не ассоциируется с мьютексом
- `explicit operator bool() const; - true`, если
  - мьютекс ассоциирован с объектом
  - и мьютекс захвачен
- `mutex_type* mutex() const;`



# Пример:

```
std::mutex m;  
std::unique_lock<std::mutex> ul1;  
if(ul1) ???
```

```
std::unique_lock<std::mutex> ul2(m); //захват  
if(ul2) ???
```

```
ul2.unlock();  
if(ul2) ???  
//или  
ul2.release();  
if(ul2) ???
```

# Пример:

```
std::mutex m;  
std::unique_lock<std::mutex> ul(m, std::defer_lock);  
bool b = ul.owns_lock(); //???
```

```
ul.lock();  
b = ul.owns_lock(); //???
```

# Модифицируем A::swap() с помощью std::unique\_lock

```
void A::swap(A& other){  
    ...  
    //обмен  
}
```

# Гранулярность защищаемого кода

Защищаемая секция должна быть  
как можно **меньше!**

=> ???

```
void A::swap(A& other){  
    std::lock(lock_this, lock_other);  
    std::unique_lock <std::mutex>  
        lock_this(this->m, std::adopt_lock );  
    std::unique_lock <std::mutex>  
        lock_other(other.m, std::adopt_lock );  
    //обмен  
} ???
```

# Дополнительные возможности `unique_lock`:

## Конструкторы:

- `template <class Rep, class Period>`  
**`unique_lock`** (`mutex_type& m`,  
`const chrono::duration<Rep,Period>& rel_time`);  
вызывает `m.try_lock_for(rel_time)`
- `template <class Clock, class Duration>`  
**`unique_lock`** (`mutex_type& m`,  
`const chrono::time_point<Clock,Duration>& abs_time`);  
– `m.try_lock_until(abs_time)`

# std::timed\_mutex

- lock()
- unlock()
- try\_lock()

Добавляет по сравнению с std::mutex методы:

- **try\_lock\_for()** — если мьютекс занят, блокируется до истечения интервала или освобождения мьютекса
- **try\_lock\_until()**

# Пример

```
void threadFunc(std::timed_mutex& m)
{
    bool b = m.try_lock_for(std::chrono::seconds (10));
    if(b){
        //...
        m.unlock();
    }
    else //интервал истек, а мьютекс так и не освободился
}
```



# Обертки для `std::timed_mutex`

- `unique_lock`
- `lock_guard` - ???

# Использование обертки для timed\_mutex

```
std::timed_mutex m;
```

```
void f() {  
    std::unique_lock<std::timed_mutex>  
        lk(m, std::chrono::milliseconds(3)); //ожидает до 3 мс  
    if(lk) {...} //если блокировка получена, обрабатываем данные  
} // ???
```

Можно (имеет смысл) использовать lock\_guard?

# Попытка использовать `lock_guard` в качестве обертки для `timed_mutex`

```
std::lock_guard<std::timed_mutex> ul1(m); //ок,  
но смысла не имеет
```

```
std::lock_guard<std::timed_mutex> ul1(m,  
std::chrono::seconds(10)); //ошибка
```

# `std::recursive_mutex`

- позволяет потоку-владельцу мьютекса многократно вызывать `lock()`
- следствие -> столько же раз нужно вызвать `unlock()`

**Рекомендация:** не использовать рекурсивные мьютексы. Если таковые понадобились, то стоит подумать об ошибках проектирования

???

std::**mutex** m; //обычный мьютекс

```
void f1() {  
    m.lock();  
    ...  
    m.unlock();  
}  
void f2() {  
    m.lock();  
    f1();  
    m.unlock();  
}
```

# Пример std::recursive\_mutex

```
std::recursive_mutex m;
```

```
void f1(){  
    m.lock();  
    ...  
    m.unlock();  
}
```

```
void f2(){  
    m.lock();  
    f1();  
    m.unlock();  
}
```

# std::recursive\_timed\_mutex

Дополнительная функциональность:

- try\_lock()
- try\_lock\_for()
- try\_lock\_until()

# C++17 shared\_mutex

## #include <shared\_mutex>

Не поддерживается в VS15

Часто возникает задача:

- позволить нескольким потокам осуществлять **чтение** (разделение мьютекса) – при этом нельзя только модифицировать
- и только одному – **запись** (владение мьютексом) – при этом нельзя ни читать, ни писать!

=> **rw - мьютексы**



# Специфика:

Для эксклюзивного использования (владение):

`lock()`, `unlock()`, `try_lock()`

Для совместного использования (разделение):

`lock_shared()`, `unlock_shared()`,  
`try_lock_shared()`

# Простой пример:

```
class Shared {  
    int m_a=0;  
    mutable std::shared_mutex m;  
public:  
    int Get()const{ //читать могут все потоки, если никто не пишет  
        m.lock_shared();//если мьютекс заблокирован для записи, то засыпаем  
        int a = m_a;  
        m.unlock_shared();  
        return a;  
    }  
    void Set(int a) {//только один поток может модифицировать  
        m.lock();  
        m_a=a;  
        m.unlock();  
    }  
};
```

# Продолжение простого примера:

```
class Shared {  
    int m_a=0;  
    mutable std::shared_mutex m;  
public:  
    Shared(const Shared& other); ///???  
    ...  
};
```

# Класс-обертка для `shared_mutex` – `std::shared_lock`

- в конструкторе - ???
- в деструкторе - ???

# Обертка для shared\_mutex – **shared\_lock – C++14**

- **конструкторы:**
  - `explicit shared_lock( mutex_type& m ); // m.lock_shared()`
  - в зависимости от параметра `defer_lock`, `try_to_lock`, `adopt_lock`
  - конструктор, принимающий интервал или момент времени
- **деструктор** – если мьютекс захвачен, `unlock_shared()`

# Переписываем пример:

```
class Shared {  
    int m_a=0;  
    mutable std::shared_mutex m;  
public:  
    int Get()const{ //читать могут все потоки  
        std::shared_lock<std::shared_mutex> lock(m);  
        return m_a;  
    }  
    void Set(int a) {//только один поток может модифицировать  
        std::unique_lock<std::shared_mutex> lock(m);  
        m_a=a;  
    }  
};
```

# C++14 shared\_timed\_mutex

## #include <shared\_mutex>

Добавляет функциональность:

Для эксклюзивного использования:

lock(), unlock(), try\_lock(), **try\_lock\_for()**,  
**try\_lock\_until()**

Для общего использования:

lock\_shared(), unlock\_shared(),  
try\_lock\_shared(), **try\_lock\_shared\_for()**,  
**try\_lock\_shared\_until()**

# Пример использования r/w мьютекса

```
class Shared{  
    mutable std::shared_timed_mutex m;  
    Data sharedData;  
public:  
    Data read()const{  
        std::shared_lock< std::shared_timed_mutex > sl(m);  
        return sharedData;  
    }  
    void write(Data newData){ //по значению безопаснее  
        std::lock_guard< std::shared_timed_mutex> sl(m);  
        sharedData = newData;  
    }  
    ...  
};
```



# Продолжение примера: ???

Реализация?

```
Shared& Shared::operator= (const Shared& );
```

# Обертка `std::scoped_lock` – C++17

## `<mutex>`

```
template< class... MutexTypes >
```

```
    class scoped_lock;
```

В конструкторе – гарантированный захват всех перечисленных в качестве параметров мьютексов посредством функции `lock()`

В деструкторе – освобождение (в обратном захвату порядке)

# Проблема - ???

```
std::vector<int> v;  
void myPush(size_t n, int x){ //потоковая функция  
    for(size_t i=0; i<n; i++){v.push_back(x);}  
}  
int main(){  
    std::thread t1(myPush,100,1);  
    std::thread t2(myPush,200,2);  
    t1.join(); t2.join();  
}
```

# Пример `scoped_lock`:

```
std::vector<int> v;  
std::mutex m;  
void myPush(size_t n, int x){ //потоковая функция  
    for(size_t i=0; i<n; i++){  
        std::scoped_lock lock{m};  
        v.push_back(x);  
    }  
}  
int main(){  
    std::thread t1(myPush,100,1);  
    std::thread t2(myPush,200,2);  
    t1.join(); t2.join();  
}
```

Переписываем  
`void A::swap(A& other);`

???

# Рекомендации –блокировки и не потерять эффективность :

- разумно выбирать гранулярность защищаемого кода
- по возможности использовать один мьютекс (a.lock(); b.lock() - гонки)
- если требуется захват нескольких мьютексов, то порядок захвата везде должен быть одинаков, иначе – гонки
- избегать вызова другого пользовательского кода в защищенной секции
- в сложном многопоточном приложении использовать **иерархические** мьютексы (назначать мьютексам разные уровни и позволять захватывать только в порядке понижения уровня).

Локальная память потока

# **C++11 СПЕЦИФИКАТОР КЛАССА ХРАНЕНИЯ THREAD\_LOCAL**

# Спецификаторы класса хранения C++ ???

- До стандарта C++11 ???
- **C++11** - спецификатор класса хранения **thread\_local** (ключевое слово => подключать заголовочные файлы не требуется)



# Время жизни

- все переменные, определенные со спецификатором `thread_local`, обладают потоковым временем жизни – **thread storage duration**
- `thread storage duration` имеет много общего со статическим временем жизни (`static storage duration`). Отличие:
  - статическое время жизни – до завершения программы
  - потоковое время жизни – до завершения потока

Инициализация “блочных” `thread_local` переменных  
(похожа на инициализацию статических переменных)

```
void ThreadFunc(){  
    for(int i=0; i<10; i++)    {  
        thread_local int z=33;  
        z++;  
        ...  
    }  
}
```

- инициализируется один раз
- при первом выполнении в каждом потоке!
- уничтожается при завершении потока

# Замечание:

```
void ThreadFunc()  
{  
    ...  
    [static] thread_local int z=33; //избыточно  
    z++;  
    ...  
}
```

# «Глобальные» thread\_local переменные

## 1.cpp

```
extern thread_local int x;  
int main()  
{  
    // x = ?  
    x++;  
    // x = ?  
    std::thread t1(ThreadFunc);  
    std::thread t2(ThreadFunc);  
    t1.join();    t2.join();  
    // x = ?  
    ...  
}
```

## 2.cpp

```
thread_local int x;  
void ThreadFunc()  
{  
    // x = ?  
    x++;  
    // x = ?  
    ...  
}
```

# Разница thread\_local и static:

**1.cpp**

```
thread_local int x;  
static thread_local int y;
```

**2.cpp**

```
extern thread_local int x;  
extern thread_local int y;
```

```
void f2(){  
    x++; //OK  
    y++; //???  
}
```

# Статические thread\_local переменные класса

```
class A {  
    static thread_local size_t count;  
public:  
    ...  
};
```

# Разница:

<b>A.h</b> <b>class A{</b> <b>static size_t count;</b>  <b>public:</b> ... <b>};</b>	<b>A.h</b> <b>class A{</b> <b>static thread_local size_t count;</b>  <b>public:</b> ... <b>};</b>
<b>A.cpp</b> <b>#include "A.h"</b> <b>size_t A::count;</b> //на этапе компиляции в единственном экземпляре	<b>A.cpp</b> <b>#include "A.h"</b> <b>thread_local size_t A::count;</b> //на этапе выполнения для каждого потока

# Нельзя объявлять со спецификатором `thread_local`:

- параметры функций
- переменные со спецификатором `register`
- не статические переменные класса

```
void f(thread_local int x); //ошибка
```

```
class A {  
    thread_local size_t count; //ошибка  
};  
int f1()  
{  
    thread_local register int y; //ошибка  
}
```



# Специфика `thread_local` объекта

- копия `thread_local` переменной, определенной вне функции, должна быть проинициализирована (сконструирована) к моменту первого использования (на момент запуска потока или непосредственно перед первым использованием – стандарт не оговаривает)
- => может вычисляться во время выполнения
- => не может быть `constexpr`
- неявная инициализация `thread_local` переменных базового типа – «0»
- если при конструировании `thread_local` переменной пользовательского типа генерируется исключение, то вызывается `terminate()`
- передавать адрес другому потоку можно, но НЕ безопасно!

# Пример использования `thread_local`

- при реализации иерархического мьютекса для решения проблемы dead-lock-ов

# Иллюстрация проблемы взаимоблокировки (псевдокод)

Thread A	Thread B
<pre>Lock(x) Lock(y) //Используем X и Y Unlock(y) Unlock(x)</pre>	<pre>Lock(y) Lock(x) //Используем X и Y Unlock(x) Unlock(y)</pre>

# Deadlock возможен, если

потоки блокируют более одного mutex-а:

- два mutex-а А и В защищают два разных ресурса, и, двум потокам, одновременно необходим доступ к этим двум ресурсам
- Блокировка одного mutex-а - атомарна, но блокировка двух mutex-ов - это два отдельных действия => если первый поток заблокирует mutex А, в то время, как второй заблокирует mutex В, оба потока «зависнут», ожидая друг друга.

# Реализация иерархического мьютекса

- В C++11-14-17 отсутствует иерархический мьютекс
- Для его реализации нужно учесть:
  - для каждого потока должен быть индивидуальный «счетчик» уровней мьютексов в данном потоке => **TLS**
  - при этом нужно гарантировать, что в потоке нет двух мьютексов с одинаковым “приоритетом”
  - уровни логично задавать не случайно, а в соответствии с важностью защищаемого участка кода

# Пример реализации иерархического мьютекса:

// hierarch\_mutex.h

```
class hierarch_mutex{
    mutable std::mutex m;
    int prev_level;
    const int this_level;
    static thread_local int thread_cur_level;
public:
    hierarch_mutex(int l) : this_level(l), prev_level(0){}
    void lock();
    void unlock();
};
```

// hierarch\_mutex.cpp

```
thread_local int hierarch_mutex ::thread_cur_level = INT_MAX;
```

# Продолжение примера

```
void hierarch_mutex ::lock() {  
    if(thread_cur_level <= this_level)  
        {throw "error level";}   
    m.lock();  
    prev_level = thread_cur_level ;  
    thread_cur_level = this_level;  
}
```

# Продолжение примера

```
void hierarch_mutex ::unlock() {  
    thread_cur_level = prev_level;  
    m.unlock();  
}
```



# Продолжение примера

```
bool hierarch_mutex ::try_lock()  
    if(thread_cur_level <= this_level)  
        {throw "error level";} //или return false;  
    if(!m.try_lock()){return false;}  
    prev_level = thread_cur_level ;  
    thread_cur_level = this_level;  
    return true;  
}
```

Джеффри Рихтер:

«Не нужно плодить потоки, потому что переключение контекстов - вещь дорогая»

## **THREAD POOL**

# Проблемы нерационального использования потоков

Если для решения часто возникающих небольших заданий для каждого задания запускается отдельный поток:

- поток создается непосредственно при поступлении задания и разрушается по завершении его обработки => порождение и уничтожение потока приводят к **значительным издержкам**
- требуется «**вручную**» задавать/определять количество одновременно работающих потоков
- для некоторых задач требуется организовать досрочное завершение потока (в классе `std::thread` не реализовано)

# Шаблон проектирования «пул потоков»

- позволяет создавать разумное количество потоков
- и многократно их использовать, пока существует пул
- при уничтожении самого пула потоки также удаляются.

В результате экономятся ресурсы, которые до этого шли на создание, переключение и уничтожение потоков.

# Идея пула потоков:

- создаем N (как выбираем N?) потоков и помещаем их в хранилище (пул).
- по мере возникновения новой задачи «достаем» из хранилища готовый поток и отдаем ему задачу на выполнение (при этом не происходит каждый раз создание/уничтожение => **при частых обращениях выигрыш**)
- как только поток отрабатывает отведенную ему задачу - он высвобождается (но продолжает существовать) => можно использовать его для выполнения следующей задачи.

# Простейшая реализация пула потоков

Сущности:

- рабочий поток,
- пул потоков,
- задание
- очередь заданий

В данной реализации не предусмотрены:

- получение результатов выполнения задания
- синхронизация с завершением задания
- генерация исключений заданиями

```
class thread_pool {  
    std::mutex m; //для безопасной многопоточной работы с очередью  
    std::queue<std::function<void()>> tasks; //извне не  
        используется, а в методах обеспечивает корректный доступ  
        к разделяемым данным посредством мьютекса  
    std::vector<std::thread> threads; //совокупность рабочих  
ПОТОКОВ  
  
    void task_thread(); //потокосная функция  
  
public:  
    thread_pool();  
    ~thread_pool();  
    void add_task(<параметры>);  
};
```

```
void thread_pool::task_thread(){  
    while (true){  
        std::function<void()> task;  
        m.lock();  
        if (tasks.empty()) {  
            m.unlock();  
            std::this_thread::yield(); //отдыхаем  
        }else {  
            task = tasks.front();  
            tasks.pop();  
            m.unlock();  
        }  
        if (task) task();  
    }  
}
```



```
thread_pool::thread_pool() {  
    size_t nThreads = <количество потоков в пуле>;  
    //Запускаем потоки:  
    for (size_t i = 0; i < nThreads; i++){  
        threads.emplace_back(&thread_pool::task_thread, this);  
        //может быть сгенерировано исключение => по-хорошему  
        нужно обработать  
    }  
}
```

```
thread_pool::~~thread_pool(){  
    ???  
}
```

# Продолжение

```
void thread_pool:: add_task(параметры)
{
    m.lock();
    tasks.push(<задание_с_полученными_параметрами>);
    m.unlock();
}
```

#include <conditional\_variable>

# УСЛОВНЫЕ ПЕРЕМЕННЫЕ

Если один поток должен дожидаться  
выполнения условия другим потоком:

```
bool ReadyFlag = false;  
std::mutex m;
```

```
void thread1()  
{  
    m.lock();  
    //подготовка данных  
    ReadyFlag = true;  
    m.unlock();  
    //...  
}
```

```
void thread2(){  
    std::unique_lock <std::mutex > l(m);  
    while(!ReadyFlag){  
        l.unlock();  
        std::this_thread::sleep_for(100ms);  
        l.lock();  
    }  
    //обработка данных  
}
```

# Проблемы:

- флаг проверяется независимо от того, произошло событие или нет
- трудно подобрать время ожидания (в среднем поток будет ждать  $\text{duration}/2$ )
- ресурсы на `unlock()/lock()`

=> это неэффективный вариант

=> использование специальных средств для решения таких задач – conditional variables!

# Цель использования условных переменных:

- когда нужно не просто избежать гонки (mutex), а синхронизировать параллельное выполнение задач
- требуется НЕ однократная передача результата от одного потока другому (future - однократная синхронизация), а многократное получение результатов от разных потоков
- => условные переменные, которые можно использовать для многократной синхронизации при обмене данными с несколькими потоками

# Реализации условных переменных

**std::condition\_variable** обеспечивает синхронизацию только посредством мьютекса

**std::condition\_variable\_any** обеспечивает синхронизацию посредством любого «мьютексоподобного» объекта => может потребовать дополнительных расходов по памяти, производительности, ресурсов ОС



# `std::condition_variable`

## `std::condition_variable_any`

это объекты синхронизации,  
предназначенные для блокировки одного  
потока, пока

- он не будет оповещен о наступлении  
некоего события из другого потока
- или не истечет заданный таймаут
- или не произойдет ложное пробуждение  
(spurious wakeup)

# Отправка/получение оповещения

Поток, формирующий условие:

- **notify\_one()** – сообщить одному потоку
- **notify\_all()** – сообщить всем потокам

Поток, ожидающий выполнения условия:

- **wait()** – ожидание выполнения условия
- **wait\_for()**, **wait\_until()**

# Важно!

- в расширение возможностей мьютекса условная переменная позволяет **одному** из ожидающих или **всем** ожидающим потокам продолжить выполнение
- для реализации ожидания на условной переменной все равно необходим мьютекс
- «ложное срабатывание» == возврат управления потоку не всегда означает выполнение условия => необходимо дополнительно проверить выполнение условия

# Ложное (spurious) пробуждение

- когда wait завершается, без участия notify\_one() или notify\_all()

То есть проверка условной переменной может сработать даже если условная переменная не осуществила уведомления => дополнительная проверка!

A.Williams: “Ложные срабатывания невозможно предсказать: с точки зрения пользователя они являются совершенно случайными. Однако они часто происходят, когда библиотека потоков не может гарантировать, что ожидающий поток не пропустит уведомления. Поскольку пропущенное уведомление делает условную переменную бесполезной, библиотека потоков активизирует поток, чтобы не рисковать”

# `void wait( std::unique_lock<std::mutex>& lock );`

- при первом вызове ,если мьютекс занят, освобождает ассоциированный мьютекс (это позволяет другим потокам изменять защищенные данные во время ожидания)
- добавляет этот поток в список ожидающих потоков во внутреннюю структуру данных условной переменной и блокирует вызвавший `wait()` поток
- при вызове для условной переменной `notify_...()` анализируется список и пробуждается один из или все ожидающие потоки, управление возвращается функции `wait()`, в которой снова захватывается мьютекс и выполнение продолжается

```
template< class Predicate >
void wait( std::unique_lock<std::mutex>& lock,
          Predicate pred );
```

- используется для защиты от ложных пробуждений, так как
- проверяет дополнительное условие, и если условие==false, то снова освобождает мьютекс и переводит поток в состояние ожидания
- эквивалентно:

```
while (!pred()) {  
    wait(lock);  
}
```

# Важно!

- на момент вызова `wait()` мьютекс должен быть захвачен! Иначе – неопределенное поведение
- во всех **ожидающих** потоках должен использоваться один и тот же мьютекс
- начиная с C++14 `wait()` не генерирует исключений

# wait\_for(), wait\_until()

возврат управления и блокировка мьютекса происходит:

- если вызвана notify\_...()
- сработало ложное пробуждение
- истек timeout - wait\_for().
- или наступил заданный момент времени - wait\_until()

Причину возврата можно узнать посредством возвращаемого значения - enum class **cv\_status**; (значения: timeout, no\_timeout)



# notify\_one(), notify\_all()

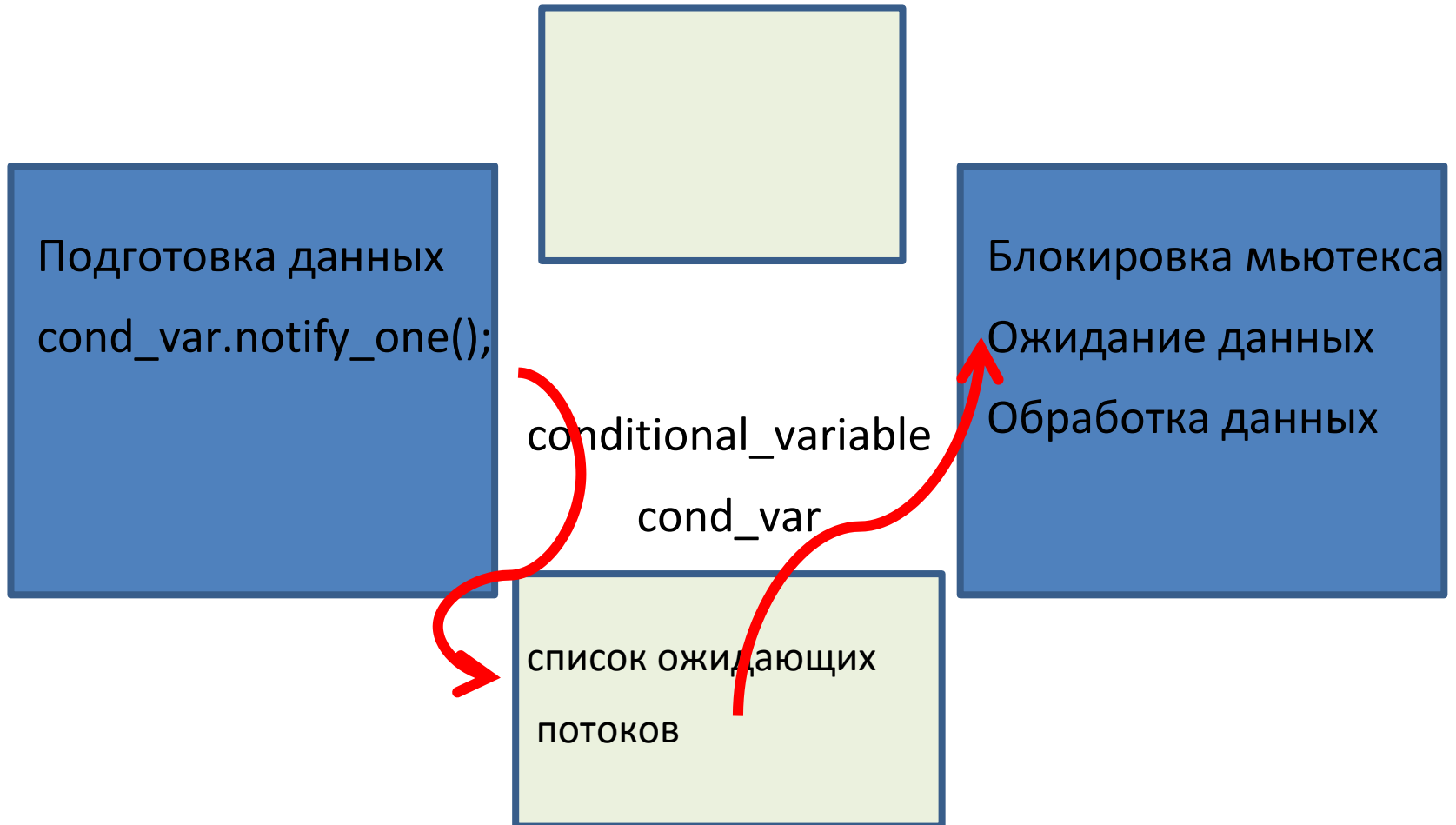
при вызове для условной переменной notify\_...() анализируется список и пробуждается:

- один из ожидающих потоков
- или все ожидающие потоки

=> управление возвращается соответствующей функции wait()

# Иллюстрация

Разделяемые данные



# Ожидание условия с помощью условной переменной в **ОДНОМ** потоке:

```
int shared = 0;  
std::condition_variable cv;
```

```
void thread_Write() {  
    //подготовка данных  
    shared = 33;  
    //уведомление о готовности  
    cv.notify_one();  
}
```

```
void thread_Read () {  
    std::mutex local_m;  
  
    std::unique_lock <std::mutex > l(local_m);  
  
    cv.wait(l);  
  
    //обработка данных  
}
```

Ожидание условия с помощью условной переменной в **одном** потоке. Защита от ложных пробуждений

```
int shared = 0;  
std::condition_variable cv;
```

```
void thread_Write() {  
    //подготовка данных  
    shared = 33;  
    //уведомление о готовности  
    cv.notify_one();  
}
```

```
void thread_Read () {  
    std::mutex local_m;  
  
    std::unique_lock <std::mutex > l(local_m);  
  
    cv.wait(l, []{ return shared == 33; });  
  
    //обработка данных  
}
```

# Ожидание условия с помощью условной переменной в **нескольких** потоках:

```
std::mutex m;  
std::conditional_variable cv;  
int shared = 0;
```

```
void thread_Write1() {  
    //подготовка данных  
    shared = 33; //гонка!  
    //уведомление о готовности  
    cv.notify_all();  
}  
  
void thread_Write2() {  
    //подготовка данных  
    shared = 44; //гонка!  
    //уведомление о готовности  
    cv.notify_all();
```

```
void thread_Read(){  
    std::unique_lock <std::mutex> l(m);  
    cv.wait(l, [] { return shared != 0; });  
    ...  
}
```

# Уведомление с помощью условной переменной в **нескольких** потоках:

```
std::mutex m;  
std::conditional_variable cv;  
int shared = 0;
```

```
void thread_Write() {  
  std::lock_guard<std::mutex> l(m);  
  //подготовка данных  
  shared = 33;  
  //уведомление о готовности  
  cv.notify_all();  
}
```

```
void thread_Read(){  
  std::unique_lock <std::mutex> l(m);  
  cv.wait(l,  
          [] { return shared == 33; });  
  ...  
}
```

А можно так:

```
std::mutex mR,mW;  
std::conditional_variable cv;  
int shared = 0;
```

```
void thread_Write() {  
  std::lock_guard<std::mutex>  
    l(mW);  
  //подготовка данных  
  shared = 33;  
  //уведомление о готовности  
  cv.notify_all();  
}
```

```
void thread_Read(){  
  std::unique_lock <std::mutex>  
  l(mR);  
  cv.wait(l,  
    [] { return shared == 33; });  
  ...  
}
```

# Глобальная функция

## `std::notify_all_at_thread_exit()`

```
void notify_all_at_thread_exit(  
    std::condition_variable& cond,  
    std::unique_lock<std::mutex> lk );
```

Пробуждает все ожидающие потоки ( `cond.wait()` ) **при завершении** текущего потока. Эквивалентно:

```
lk.unlock();  
cond.notify_all();
```

Важно! Действия выполняются гарантированно **после** вызова деструкторов всех потоко-локальных объектов (thread local storage duration)



# Пример std::notify\_all\_at\_thread\_exit

```
std::mutex m;  
std::conditional_variable cv;  
int shared = 0;
```

```
void thread_Write() {  
    std::unique_lock<std::mutex> l(m);  
  
    //использование thread_locals  
  
    //уведомление о готовности выдать только  
    после завершения потока  
    std::notify_all_at_thread_exit(cv,  
        std::move(l));  
    ...  
} //1. деструкторы для thread_locals, 2.  
unlock mutex, 3. notify_all()
```

```
void thread_Read(){  
    std::unique_lock<std::mutex>  
        l(m);  
    cv.wait(l, <проверка усл.>);  
    ...  
}
```

# Пример std::condition\_variable

```
std::vector<int> data;  
std::condition_variable cv;  
std::mutex m;
```

```
void thread_func1() {  
    std::lock_guard<std::mutex> lock(m);  
    data.push_back(<вычисленное_значение>);  
    cv.notify_one();  
}  
  
void thread_func2() {  
    std::unique_lock<std::mutex> lock(m);  
    cv.wait( lock, []() { return !data.empty(); } );  
    std::cout << data.back() << std::endl;  
}
```

`std::condition_variable::native_handle()`

`native_handle_type native_handle();`

В ОС Windows условные переменные на системном уровне поддерживаются объектом синхронизации **CONDITION\_VARIABLE**

Системные функции для работы с CONDITION\_VARIABLE:

`InitializeConditionVariable(),`

`SleepConditionVariableSRW (), WakeConditionVariable()`

```
#include <mutex>
```

```
STD::CALL_ONCE  
STD::ONCE_FLAG
```

# Защита разделяемых данных во время инициализации

Для «дорогостоящих» разделяемых данных (например, получение доступа к БД или установка сетевого соединения) актуальна проблема «отложенной инициализации» (сначала проверка: была ли произведена инициализация, и выполнение инициализирующих действий только при первом вызове).

В однопоточной программе – проблем нет.

В многопоточной программе – защита!

# Совсем неэффективное решение:

```
std::shared_ptr<Resource> resource_ptr;  
std::mutex resource_mutex;
```

```
void f()  
{  
    resource_mutex.lock();  
    if(! resource_ptr)  
    {  
        resource_ptr.reset(new Resource);  
    }  
    resource_mutex.unlock();  
    ...  
}
```

# Решение – Double-Checked Locking

```
std::shared_ptr<Resource> resource_ptr;
```

```
std::mutex resource_mutex;
```

```
void f()
```

```
{
```

```
    if(! resource_ptr) /*1
```

```
    {
```

```
        std::lock_guard<std::mutex> lk(resource_mutex);
```

```
        if(! resource_ptr)
```

```
        {
```

```
            resource_ptr.reset(new Resource); /*2
```

```
        }
```

```
    }
```

```
...
```

```
}
```

# Проблемы Double-Checked Locking

может привести к гонкам (data race), так как:

- чтение без мьютекса **\*1** не синхронизировано
- с защищенной мьютексом записью в другом потоке **\*2**

=> в C++11 включены

- `std::call_once()`
- `std::once_flag`



# std::call\_once()

- std::call\_once создана для того, чтобы защищать общие данные во время инициализации
- это техника, позволяющая вызвать нечто callable один раз, независимо от количества потоков, которые пытаются выполнить этот участок кода
- std::call\_once - быстрый и удобный механизм для создания потокобезопасных singleton-ов

# Специфика – variadic template:

```
template< class Callable, class... Args >  
void call_once( std::once_flag& flag,  
                Callable&& f, Args&&... args );
```

Напоминание: параметры передаются по значению. Если нужно передать ссылку, то `std::ref()` или `std::cref()`

# std::once\_flag

- если один и тот же объект `std::once_flag` передается при разных вызовах `std::call_once()` – функции, гарантируется, что тело указанной функции будет выполнено один раз!

# Эффективность

Утверждается, что издержки использования `std::call_once()` ниже, чем явное использование мьютекса

# Модификация примера с использованием `std::call_once()`

```
std::shared_ptr<Resource> resource_ptr;
```

```
std::once_flag resource_flag;
```

```
void init_resource()
```

```
{ resource_ptr.reset(new Resource); }
```

```
void f()
```

```
{  
    std::call_once(resource_flag, init_resource);
```

```
    ...
```

```
}
```

???

```
std::once_flag flag;
```

```
void do_once(char c){  
    std::call_once(flag, [c]() { std::cout << c << std::endl; });  
}
```

```
int main()  
{  
    std::thread t1(do_once, 'A');  
    std::thread t2(do_once, 'B');  
    std::thread t3(do_once, 'C');  
  
    t1.join();  t2.join();  t3.join();  
}
```

# Пример `std::call_once()` для методов класса

```
class A {  
    std::once_flag fl;  
    void open_connection();  
public:  
    void send(<парам>){  
        std::call_once(fl, &A:: open_connection, this);  
        //отправка данных  
    }  
    <возвр_знач> receive(){  
        std::call_once(fl, &A:: open_connection, this);  
        //прием данных  
    }  
    ...  
};
```

# **ПРОГРАММИРОВАНИЕ НА ОСНОВЕ ЗАДАЧ**



Скотт Мейерс:

«Предпочитайте  
программирование на основе  
задач программированию на  
основе потоков»

# Проблемы подхода на основе ПОТОКОВ:

- Получение формируемого потоком результата?
- Обработка сгенерированного потоком исключения?
- Количество запускаемых в системе потоков – ограниченный ресурс => при превышении – `std::system_error`
- «Превышение подписки» - количество программных потоков (активных и в состоянии готовности) значительно превышает количество аппаратных потоков => частые переключения контекста => время на переключение + перезагрузка кэшей процессора => неэффективно

# Большинство перечисленных проблем

- Решается посредством распараллеливания на основе задач =>
- Переложить проблемы на средства, предоставляемые для этих целей стандартной библиотекой

Когда актуально программирование  
посредством ПОТОКОВ (требуется управление «вручную»):

- Доступ к нижележащим средствам ОС (например, задание приоритетов)
- Выигрыш по эффективности в задачах с фиксированным (заранее известным числом потоков => можно оптимизировать)
- Реализация пулов потоков

Высокоуровневые средства запуска и взаимодействия потоков. Асинхронное программирование. Ожидание одноразовых результатов

```
#include <future>
```

**STD::ASYNC()**

**STD::FUTURE**

**STD::SHARED\_FUTURE**

# Результат работы потоковой функции посредством `std::thread`?

в качестве одного (нескольких) параметра в потоковую функцию можно отправить адрес, по которому функция «положит» сформированный ею результат

Проблемы:

- при ожидании завершения запущенного потока???
- если `detach()`???

# Задача:

Требуется запуск длительной операции и получение результата. При этом результат понадобится позже (а может быть и вообще не понадобится...)

Варианты:

- вызвать синхронно =>
  - ‘+’ - результат гарантированно сформирован после вызова
  - ‘-’ – время, потраченное на вычисление, можно было бы потратить на выполнение какой-нибудь другой полезной работы, не требующей результата

# Проблемы получения результата **detached** потока:

- как узнать о том, что результат сформирован?
- как обеспечить существование объекта, адрес которого передан в поток?



# Асинхронное программирование

стиль программирования, при котором :

- «тяжеловесные» задачи исполняются в detached дочерних (фоновых) потоках,
- и результат выполнения которых может быть получен родительским потоком, когда он того пожелает (при условии, что результат доступен)

# шаблон функции `std::async()`

- позволяет **синхронно** или **асинхронно** запустить задачу (потенциально в отдельном потоке)
- передать ей любое количество параметров (аналогично `std::thread`)
- и получить возвращаемое потоком значение с помощью шаблона `std::future` (когда оно будет сформировано)

# Формы `async()` – C++14

```
template< class Function, class... Args>
std::future<std::result_of_t<std::decay_t<Function>
(std::decay_t<Args>...)>>
    async( Function&& f, Args&&... args );
```

```
template< class Function, class... Args >
std::future<std::result_of_t<std::decay_t<Function>(std::decay_t<Args>...)>>
    async( std::launch policy, Function&& f,
            Args&&... args );
```

# Специфика **std::launch** :

- По умолчанию **std::launch::deferred** | **std::launch::async** реализация решает сама: запускать новый поток или выполнять синхронно
- **std::launch::deferred** – отложить вызов функции до вызова методов `wait()` или `get()` объекта `future` ( *lazy evaluation* )
- **std::launch::async** – запуск функции в отдельном потоке

Для асинхронного взаимодействия и обмена данными thread support library предоставляет:

- **futures** – средства для однократного
  - получения возвращаемых потоками значений
  - и обработки исключений при асинхронном выполнении

## Будущие результаты (future)



`std::future<> & std::shared_future<>`

# Важно!

- Будущие результаты сами по себе не обеспечивают синхронизацию доступа к разделяемым данным
- => если несколько потоков используют один и тот же объект `future`,
  - то они сами должны синхронизировать доступ с помощью мьютекса или других средств синхронизации
  - или нужно позволить нескольким потокам пользоваться одним и тем же объектом **`std::shared_future`** или каждому потоку работать со своей копией - **`std::shared_future`** без дополнительной синхронизации

# Шаблон класса `std::future`:

- формально является **оберткой** для значения любого типа, вычисление или получение которого происходит отложено (то есть в неизвестном будущем)
- фактически поддерживает механизм однократного уведомления, является получателем будущего значения
- сама по себе обертка пассивна, она просто предоставляет доступ к некоторому **разделяемому состоянию**, которое состоит из 2-х частей: собственно интересующие **данные** + **флаг готовности** (как только значение вычислено, устанавливается флаг)
- для многих задач менее ресурсоемко, чем `condition variable` + `mutex`



# Шаблон `std::future`

предоставляет механизм для однократного получения результата асинхронной операции, инициированной: **`std::async()`**, **`std::packaged_task()`**, **`std::promise()`**

- готовность результата посредством методов **`std::future`**:
  - **`wait()`** - блокирует вызвавший поток до получения результата
  - **`wait_for()`** — ждет завершения или истечения timeout-а
  - **`wait_until()`** - ждет завершения или наступления момента
  - **`get()`** — для ожидания вызывает **`wait()`** и возвращает результат

# std::future::get()

- **T get();** //генеральный шаблон (возвращаемое значение формируется посредством std::move()) => повторный вызов == неопределенное поведение
- **T& get();**//для специализации future<T&>
- **void get();** // для специализации future<void>

Важно!

- Любой из вариантов дожидается результата
- Если было сохранено исключение, оно заново генерируется

# std::future::get()

При вызове get() могут быть три ситуации:

- если выполнение функции происходило в отдельном потоке и уже закончилось (или результат уже сформирован) – сразу получаем результат
- если выполнение функции происходит в отдельном потоке и еще не закончилось (или результат еще не сформирован) – поток-получатель блокируется
- если выполнение функции еще не начиналось, то вызов происходит синхронно

Важно: вторичный вызов get() – неопределенное поведение!

# Пример `std::async()`.

## Поведение по умолчанию

```
int sum(const std::vector<int>& v){  
    int res = 0;  
    for (auto i:v) { res += i;}  
    return res;  
}
```

```
int main(){  
    std::vector<int> v = { 1, 2, 3, 4 };  
    std::future<int> f = std::async(sum, std::cref(v));  
    //...какие-то вычисления  
    std::cout<<f.get(); //ждемся окончания,  
                        получаем результат  
}
```

# Пример `std::async()` Явное задание условий вызова:

```
int sum(const std::vector<int>& v){
    int res = 0;
    for (auto i:v) { res += i;}
    return res;
}

int main(){
    std::vector<int> v = { 1, 2, 3, 4 };
    std::future<int> f = std::async(std::launch::async, sum,
                                   std::cref(v));

    //...какие-то вычисления

    std::cout<<f.get(); //дожидаемся окончания,
                        получаем результат
}
```

# Пример `std::async()` Явное задание условий вызова:

```
int sum(const std::vector<int>& v){  
    int res = 0;  
    for (auto i:v) { res += i;}  
    return res;  
}
```

```
int main(){  
    std::vector<int> v = { 1, 2, 3, 4 };  
    std::future<int> f = std::async( std::launch::deferred, sum,  
                           std::cref(v));  
  
    //...какие-то вычисления  
    std::cout<<f.get(); //синхронный вызов + прием возвращаемого  
                        значения  
}
```

Специфика: если функция ничего не  
возвращает

```
void func(void);
```

```
int main(){
```

```
    std::future<void> f = std::async( func);
```

```
    //...какие-то вычисления
```

```
    f.get(); //эквивалентно f.wait();
```

```
}
```

# Пример использования wait\_for()

```
while  
(f.wait_for(1s)==std::future_status::timeout) {  
    // выполняем другую работу, не требующую результата  
}  
//работаем с результатом
```



# std::future::valid()

```
int func(int);  
int main(){  
    int res=0;  
    std::future<int> f1 = std::async( func);  
    bool b = f1.valid(); //true  
    std::future<int> f2 = std::move( f1);  
    b=f1.valid(); //false  
    if(f2.valid()){ res = f2.get();} //true  
    b=f2.valid(); //false  
}
```

# Сохранение исключения в объекте future

Если функция, вызванная посредством `async()`, генерирует исключение,

- это исключение сохраняется в объекте `future` вместо значения (результата)
- вызов `get()` повторно возбуждает **сохраненное исключение** (при этом стандарт не оговаривает: в объекте `future` создается копия исключения или сохраняется ссылка на оригинал)

# Сохранение исключения в объекте future. Пример:

```
double square_root(double x) {  
    if (x<0) { throw std::out_of_range("x<0"); }  
    return sqrt(x);  
}  
int main(){  
    std::future<double> f = std::async(square_root, -1);  
    double y=0;  
    try {  
        y = f.get();  
    }  
    catch (std::out_of_range& e){ std::cout << e.what(); }  
    //или catch (std::exception& e){ std::cout << e.what(); }  
}
```

# Пример более жизненный:

```
int main() {  
    std::vector<std::future<int>> futures;  
    const size_t N = <большое!!!>;  
    futures.reserve(N);  
  
    for(int i = 0; i < N; ++i) {  
        futures.push_back (std::async([](int x){return x*x;},i));  
    }  
    //... занимаемся полезной работой  
    //нужны ВСЕ результаты!  
    for(auto &f : futures) {  
        std::cout << f.get() << std::endl;  
    }  
    ...  
}
```

**STD::SHARED\_FUTURE**

Если требуется в нескольких потоках ожидать  
получения одного и того же результата ???

```
int threadFunc(int x, int y){ return x + y; }
```

```
void thread1(std::future<int>& f){ int res = f.get(); }
```

```
void thread2(std::future<int>& f){ int res = f.get(); }
```

```
int main(){
```

```
    std::future<int> f = std::async(threadFunc, 1, 2);
```

```
    std::thread t1(thread1, std::ref(f));
```

```
    std::thread t2(thread2, std::ref(f));
```

```
    t1.join();  t2.join();
```

```
}
```

# Шаблон класса `std::shared_future`

- позволяет получить результат асинхронной операции, но,
  - в отличие от `future` (вторичный вызов `get()` – неопределенное поведение)
  - **`shared_future`** допускает несколько вызовов `get()` и возвращает один и тот же результат или генерирует одно и то же исключение
- создать и проинициализировать можно:
  - посредством конструкторов `shared_future`
  - или методом `future::share()`

# Формирование std::shared\_future

```
std::future<int> f1 = ...;  
std::shared_future<int> sf1 ( std::move(f1));  
bool b1 = f1.valid(); //???
```

```
std::future<int> f2 = ...;  
std::shared_future<int> sf2= f2.share();  
bool b2 = f2.valid(); //???
```



# Объекты `shared_future` можно:

- копировать
- присваивать
- перемещать

# Модифицируем пример посредством `std::shared_future`

```
int threadFunc(int x, int y){ return x + y; }
```

```
void thread1(std::shared_future<int>& f){ int res = f.get(); ...}
```

```
void thread2(std::shared_future<int>& f){ int res = f.get(); ...}
```

```
int main(){
```

```
    std::future<int> f = std::async(threadFunc, 1, 2);
```

```
    std::shared_future<int> sf(std::move(f));
```

```
    std::thread t1(thread1, sf);
```

```
    std::thread t2(thread2, sf);
```

```
    t1.join();  t2.join();
```

```
}
```