# Vandex Taxi

# Шустрый, простой и современный C++

### Полухин Антон

Antony Polukhin



### Очём поговорим

Зачем вообще об этом говорить?

Алгоритмы

Структуры данных

Move семантика

Оптимизации и микроптимизации

Многопоточность

# Почему и Зачем?!

```
prec ^= (prec >> 31) & 0x7fffffff;
```

```
int prec = fi.i;
prec ^= (prec >> 31) & 0x7fffffff;
```

```
union { float f; int i; } fi;
fi.f = f;
int prec = fi.i;
prec ^= (prec >> 31) & 0x7fffffff;
```

```
unsigned apply(float f) {
   union { float f; int i; } fi;
   fi.f = f;
   int prec = fi.i;
   prec ^= (prec >> 31) & 0x7ffffffff;
```

Что это за дичь!?!?!?

```
struct mapper {
    unsigned apply(float f) {
        union { float f; int i; } fi;
        fi.f = f;
        int prec = fi.i;
        prec ^= (prec >> 31) & 0x7fffffff;
        for (unsigned i = 0; ; ++i) {
            if (prec < ranges[i]) return i;</pre>
    std::vector<int> ranges;
};
```

```
struct mapper {
    unsigned apply(float f) {
        union { float f; int i; } fi;
        fi.f = f;
        int prec = fi.i;
        prec ^= (prec >> 31) & 0x7fffffff;
        for (unsigned i = 0; ; ++i) {
            if (prec < ranges[i]) return i;</pre>
    std::vector<int> ranges;
};
```

### Оно того стоило?

### Int vs Float

```
bool test_ints(int lhs, int rhs) {
    return lhs < rhs;
}

bool test_floats(float lhs, float rhs) {
    return lhs < rhs;
}</pre>
```

### Int vs Float

https://godbolt.org/#

### Int vs Float

```
test_ints(int, int): # @test_ints(int, int)
 cmp edi, esi
 setl al
  ret
test_floats(float, float): # @test_floats(float, float)
 ucomiss xmm1, xmm0
 seta al
  ret
```

### Instruction tables

Lists of instruction latencies, throughputs and micro-operation breakdowns for Intel, AMD and VIA CPUs

By Agner Fog. Technical University of Denmark.

#### Latency:

This is the delay that the instruction generates in a dependency chain. The numbers are minimum values. Cache misses, misalignment, and exceptions may increase the clock counts considerably. Where hyperthreading is enabled, the use of the same execution units in the other thread leads to inferior performance. Denormal numbers, NAN's and infinity do not increase the latency. The time unit used is core clock cycles, not the reference clock cycles given by the time stamp counter.

### Reciprocal throughput:

The average number of core clock cycles per instruction for a series of independent instructions of the same kind in the same thread.

#### Integer instructions

Instruction	Operands	μοps fused domain	µops unfused domain		Latency	Recipro- cal through put	Comments
Move instruc- tions							
MOV	r,i	1	1	p0156		0.25	
MOV	r8/16,r8/16	1	1	p0156	1	0.25	
		_					

#### Haswell

ADD SUB	m,r/i	2	4	2p0156 2p237 p4	6	1
ADC SBB	r,r/i	2	2	2p0156	2	1
ADC SBB	r,m	2	3	2p0156 p23		1
ADC SBB	m,r/i	4	6	3p0156 2p237 p4	7	2
CMP CMP	r,r/i m,r/i	1	1	p0156 p0156 p23		0.25
INC DEC NEG NOT	r	1	1	p0156		0.25
INC DEC NOT	m	3	4	p0156 2p237 p4	6	1
1		^		04500007 4	^	

1 ( ) ( ) ( )	7,111120	•	_	PO P20		"		
VRCPPS	у,у	3	3	2p0 p15	7	2	AVX	
VRCPPS	y,m256	4	4	2p0 p15 p23		2	AVX	
CMPccSS/D CMPccPS/D	x,x / v,v,v	1	1	p1	3	1		
CMPccSS/D								
CMPccPS/D	x,m / v,v,m	2	2	p1 p23		1		
(U)COMISS/D	X,X	1	1	p1		(1)		
(U)COMISS/D	x,m32/64	2	2	p1 p23		1		
MAXSS/D PS/D MINSS/D PS/D	x,x / v,v,v	1	1	p1	3	1		
MAXSS/D PS/D MINSS/D PS/D	x,m / v,v,m	1	2	p1 p23		1		

Разницы практически нет

Разницы практически нет

Нужны инструменты покруче

Intel Architecture Code Analyzer

# Идём глубже

```
#include "iacaMarks.h"
template <class T>
unsigned test(const T* range, T prec) {
    for (unsigned i = 0; ; ++i) {
        if (prec < ranges[i]) return i;</pre>
void testing_cmps(unsigned& a, const int* range, int prec) {
    IACA_START
    a = test(range, prec);
    IACA_END
```

# Идём глубже

```
#include "iacaMarks.h"
template <class T>
unsigned test(const T* range, T prec) {
    for (unsigned i = 0; ; ++i) {
        if (prec < ranges[i]) return i;</pre>
void testing_cmps(unsigned& a, const int* range, int prec) {
    IACA_START
    a = test(range, prec);
    IACA_END
```

```
Throughput Analysis Report

Block Throughput: 2.00 Cycles Throughput Bottleneck: Dependency chains

Total Num Of Uops: 10
```

# Идём глубже

```
#include "iacaMarks.h"
template <class T>
unsigned test(const T* range, T prec) {
    for (unsigned i = 0; ; ++i) {
        if (prec < ranges[i]) return i;</pre>
void testing_cmps(unsigned& a, const float* range, float prec) {
    IACA_START
    a = test(range, prec);
    IACA_END
```

```
Throughput Analysis Report
-----
Block Throughput: 2.52 Cycles Throughput Bottleneck: FrontEnd
```

<...>

Total Num Of Uops: 12

Давайте воспользуемся крутым бенчмарком и определим разницу между стр и ucomis на **практике** 

measure	naive	optim
*** GCC 7		
8	1.00	1.13
64	1.00	1.02
512	1.00	1.21
4096	1.00	0.98
8192	1.00	1.18
*** Clang 5		
8	1.00	1.19
64	1.00	1.02
512	1.00	1.00
4096	1.00	1.18
8192	1.00	0.90
***		

# Стоило ли это так оптимизировать?

## Стоило ли это так оптимизировать?

Мы потратили кучу времени, пытаясь осознать, что написано в коде

Эта функция на критическом пути?

Мы потратили кучу времени, пытаясь осознать, что написано в коде

Эта функция на критическом пути?

Выиграли 20% производительности в **лучшем** случае

Профилирование всего приложения спец инструментами

Профилирование всего приложения спец инструментами

Знаете и без профилирования, что тормозит?

Профилирование всего приложения спец инструментами

Знаете и без профилирования, что тормозит?

Замените тело функции на пустое и перепроверьте!

Так что, можно весь код писать тяпляп, хорошо оптимизируя только несколько основных функций? Нельзя! Кеши, оптимизаторы, рост данных...

#### Не пессимизируем

```
struct mapper {
    unsigned apply(float f) {
        union { float f; int i; } fi;
        fi.f = f;
        int prec = fi.i;
        prec ^= (prec >> 31) & 0x7fffffff;
        for (unsigned i = 0; ; ++i) {
            if (prec < ranges[i]) return i;</pre>
    std::vector<int> ranges;
};
```

#### Не пессимизируем

```
struct mapper {
    unsigned apply(float f) {
        union { float f; int i; } fi;
        fi.f = f;
        int prec = fi.i;
        prec ^= (prec >> 31) & 0x7fffffff;
        for (unsigned i = 0; ; ++i) {
            if (prec < ranges[i]) return i;</pre>
    std::vector<int> ranges;
```

#### Правильное решение

```
struct mapper_fixed {
   unsigned apply(float f) {
        const auto it = std::lower_bound(ranges.cbegin(), ranges.cend(), f);
        return it - ranges.cbegin();
    std::vector<float> ranges;
};
```

Какие выводы мы сделали? Что запомнили?

#### Выводы

Перед микрооптимизациями нужно оптимизировать алгоритм

Перед любыми оптимизациями нужна профилировка

Надо писать комментарии к микрооптимизированному коду

Не надо пессимизировать

Прежде чем мы перейдём к алгоритмам...

# Оставим за бортом

ІО (сеть, диск, устройства со своей памятью и т.д.)

Архитектура

А стоит ли делать маленькие и понятные микрооптимизации?

#### Давайте распрощаемся с иллюзиями

```
https://godbolt.org/#
unsigned foo(unsigned i)
  return i * 2;
                                 // mul или imul - умножение
  return i << 1;
                                 // shl или + - сдвиг влево
  return i / 64;
                                 // div
                                               - деление
  return i / 11;
                                 // shr
                                              – сдвиг вправо
   return i % 2;
                                  // add и sub
                                               - сложение и вычитание
   return i * i * i * i * i;
```

## Оговорки

https://godbolt.org/z/H-OTA-

https://godbolt.org/z/4MW8LL

# Алгоритмы

"O" большое — время работы алгоритма/функции в зависимости от количества входных элементов N

"О" большое — время работы алгоритма/функции в зависимости от количества входных элементов N

for 
$$(size_t i = 0; i < N; ++i) => O(N)$$

"О" большое — время работы алгоритма/функции в зависимости от количества входных элементов N

N	N*log(N)	N*N
2	2	4
4	8	16

N	N*log(N)	N*N
2	2	4
4	8	16
8	24	64
16	64	256
32	160	1,024
64	384	4,096
128	896	16,384
256	2,048	65,536
512	4,608	262,144
1,024	10,240	1,048,576

## std::sort(beg, end)

Сортирует диапазон



# it std::lower\_bound(beg, end, value)

Бинарный поиск в сортированном диапазоне

Позиция элемента value или элемент сразу след за ним

Найти позицию для вставки value в сортированный диапазон

it std::lower\_bound(beg, end, 5)

0	1	2	3	4	5	6	7	8	9	0
0	1	2	3	4	4	6	7	8	9	

## pair std::equal\_range(beg, end, value)

Бинарный поиск диапазона в сортированном диапазоне

pair std::equal\_range(beg, end, 5)

0	1	2	3	4	5	6	7	8	9	
0	1	2	3	4	4	6	7	8	9	
0	1	2	5	5	5	7	7	8	9	

std::sort =>  $O(N*log_2(N))$ 

```
std::sort => O(N*log_2(N))
std::stable_sort => O(N*log_2^2(N))
```

```
=> O(N^*\log_2(N))
std::sort
                       => O(N*log_2(N))
std::stable sort
std::minmax_element => O(N)
std::partition
                       => O(N)
std::nth_element
                  => \sim O(N)
std::partial_sort
                       => O(N*log<sub>2</sub>(S))
```

# result std::partition(beg, end, pred)

Слева от result будет true, справа false:

```
[beg, result) == true
```

4031295876

# std::nth\_element(beg, mid, end)

Выставить значение по итератору mid так чтобы:

Если отсортировать [beg, end) то значение mid не изменится

Слева от mid — значения *меньшие* или *равные* mid

Справа от mid - значения большие или равные mid

4031259876

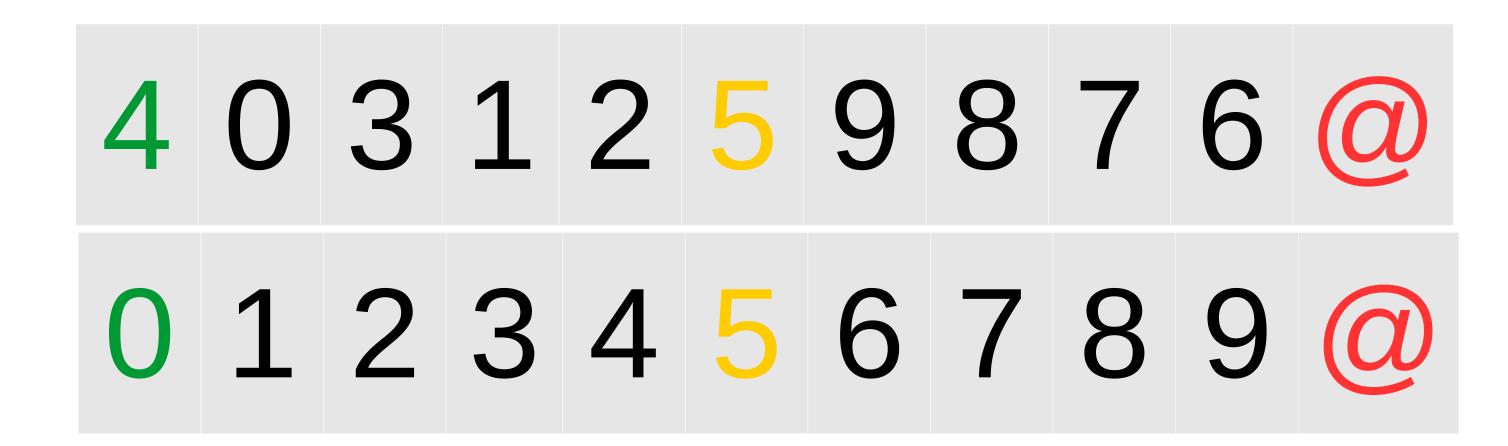
# std::nth\_element(beg, mid, end)

Выставить значение по итератору mid так чтобы:

Если отсортировать [beg, end) то значение mid не изменится

Слева от mid — значения *меньшие* или *равные* mid

Справа от mid - значения большие или равные mid



#### std::nth\_element

Найти 5 людей с наименьшим балансом std::nth\_element(v.begin(), v.begin() + 4, v.end());

Найти 5 людей с наибольшим балансом std::nth\_element(v.begin(), v.begin() + 4, v.end(), std::greater<>{});

Найти 1001 позвонившего std::nth\_element(v.begin(), v.begin() + 1000, v.end());

## std::partial\_sort(beg, mid, end)

Выставить значение по итератору mid так чтобы:

[beg, mid) не изменятся, если отсортировать [beg, end)

[beg, mid) - отсортированы

0 1 2 3 4 9 5 8 7 6 @

## std::partial\_sort

Pacпределить 5 призовых мест по наименьшему кол-ву штрафных баллов std::partial\_sort(v.begin(), v.begin() + 5, v.end());

Покарать 5 школьников, пришедших последними на урок std::partial\_sort(v.begin(), v.begin() + 5, v.end(), std::greater<>{});

# std::minmax\_element

Найти самого бедного и самого богатого клиента банка auto mm = std::minmax\_element(v.begin(), v.end()); std::cout << \*mm.first << ' ' << \*mm.second << '\n';

# К компилятору!

https://github.com/apolukhin

# Контейнеры

"O" большое — время работы алгоритма/функции в зависимости от количества входных элементов N

# Знай микро-врага в лицо!

ithare.com	Operation Cost in CPU Cycles	10°	10¹	10 <sup>2</sup>	10 <sup>3</sup>	10⁴	<b>10</b> <sup>5</sup>	10 <sup>6</sup>
"Simple" r	egister-register op (ADD,OR,etc.)	<1						
	Memory write	~1						
	Bypass delay: switch between							
	integer and floating-point units	0-3						
	"Right" branch of "if"	1-2						
	Floating-point/vector addition	1-3						
N	Multiplication (integer/float/vector)	1-7						
	Return error and check	1-7						
	L1 read		3-4					
	TLB miss		7-21					
	L2 read		10-12					
"Wrong" bra	anch of "if" (branch misprediction)		10-20					
	Floating-point division		10-40					
	128-bit vector division		10-70					
	Atomics/CAS		15-30					
	C function direct call		15-30					
	Integer division		15-40					
	C function indirect call		20-50					
	C++ virtual function call		30-6	60				
	L3 read		30-7	70				
	Main RAM read			100-150				
NUI	MA: different-socket atomics/CAS			400.000				
	(guesstimate)			100-300				
	NUMA: different-socket L3 read			100-300				
Allocation	+deallocation pair (small objects)			200-50	00			
NUMA	: different-socket main RAM read			300	-500			
	Kernel call				1000-1500			
Th	read context switch (direct costs)				2000			
	C++ Exception thrown+caught				5000	-10000		
T	hread context switch (total costs,					10000 - 1	million	
	including cache invalidation)					10000 - 1		

## Знай микро-врага в лицо!

200-500		
300-500		
1000-1	500	
200	0	
	000-10000	
	40000 4 '11'	
	10000 - 1 million	
	300-500 1000-15 200	

# Знай нано-врага в лицо!

L2 read	10-12
"Wrong" branch of "if" (branch misprediction)	10-20
Floating-point division	10-40
128-bit vector division	10-70
Atomics/CAS	15-30
C function direct call	15-30
Integer division	15-40
C function indirect call	20-50
C++ virtual function call	30-60
L3 read	30-70
Main RAM read	100-150

# Знай нано-врага в лицо!

L2 read 💢	10-12
"Wrong" branch of "if" (branch misprediction) X	10-20
Floating-point division X	10-40
128-bit vector division 💢	10-70
Atomics/CAS X	15-30
C function direct call X	15-30
Integer division 💢	15-40
C function indirect call X	20-50
C++ virtual function call 💢	30-60
L3 read	30-70
Main RAM read	100-150

#### Кеш линия

В одной кеш линии х86 — 64 Байта

В одну кеш линию помещается 16 int

# Знай нано-врага в лицо!

L2 read 💢	10-12
"Wrong" branch of "if" (branch misprediction) X	10-20
Floating-point division X	10-40
128-bit vector division 💢	10-70
Atomics/CAS X	15-30
C function direct call X	15-30
Integer division 💢	15-40
C function indirect call X	20-50
C++ virtual function call 💢	30-60
L3 read	30-70
Main RAM read	100-150

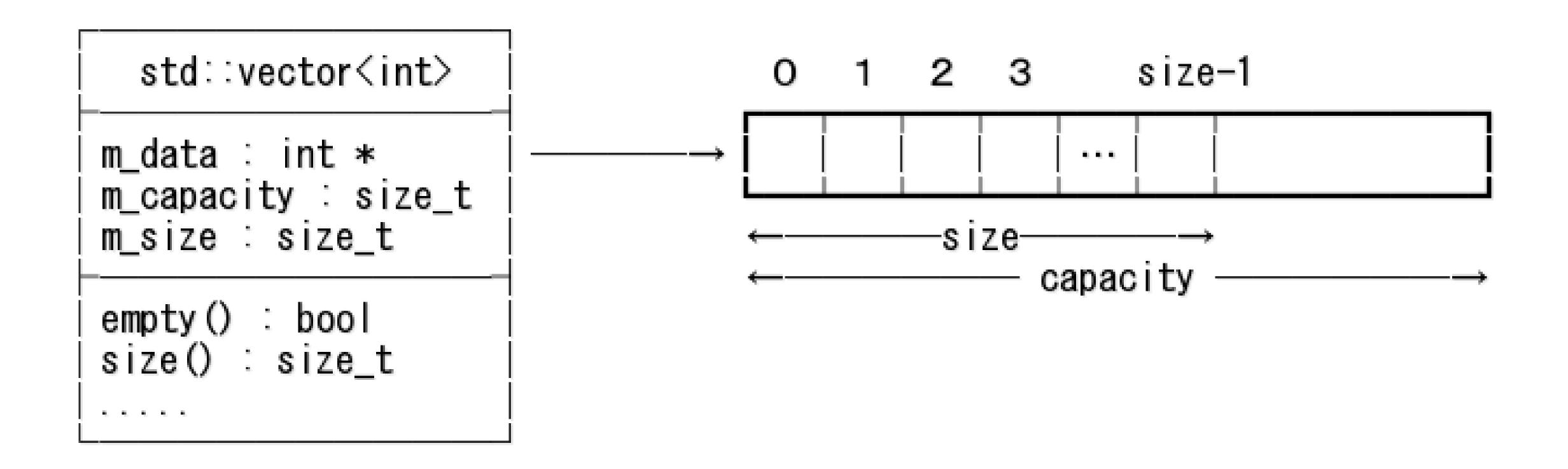
# Все контейнеры, для хранения данных (не ассоциативные)

# Все контейнеры, для хранения данных

std::vector

# container\_1.hpp

#### std::vector



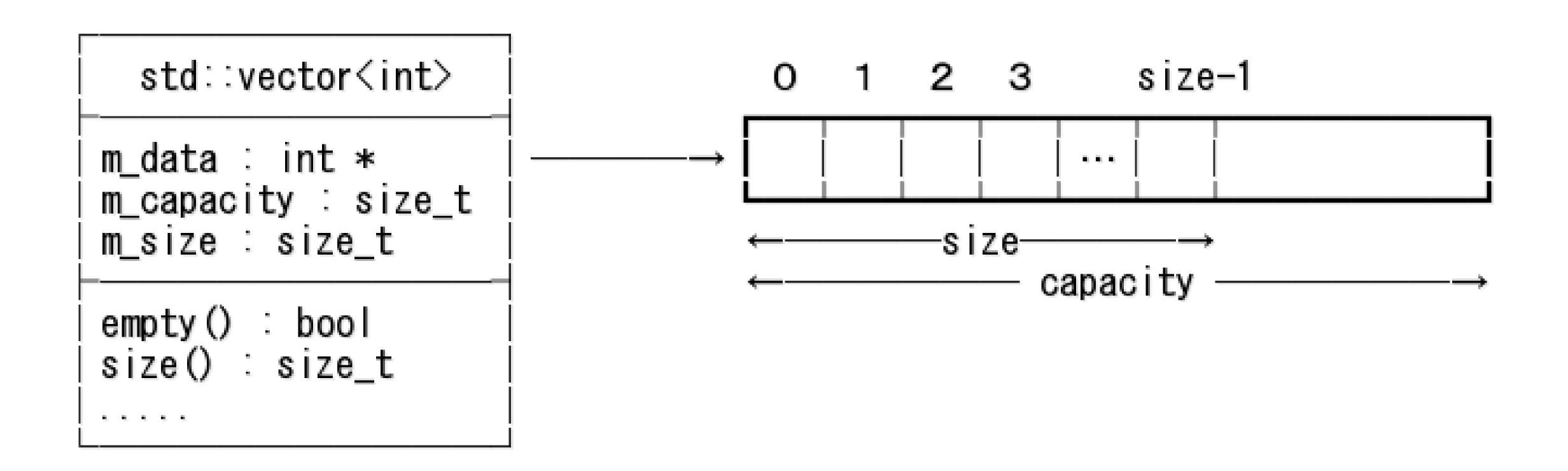
# container\_2.hpp

#### Баги Фишки std::vector

#### Фишки

Move конструкторы и noexcept

#### Move construction & noexcept



#### Фишки

Move конструкторы и noexcept

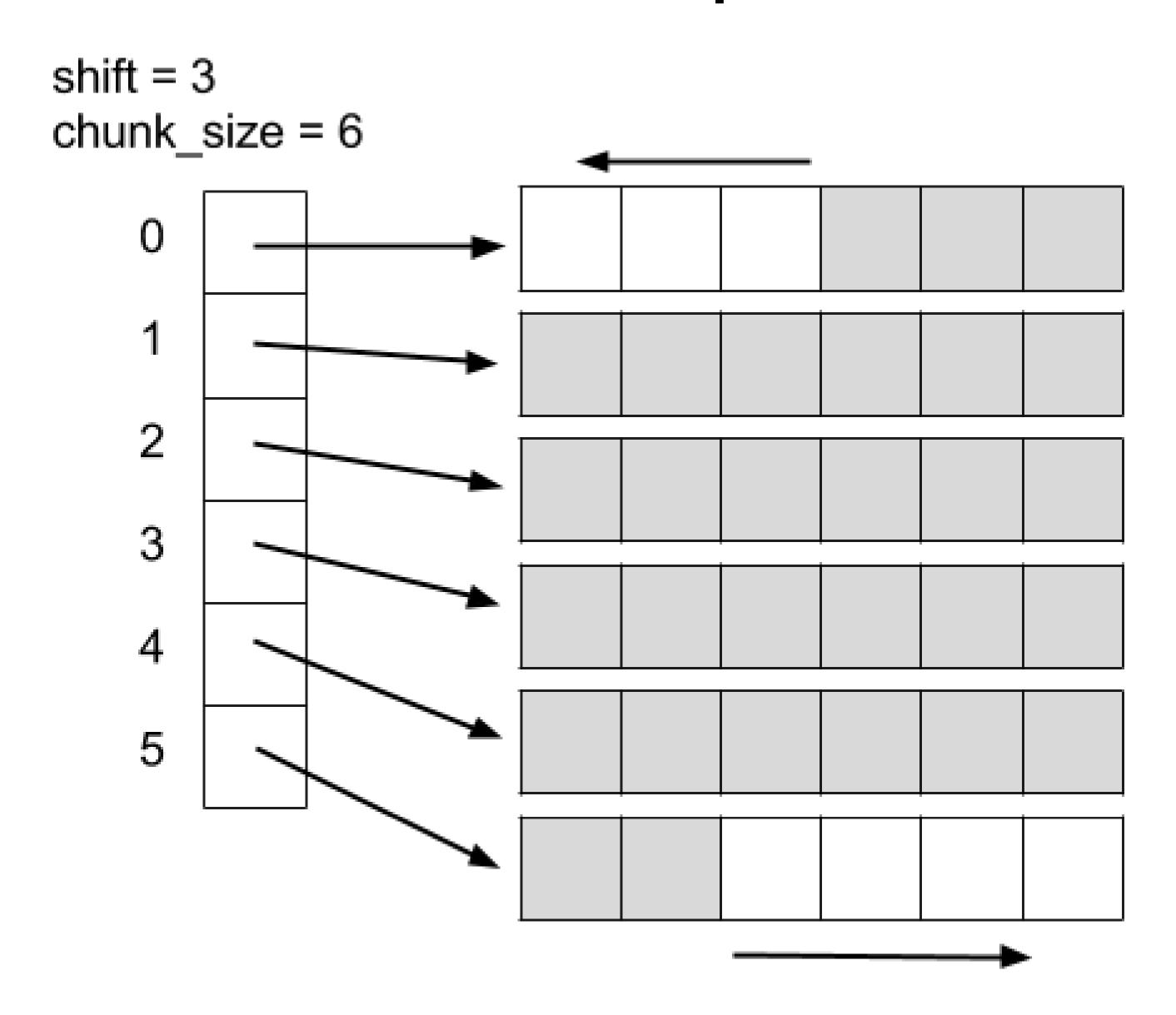
TriviallyCopyable

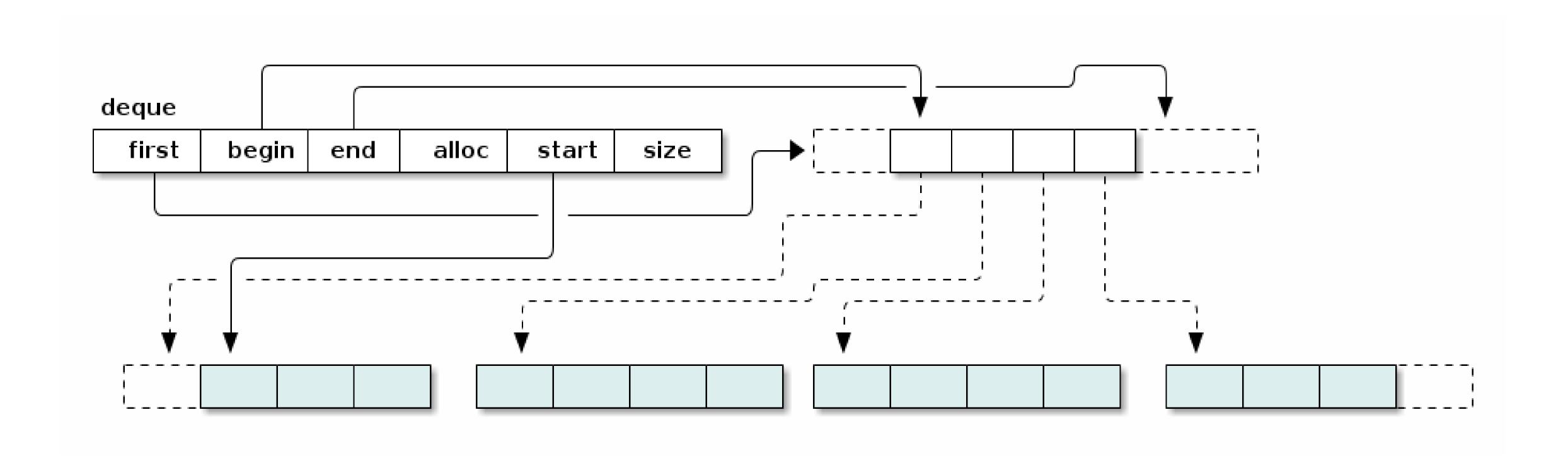
std::vector::reserve

std::vector::at

std::vector<bool>

container\_[3-5].hpp





https://godbolt.org/g/4cJ5HF

```
https://godbolt.org/g/4cJ5HF
```

https://gcc.gnu.org/bugzilla/show\_bug.cgi?id=81461

# Мы уже пробовали использовать short?

container\_[1-2].hpp

## Особые случаи, когда vector не очень

Огромные объёмы данных

Огромные объёмы данных

Многопоточность и особая нагрузка

Огромные объёмы данных

Многопоточность и особая нагрузка

Ассоциативные контейнеры (иногда)

Огромные объёмы данных

Многопоточность и особая нагрузка

Ассоциативные контейнеры (иногда)

std::string

Огромные объёмы данных

Многопоточность и особая нагрузка

Ассоциативные контейнеры (иногда)

std::string

std::array

Огромные объёмы данных

Многопоточность и особая нагрузка

Ассоциативные контейнеры (иногда)

std::string

std::array

Использование сторонних библиотек со стековыми векторами

boost::container::small\_vector

container\_[1-5].hpp

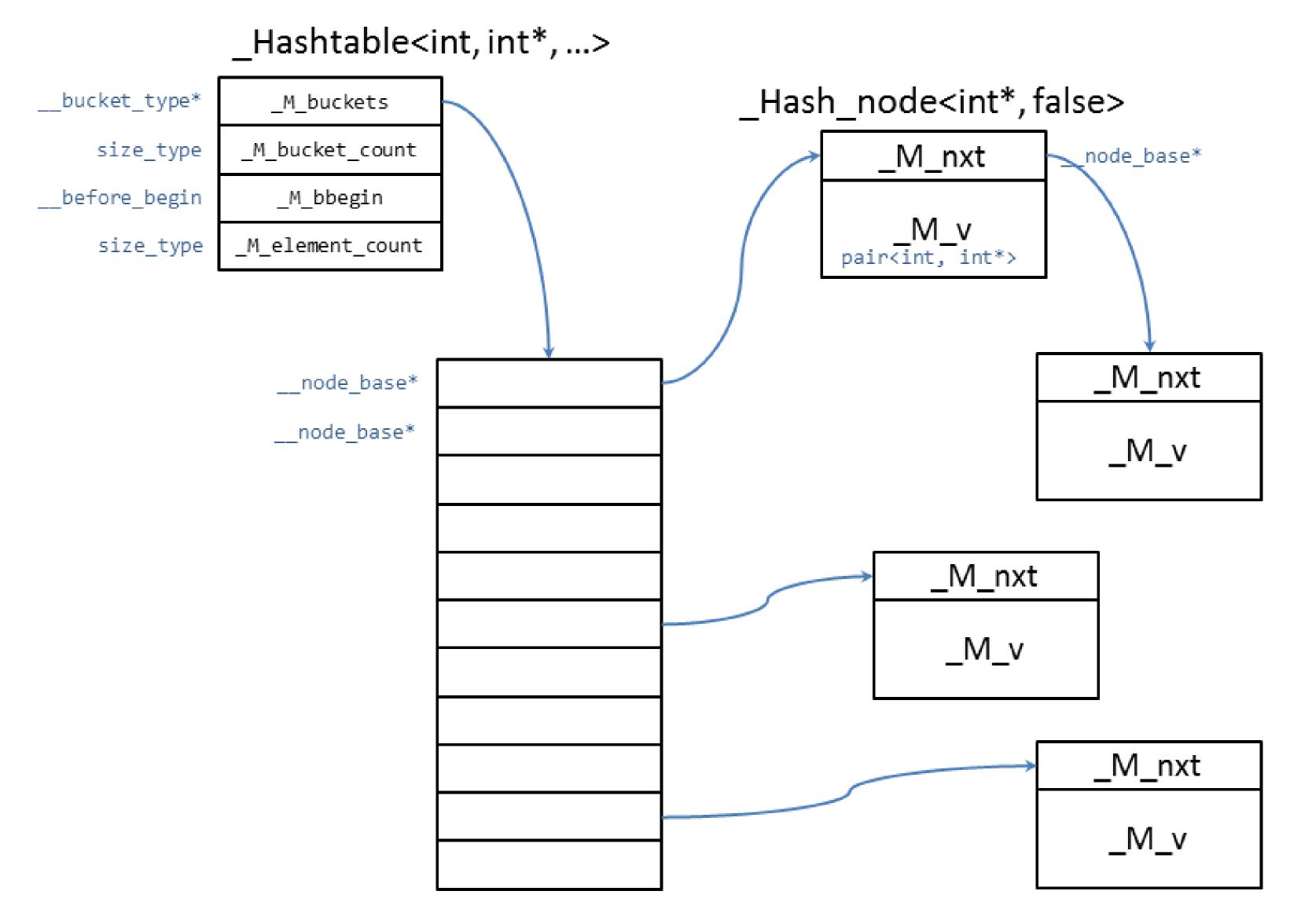
# Ассоциативные контейнеры

## Все ассоциативные контейнеры

#### Все ассоциативные контейнеры

std::unordered\_set / std::unordered\_map

#### std::unordered\_map<int, int\*>



#### Внимание

std::unordered\_\* контейнеры инвалидируют итераторы при вставке! Лучше использовать ссылки/указатели на элементы

He нужен value? Используйте unordered\_set<key>

container\_6.hpp

Нужно хранить данные именно упорядоченно

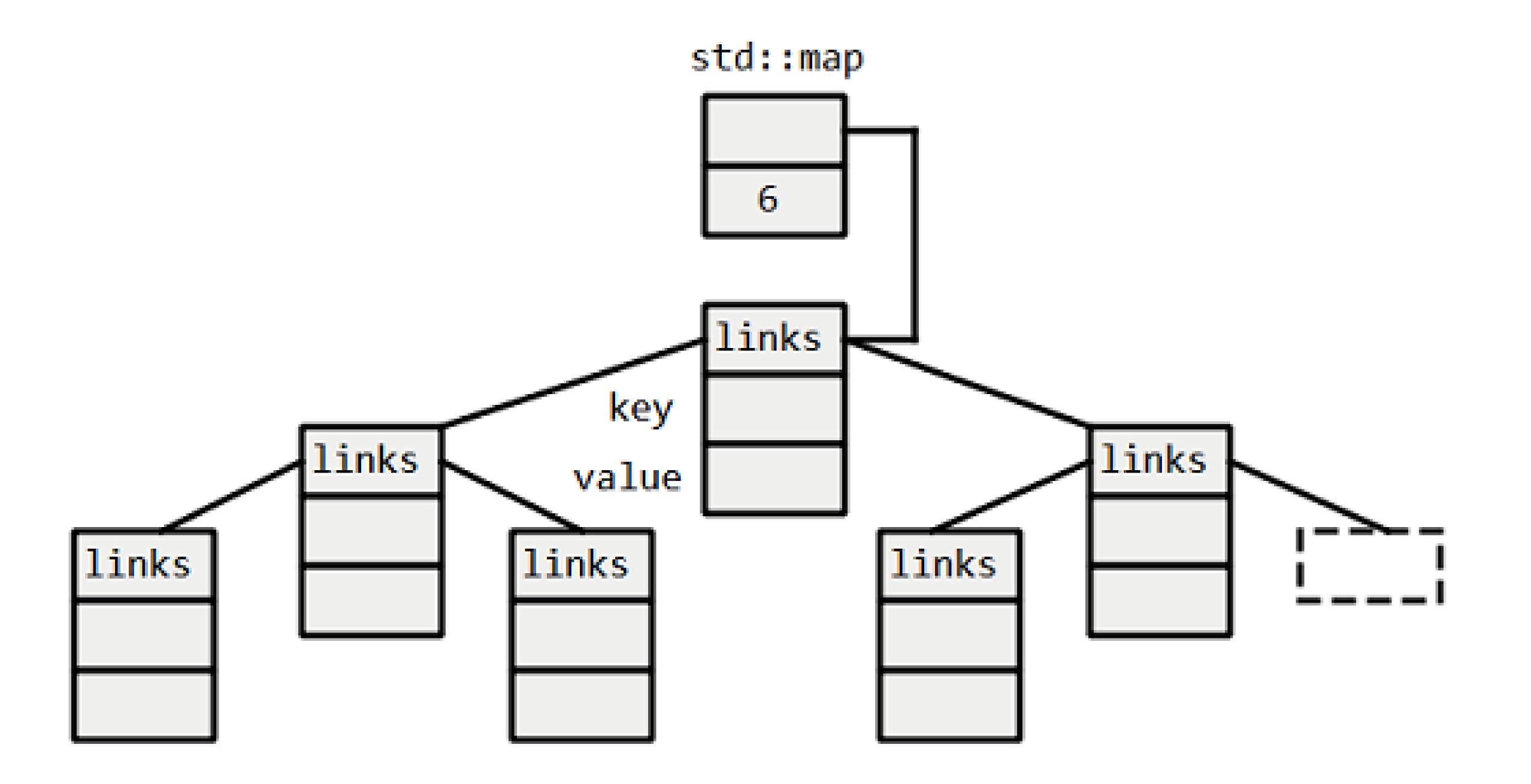
Нужно хранить данные именно упорядоченно

Малые объёмы данных

Нужно хранить данные именно упорядоченно

Малые объёмы данных

Использование сторонних библиотек с "плоскими" ассоциативными контейнерами



flat\_set / flat\_map

container\_6.hpp

#### Move семантика

#### Rvalue & Ivalue

```
void examples() {
    std::string s0{"Hello word. Nice to be here!"};
                                        // Конструктор копирования для s1
    std::string s1 = s0;
    std::string s2 = std::move(s0); // Koнструктор перемещения для s2,
                                        // s0 становится пустым
    s0 = s1;
                                        // Оператор присваивания для s0
   s2 = std::move(s0);
                                        // Оператор перемещения для s2
```

#### Rvalue & Ivalue args

# Forwarding reference

```
template <class T>
void foo(T&& value) { // forwarding reference - любая ссылка (const&, &, && и т. д.)

std::string s = std::forward<T>(value); // превращает бывшую && в настоящую

// rvalue ТОЛЬКО в случае forwarding reference
}
```

#### Move элеметнов из контейнера

```
void example_appending_data(std::vector<std::string>&& data) {
    some_vector.insert(
        some_vector.end(),
        data.begin(),
        data.end()
    );
}
```

#### Move элеметнов из контейнера

```
#include <iterator>
void example_move_iterator(std::vector<std::string>&& data) {
    some_vector.insert(
        some_vector.end(),
        std::make_move_iterator(data.begin()),
        std::make_move_iterator(data.end())
```

#### Для производительности:

std::move объект если вы его больше не используете

И если вы его

не возвращаете из функции return str;

не создали прямо во время вызова функции rvalue(std::string{"Hello"})

**не** возвращаете его из функции rvalue(foo())

#### Move семантика

```
v_new.push_back(std::move(v.back()); v.pop_back();
v_new.front() = std::move(v[10]);
v_new.insert(
   v new.end(),
   std::make_move_iterator(v.begin()),
   std::make_move_iterator(v.end())
```

# container\_7.hpp

# Не пессимизируйте

std::vector

std::vector

std::vector::reserve()

```
std::vector
std::vector::reserve()
++it;
```

```
std::vector
std::vector::reserve()
++it;
for (auto& v: container)
std::unordered_map
```

```
std::vector
std::vector::reserve()
++it;
for (auto& v: container)
std::unordered_map
<algorithm>
```

```
std::vector
std::vector::reserve()
++it;
for (auto& v: container)
std::unordered_map
<algorithm>
??? std::move & std::make_move_iterator ???
```

```
std::vector
std::vector::reserve()
++it;
for (auto& v: container)
std::unordered_map
<algorithm>
??? std::move & std::make_move_iterator ???
string_view
```

```
std::vector
std::vector::reserve()
++it;
for (auto& v: container)
std::unordered_map
<algorithm>
??? std::move & std::make_move_iterator ???
string_view
Пишите выражения просто и используйте целые числа если float вам не нужен
```

#### Многопоточность

# Доступ (неправильный) к общим переменным

```
int shared_i = 0;
void do_inc() {
    for (std::size_t i = 0; i < 500000; ++i) {</pre>
        // ...
        const int i_snapshot = ++shared_i;
void do_dec() {
    for (std::size_t i = 0; i < 5000000; ++i) { const int i_snapshot = --shared_i;</pre>
```

# Доступ (неправильный) к общим переменным

```
int shared_i = 0;
void do_inc();
void do_dec();
void run() {
    std::thread t1(&do_inc);
    do_dec();
    t1.join();
    // assert(shared_i == 0); // Oops!
    std::cout << shared_i << std::endl;</pre>
```

# Доступ к общим переменным

```
#include <mutex>
int shared_i = 0;
std::mutex mutex_i;
void do_inc() {
    for (std::size_t i = 0; i < 500000; ++i) {</pre>
        // ...
            std::lock_guard<std::mutex> lock(mutex_i);
            const int i_snapshot = ++shared_i;
```

#### Правильный и быстрый доступ

```
#include <atomic>
std::atomic<int> shared_i{0};
void do_inc() {
    for (std::size_t i = 0; i < 500000; ++i) {</pre>
        // ...
        const int i_snapshot = ++shared_i;
        // ...
```

# Модель памяти

```
std::atomic<int> i{0};
int a, b, c, d;
```

```
std::atomic<int> i{0};
int a, b, c, d;
void dependency() {
    a = 0;
    b = a + 1;
    c = b + a;
```

```
std::atomic<int> i{0};
int a = 10, b = 11, c = 12;
void dependency() {
     a = 0; // Для других потоков \uparrow \uparrow \downarrow \downarrow
     b = a + 1; // Для других потоков \uparrow \uparrow \downarrow \downarrow
     c = b + a; // Для других потоков \uparrow \uparrow \downarrow \downarrow
```

```
std::atomic<int> i{0};
int a, b, c, d;
void seq_cst() {
    /* \uparrow \uparrow */ a = 0;
    i.fetch_add(1, std::memory_order_seq_cst);
    /* ↑ до seq_cst, ↓ до seq_cst */ b = 0;
    /* ↑ до seq_cst, ↓ до seq_cst */ c = 0;
    i.fetch_sub(1, std::memory_order_seq_cst);
```

```
std::memory_order_acquire / std::memory_order_release
std::memory_order_relaxed
```

## Очереди

## conditional\_variable

```
#include <vector>
#include <condition_variable>
class mt_stack {
    std::vector<int>
                            data_;
                            data_mutex_;
    std::mutex
    std::condition_variable cond_;
public:
    void push(int data);
    int pop();
```

### conditional\_variable

```
void mt_stack::push(int data) {
    std::unique_lock<std::mutex> lock(data_mutex_);
    data_.push_back(data);
    lock.unlock();

cond_.notify_one();
}
```

### conditional\_variable

```
int mt_stack::pop() {
    std::unique_lock<std::mutex> lock(data_mutex_);
   while (data_.empty()) {
        cond_.wait(lock);
    int ret = data_.back();
   data_.pop_back();
    return ret;
```

# Спасибо

## Полухин Антон

Старший разработчик Yandex. Тахі



antoshkka@gmail.com



antoshkka@yandex-team.ru



https://github.com/apolukhin



https://stdcpp.ru/



## Бонусы

shared\_ptr

#### shared\_ptr

```
#include <memory>
std::shared_ptr<int> p = std::make_shared<int>();
auto p1 = p;
*p1 = 42;
assert(*p == 42);
```

std::shared ptr<T> T Object Ptr to T Ptr to Control Block Control Block Reference Count Weak Count Other Data (e.g., custom deleter, allocator, etc.)

#### throw



memcpy/memmove

