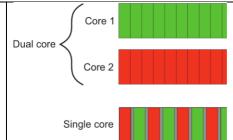
Область распределенных вычислений — раздел теории вычислительных систем, изучающий теоретические вопросы организации распределенных систем.  Параллельные вычисления — форма вычислений, при которой несколько вычислений выполняются одновременно (в течение перекрывающихся периодов времени), вместо последовательных (с завершением одного до начала следующего). Являются частью распределённых вычислений.	Распределенная система — это такая систем в которой взаимодействие и синхронизация программных компонентов, выполняемых независимых сетевых компьютерах, осуществляется посредством передачи сообщений.  Процесс (операционной системы) — это выполнение программных инструкций в рамках независимого виртуального адресно пространства, объединяющего код и данных компьютерной программы.	которой вычисления могут выполняться без ожидания завершения других вычислений. Современные ОС поддерживают параллелизм как минимум на уровне запускаемых программ.  Поток (выполнения) — программный механизм, предоставляемый ОС для параллельного выполнения программных
Адресное пространство пользователя — объединение кода и данных выполняемой программы в рамках запущенного в ОС процесса, защищённые от области ядра ОС и других процессов.	Параллельная машина с произвольным доступом к памяти PRAM (наиболее популярная) — объединены р процессоров, общая память и устройство управления, которое передает команды программы П процессорам.	Устройство называется многопроцессорным, если в его составе используется два или более физических процессора. Операционная система или программа называется многопроцессорной, если способна распределять задачи между процессорами
Классификация Флинна поток определяется как последовательность команд или данных,выполняемых или обрабатываемых процессором. В этой модели программа может предоставлять процессору не только потокинструкций для исполнения, но и поток данных для обработки. Потоки команд и потоки данных предполагаются независимыми.	Классификация Флинна 1. Один поток команд, один поток данных. 2. Один поток команд, несколько потоков данных — одна и та же операция выполняется одновременно над различными данными. 3. Несколько потоков команд, один поток данных — полезны в некоторых узкоспециальных задачах. 4. Несколько потоков команд, несколько потоков данных — большинство систем.	Среди <b>MIMD-систем</b> выделяют:
В системах с распределенной памятью каждый вычислительный узел имеет доступ только к принадлежащему ему участку локальной памяти. Для межпроцессорного обмена данными предусматривается возможность отправки и приема сообщений по коммуникационной сети, объединяющей вычислительную систему. Эта модель программирования — модель передачи сообщений.	Средства параллельного программировани 1. Автоматическое распараллеливание последовательной версии программы средствами компилятора; 2. Использование специализированных язы для параллельного программирования(Erlar 3. Использование библиотек, предоставляющих возможности параллельного исполнения кода; 4. Программирование с использованием особых расширений языка — средств распараллеливания.	RAM:     1. Система, состоит из процессора, устройства доступа к памяти (системной шины) и памяти, состоящей из конечного числа ячеек.
Системная шина  1.чтение данных и регистров гі, где  2.выполнение ари операции над со 3.запись данных и некоторую ячей	іствие процессора состоит из	Устройство управления Программа П $r_1^{(1)}$ $r_2^{(2)}$ $r_2^{(2)}$ $r_N^{(2)}$ $r_N^{(2)}$ $r_N^{(2)}$ $r_N^{(2)}$ $r_N^{(2)}$ $r_N^{(2)}$
Конфликты чтения. Если несколько процессоров пытаются прочесть данные из одной ячейки, то возможны два варианта дальнейших действий:  1. Исключающее чтение (ER). В данный момент времени чтение разрешено только одному процессору, в противном случае происходит ошибка выполнения программы.  2. Одновременное чтение (CR). Количество процессоров, получающих доступ к одной ячейке памяти, не ограничивается.	Конфликты записи. Если более одного процессора пытаются записать данные по одному адресу, разделяют два способа действия.  1. Исключающая запись (EW). Только одно процессору разрешено осуществлять запис данную ячейку в конкретный момент време.  2. Одновременная запись (CW). Несколько процессоров сразу получают доступ для зап к одной ячейке памяти.	ь в EREW-системе с р процессорами, который может быть исполнен за время O(T(N)log2 N). При этом объем памяти PRAM должен быть
Ускорение вычислений (PRAM) с $p$ процессорами $S_p(N) = \frac{T_1(N)}{T_p(N)}$ Это мера прироста производительности по сравнению с наилучшим последовательным алгоритмом.	Закон Амдала (PRAM). Пусть $f$ — доля последовательных вычислений в алгоритме Тогда ускорение Sp при использовании A на системе из p процессоров удовлетворяет неравенству $S_p\leqslant \frac{1}{f+(1-f)/p}$	* *

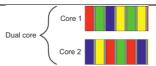
### **Искусственный параллелизм** — имитация параллельного выполнения задач на компьютерах с единственным ядром. **Истинный параллелизм** — возможность <mark>использовать несколько ядер для достижения</mark> параллельного выполнения задач. **Многопоточность** — искусственный или истинный параллелизм, реализованный в рамках одного процесса.

#### Задержки:

1. Чтобы обеспечить чередование, система производит контекстное переключение при каждом переходе от одной задачи к другой. 2. Чтобы переключить контекст, ОС должна сохранить состояние процессора и счетчик команд для текущей задачи, определить, какая задача следующая, и загрузить в процессор состояние новой задачи.



3. Загрузка команды и данных новой задачи в кэш-память.



С истинным параллелизмом количество задач может превышать число ядер, тогда будет применяться механизм переключения задач

#### Межпроцессорный параллелизм Плюсы:

- разделение обязанностей;
- высокая надёжность:
- в шаге от распределённой системы.

#### Минусы:

- используется больше ресурсов;
- взаимодействие занимает много времени.

#### Многопоточность. *Плюсы:*

- разделение обязанностей,
- высокая производительность<mark>,</mark>
- потребляется меньше ресурсов системы.

**Минус:** слабая надёжность.

Используется для: разделения обязанностей, повышения производительности, экономии ресурсов, контроль надежности программы.

Гонка данных — это ситуация, когда по Когда вредна многопоточность Затраты на распараллеливание не меньшей мере два потока имеют одновременный доступ к некоторой переменной и хотя бы один из потоков производит её запись.

# Проектирование многопоточности

**Распараллеливание по задачам** — разделение программы на задачи и запуск их параллельно. Распараллеливание по данным —выполнение одинаковых операций с различными данными. Многопоточность требует изменения подхода к проектированию.

компенсируют получаемый эффект: слишком мелкие и быстро выполняемые задачи (может помочь гетерогенный параллелизм, асинхронность); задач слишком много (может помочь пул потоков, асинхронность)

Мёртвая блокировка, взаимоблокировка,

двух или более потоков заблокирован в

и до своей разблокировки не может

**дедлок** — это состояние, в котором каждый из

ожидании ресурса, занятого другим потоком,

освободить ресурс, которого ожидает другой

остановка работы потоков, а чаще всего и всей

**Результат мёртвой блокировки** — полная

### Правила работы с мьютексами

1. Если функция выбросит исключение, мьютекс m никогда не будет освобождён. 2. Ни в коем случае нельзя вызывать из-под мьютекса функцию, внутреннее устройство которой неизвестно. В большинстве случаев <mark>неопределённое поведение выражается в</mark> <mark>мёртвой блокировке.</mark>

Мьютекс — примитив синхронизации, обеспечивающий взаимное исключение исполнения критических участков кода.

- 1. В каждый конкретный момент только один поток может владеть объектом, защищённым
- 2. Если другому потоку будет нужен доступ к данным, защищённым мьютексом, то этот поток блокируется до тех пор, пока мьютекс не будет освобождён.

Блокировщики управляют захватом и освобождением ресурса посредством идиомы **RAII** - захват ресурса есть инициализация. При объявлении объекта класса, работающего с файлами, на стеке происходит и его инициализация с вызовом конструктора, захватывающего ресурс. При выходе из областивидимости объект выталкивается из стека, но перед этим вызывается деструктор объекта, который и освобождает захваченный программы. Навечно. Если значение переменной никогда не изменяется, нет нужны синхронизировать доступ к ней с помощью механизмов блокировки. Нужно лишь присвоить ей начальное значение потокобезопасным способом. В языке С++ есть три способа потокобезопасной инициализации:

- константные выражения (constexpr);
- функция std::call once вместе с флагом std::once flag;
- локальная статическая переменная.

#### Потокобезопасная инициализация Локальная статическая переменная

Статические переменные, локальные для блока, инициализируются один раз ленивым образом.

В стандарте С++ 11 для локальных статических переменных появилась новая гарантия: их инициализация потокобезопасна

### Переменные условия

- 1. Позволяют синхронизировать потоки посредством обмена сообщениями
- 2. Один поток выступает отправителем сообщения, а другой поток (или несколько потоков) – получателем.
- 3.Получатель ждёт, пока не придёт сообщение. 4. Чаще всего применяются, когда нужно реализовать способ обработки данных по типу издателя и подписчика или производителя и потребителя.
- 5. Служит связующим звеном между отправителем и получателем сообщения.

### Переменные условия. Методы класса. notify one - оповестить один ожидающий поток о настуалении события; notify\_all – оповестить все потоки;

wait – ожидать сообщения, держа блокировщик открытым;

wait\_for – ожидать сообщения, держа блокировшик открытым, но не более заданного промежутка времени; wait until - ожидать сообщения, держа блокировщик открытым, но не более, чем до

заданного момента времени: native\_handle - возвращает системный дескриптор переменной условия

Переменные условия. Значение предиката

Предикат наделяет переменную условия состоянием. Функция ожидания всегда должна сначала проверить истинность предиката.

Предикат, таким образом, помогает бороться с двумя известными слабыми местами переменных условия: утерянным пробуждением и ложным пробуждением.

#### Переменные условия. Утерянные и ложные пробуждения

Утерянным пробуждением называется ситуация, когда поток-отправитель успевает послать оповещение до того, как получатель начинает его ожидать. Как следствие оповещение оказывается утерянным. **Ложное пробуждение** — это пробуждение ожидающего потока, когда отправители никаких оповещений не посылали. (одна из причин такого явления похищенное пробуждение: перед тем как пробуждённый поток-адресат получает шанс запуститься, другой поток успевает вклиниться первым и начинает выполнение)

Семафор — примитив синхронизации работы процессов и потоков, в основе которого лежит счётчик, над которым можно производить две атомарные операции: увеличение уменьшение значения на единицу, при этом операция уменьшения для нулевого значения счётчика является недопустимой. Служит для построения сложных механизмов

синхронизации и используется для синхронизации параллельно работающих задач, для защиты передачи данных через разделяемую память, для защиты критических секций, а также для управления доступом к аппаратному обеспечению.

- 1. Семафор это структура данных, содержащая очередь и счётчик.
- 2. Счётчик инициализируется значением, большим или равным нулю.
- 3. Операция wait захватывает семафор, уменьшая значение счётчика, если оно положительно, или блокирует поток в противном случае.
- 4. Операция signal освобождает семафор путём увеличения счётчика и, если очередь заблокированных потоков не пуста, пробуждает первый поток из неё.
- 5. Постановка заблокированных потоков в очередь необходима для предотвращения ресурсного голода.

Критическая секция — участок исполняемого кода программы, в котором производится доступ к общему ресурсу (данным или устройству), который не должен быть одновременно использован более чем одним потоком выполнения. При нахождении в критической секции двух (или более) потоков возникает состояние «гонки».

Ресурсный голод — ситуация, когда поток или процесс не может продолжить работу и проводит неопределённо долгое время в ожидании из-за того, что система каждый раз отказывает ему в предоставления некоторого ресурса

Асинхронные задания (механизм будущих результатов) — позволяет выполнять две операции параллельно, пока результат выполнения одной операции не понадобятся для другой.

Задание характеризуется пакетом работы, которую предстоит выполнить, и обладает двумя частями: **обещанием** и **будущим**.

Асинхронные задания ведут себя подобно каналам, по которым данные проходят от входного разъёма до выходного. Один конец канала называется обещанием (помещает данные в канал), другой — будущим (в неопределённый момент в будущем извлекает из канала результат их преобразования). Этими разъёмами может управлять один и тот же канал, а могут — разные.

Критерий	Потоки	Задания
Основные сущности	Родительский и дочерний потоки	Обещание и фьючерс
Способ передачи данных	Общая переменная	Канал
Отдельный поток	Всегда	Иногда
Синхронизация	Функция join ожидает завершения потока	Функция get блокирует выполнение
Исключение в дочернем потоке (задании)	Оба потока завершаются вместе со всем процессом	Передаётся через обещание и фьючерс
Передаваемые данные	Значения	Значения, оповещения и исключения

Всегда следует предпочитать асинхронные вызовы. С++ сам решает, выполнять асинхронный вызов в отдельном потоке или нет (зависит от доступных ядер процессора, загруженности системы и т.п.). Вызывая функцию std::аsync, программист лишь передаёт ей задание, которое должно быть выполнено (+ можно передать политику запуска). Вся работа по возможному созданию потока и управлению временем его жизни перекладывается на внутренние механизмы реализации.

Асинхронные задания. Политика запуска.

Можно в явном виде указать, каким образом реализации следует выполнить асинхронный вызов: в том же потоке, который создал вызов, или в другом потоке.

Обещание запускается только тогла, когла

Обещание запускается только тогда, когда фьючерс в явном виде запрашивает его результат (ленивые вычисления).

При строгой (жадной) стратегии выражение вычисляется немедленно, тогда как при ленивой (отложенной) стратегии вычисление откладывается до тех пор, пока значение не станет необходимо.

Особый случай составляют фьючерсы, о которых забывают сразу после создания. Они не сохраняются в каких-либо переменных и должны запускаться немедленно в момент создания. Существенно, что обещания таких фьючерсов должны выполняться в отдельном потоке, чтобы фьючерс мог начать работу

Короткоживущие фьючерсы выглядят удобными, но имеют существенный недостаток — они выполняются на самом деле последовательно.

Модель памяти — это договорённость о том, что: разработчик программы предотвращает гонку данных, а компилятор гарантирует строгость последовательности выполнения. Модель памяти С++ была создана таковой, чтобы обеспечивать оптимальный многопоточный код под множество архитектур процессоров.

Строгость последовательности выполнения — это результат любого выполнения, как если бы: — операции всех потоков выполняются в некотором последовательном порядке; — операции каждого потока проявляются в этой последовательности в порядке, указанном их программой.

Механизмы синхронизации потоков std::atomic<>. Наиболее тонкий, сложный и низкоуровневый механизм синхронизации и реализации атомарных операций. Переменные, защищённые от гонки данных.

Атомарная операция — это операция, которую невозможно наблюдать в промежуточном состоянии, она либо выполнена, либо нет. Атомарные операции могут состоять из нескольких операций.

Обеспечивает синхронизацию потоков.

# **Нежелательное блокирование** состояние потока, когда он не может

продолжать выполнение, так как чего-то ждет. Взаимоблокировка — когда один поток ждет другого, а тот, в свою очередь, ждет первого. Блокировка в ожидании завершения ввода/вывода или поступления данных из внешнего источника — если поток блокируется в ожидании данных из внешнего источника, то он не может продолжать работу, даже если данные так никогда и не поступят.

**Сопрограмма** (корутина) — блок кода, который работает асинхронно с вызывающим кодом, то есть по очереди.

**Сопрограмма** (корутина) — это функция, которая может приостановить выполнение, чтобы возобновить его позже.

#### Какие бывают:

**Стековые** сопрограммы — это сопрограммы, у которых есть свой стек (как у потоков).

**Безстековые** сопрограммы — никак не зависят от операционной системы, и реализуются исключительно средствами компилятора.

#### Где применяются:

**Конечные автоматы** в рамках одной подпрограммы;

**Акторная модель параллелизма**, например в видеоиграх;

#### Генераторы;

Передача последовательности процессов, где каждый подпроцесс является сопрограммой; Обратная связь, обычно используется в математическом ПО.

Условие (порядок) синхронизации std::memory\_order\_seq\_cst

Наиболее строгий порядок синхронизации. Свойства:

- порядок модификаций разных атомарных переменных в потоке thread1 сохранится в потоке thread2;
- потоке (птеаи2,
   все потоки будут видеть один и тот же
  порядок модификации всех атомарных
  переменных (сами модификациимогут
  происходить в разных потоках);
- все модификации памяти (не только модификации над атомиками) в потоке thread1, выполняющей store на атомарной переменной, будут видны после выполнения load этой же переменной в потоке thread2.

Таким образом можно представить seq\_cst операции, как барьеры памяти, в которых состояние памяти синхронизируется между всеми потоками программы.

Этот флаг синхронизации памяти в C++ используется по умолчанию

# Условие (порядок) синхронизации std::memory\_order\_relaxed

Он гарантирует только свойство атомарности операций, при этом не может участвовать в процессе синхронизации данных между потоками.

#### Свойства:

модификация переменной "появится" в другом потоке не сразу;

-поток thread2 "увидит" значения одной и той же переменной в том же порядке, в котором происходили её модификации в потоке thread1;

-порядок модификаций разных переменных в потоке thread1 не сохранится в потоке thread2. Гонка за данными — это особый тип гонки, который приводит к неопределенному поведению из-за несинхронизированного одновременногодоступа к разделяемой ячейке памяти.

– Обычно гонка за данными возникает вследствие неправильного использования атомарных операций для синхронизации потоков или в результате доступа к разделяемым данным, не защищенного подходящим мьютексом.

Нарушение инвариантов — такие гонки могут проявляться в форме висячих указателей (другой поток уже удалил данные, к которым мы пытаемся обратиться), случайного повреждения памяти (из-за того, что поток читает данные, оказавшиеся несогласованными в результатечастичного обновления) и двойного освобождения (например, два потока извлекают из очереди одно и то же значение, и потом оба удаляют ассоциированные с ним данные).

Проблемы со временем жизни — такого рода проблемы можно было бы отнести к нарушению инвариантов, но на самом деле это отдельная категория. Основная проблема в том, что поток живет дольше, чем данные, к которым он обращается, поэтому может попытаться получить доступ к уже удаленным или разрушенным иным способом данным. Не исключено также, что когда-то отведенная под эти данные память уже занята другим объектом. Обычно такие ошибки возникают, когда поток хранит ссылки на локальные переменные, которые вышли из области видимости до завершения функции потока, но это не единственный сценарий.

Флаг синхронизации памяти acquire/release

является более тонким способом синхронизировать данные между парой потоков (работают только в паре над одним атомарным объектом).

#### Свойства:

– модификация атомарной переменной с release будет видна в другом потоке, выполняющем чтение этой же атомарной переменной с acquire;
 – все модификации памяти в потоке thread1, выполняющей запись атомарной переменной с release, будут видны после выполнения чтения той же переменной с acquire в потоке

- процессор и компилятор не могут перенести операции записи в память раньше release - операции в потоке thread1, и нельзя перемещать выше операции чтения из памяти позже acquire операции в потоке thread2. Пул потоков – поддерживает несколько потоков, ожидающих выделения задач для одновременного выполнения управляющей программой.

Очередь задач — позволяет развязать основной алгоритм и другие шаблоны параллельного программирования (часто применяется с пулом потоков). Блокиратор — безопасное использование

блокировки (как правило на основе мьютекса) с применением идиомы RAII в C++. Активный объект — шаблон проектирования, который отделяет поток выполнения метода от потока, в котором он был вызван (см. модель

акторов).

Блокировка с двойной проверкой —
предназначен для уменьшения накладных расходов, связанных с получением блокировки.

Монитор – объект, предназначенный для безопасного использования более чем одним потоком

Планировщик - обеспечивающий механизм реализации политики планирования, но при этом не зависящий ни от одной конкретной политики.

Пул потоков — шаблон параллельного программирования, который поддерживает несколько потоков, ожидающих выделения задач для одновременного выполнения управляющей программой.

#### Преимущества:

Накладные расходы на создание и уничтожение потоков ограничены -> растет производительность; Обеспечивает управление параллелизмом и

Обеспечивает управление параллелизмом и синхронизацией потоков на более высоком уровне абстракции

#### Недостатки:

При активном использовании синхронизациимежду задачами в пуле потоков понижается его эффективность.

Активная блокировка — взаимоблокировка при цикле активной проверки, например спинлоке. В одних случаях программа потребляет много процессорного времени, потому что потоки блокируют друг друга, продолжая работать. В других — блокировка рано или поздно может «рассасаться» из-за вероятностной природы планирования, но заблокированная задача испытывает ощутимые задержки, характеризуемые высоким потреблением процессорного времени.

# Использование доступных инструментов (желательно в CI/CD):

- «выкрутить» диагностику компилятора намаксимум;
- прогнать через статический анализатор;
- прогнать через специальный профилировщик.
   «Ручной» просмотр:
- посмотреть самому;
- показать другу:
- объяснить кому-то, как работает ваш код (можно плюшевому мишке или утёнку).

# Поиск связанных с параллелизмом ошибок путем тестирования:

проверить в однопоточном режиме;
 протестировать в многопоточном.

Г**естирование грубой силой** — код

прогоняется многократно на большом количестве потоков (например, стресстест). Позволяет «поймать» редкие ошибочные ситуации. Недостаток метода грубой силы в том, что он может вселять ложную уверенность.

#### Комбинаторное имитационное тестирование

- прогонять код под управлением специальной программы, которая имитирует реальную среду.

Обнаружение возникающих во время тестирования проблем с помощью специальной библиотеки.

## Нагрузочное тестирование

Для исследования времени отклика системы на высоких или пиковых нагрузках. Моделирования ожидаемого использования приложения с помощью эмуляции работы нескольких пользователей одновременно. Таким образом, подобное тестирование больше всего подходит для многопользовательских систем, чаще — использующих клиент-серверную архитектуру (например, веб-серверов).

#### Принципы:

- Уникальность запросов;
- Время отклика системы;
- Зависимость времени отклика системы от степени распределённости этой системы;
- Разброс времени отклика системы;
- Точность воспроизведения профилей нагрузки.

#### Регрессионное тестирование

Направленны на обнаружение ошибок в уже протестированных участках исходного кода. Такие ошибки — когда после внесения изменений в программу, перестаёт работать то, что должно было продолжать работать, — называют регрессионными ошибками.

#### Классификация:

- Минимизация набора тестов
- Задача минимизации наборов
- Задача с определением приоритетов
- Задача выбора тестов