

Nebenläufigkeit in Java

Concurrency in Java

Dozent: Prof. Dr. Michael Eichberg
Kontakt: michael.eichberg@dhbw-mannheim.de
Version: 2024-03-29



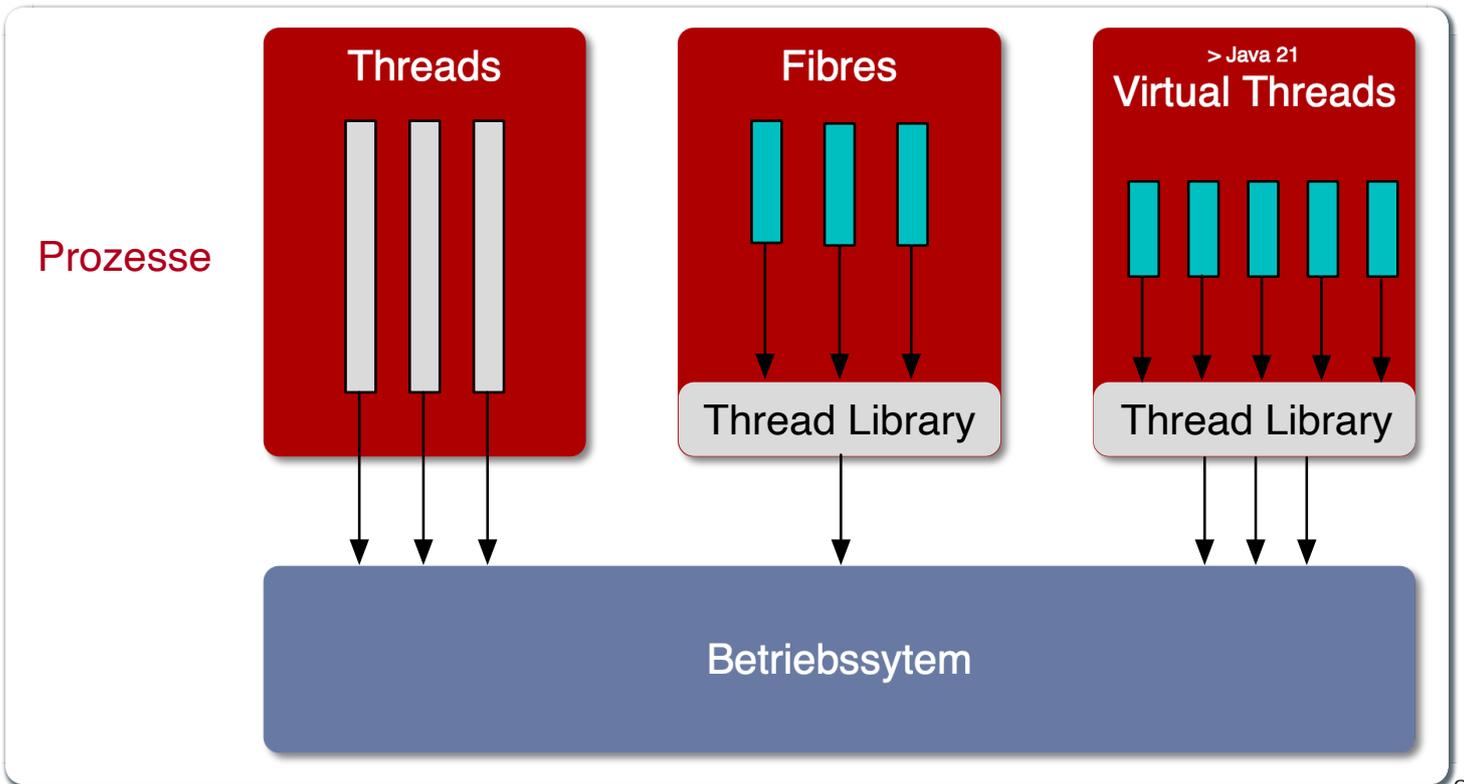
1

Folien: <https://delors.github.io/ds-nebenlaeufigkeit-in-java/folien.rst.html>
<https://delors.github.io/ds-nebenlaeufigkeit-in-java/folien.rst.html.pdf>

Fehler auf Folien melden:
<https://github.com/Delors/delors.github.io/issues>

Ein gutes Verständnis von nebenläufiger Programmierung ist für die Entwicklung von verteilten Anwendungen unerlässlich, da Server immer mehrere Anfragen gleichzeitig bearbeiten.

Prozesse vs. Threads

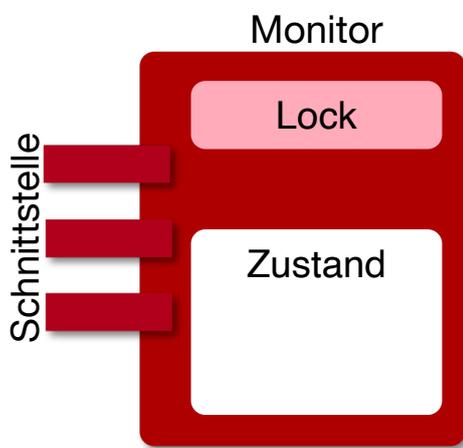


3

- Prozesse sind voneinander isoliert und können nur über explizite Mechanismen miteinander kommunizieren; Prozesse teilen sich nicht denselben Adressraum.
- Alle Threads eines Prozesses teilen sich denselben Adressraum. *Native Threads* sind vom Betriebssystem unterstützte Threads, die direkt vom Betriebssystem verwaltet werden. Standard Java Threads sind *Native Threads*.
- *Fibres* (auch *Coroutines*) nutzen immer kooperatives Multitasking. D. h. ein Fibre gibt die Kontrolle an eine andere Fibre explizit ab. (Früher auch als *Green Threads* bezeichnet.) Diese sind für das Betriebssystem unsichtbar.
- Ab Java 21 unterstützt Java nicht nur klassische (native) Threads sondern zusätzlich auf Virtual Threads. Letztere erlauben insbesondere eine sehr natürliche Programmierung von Middleware, die sich um die Parallelisierung/Nebenläufigkeit kümmert.

Kommunikation und Synchronisation mit Hilfe von Monitoren

Ein *Monitor* ist ein Objekt, bei dem die Methoden im wechselseitigen Ausschluss (engl. *mutual exclusion*) ausgeführt werden.



Bedingungs-Synchronisation

- drückt eine Bedingung für die Reihenfolge der Ausführung von Operationen aus.
- z.B. können Daten erst dann aus einem Puffer entfernt werden, wenn Daten in den Puffer eingegeben wurden.
- Java unterstützt pro Monitor nur eine (anonyme) Bedingungs-Variable, mit den klassischen Methoden `wait` und `notify` bzw. `notifyAll`.

4

Warnung

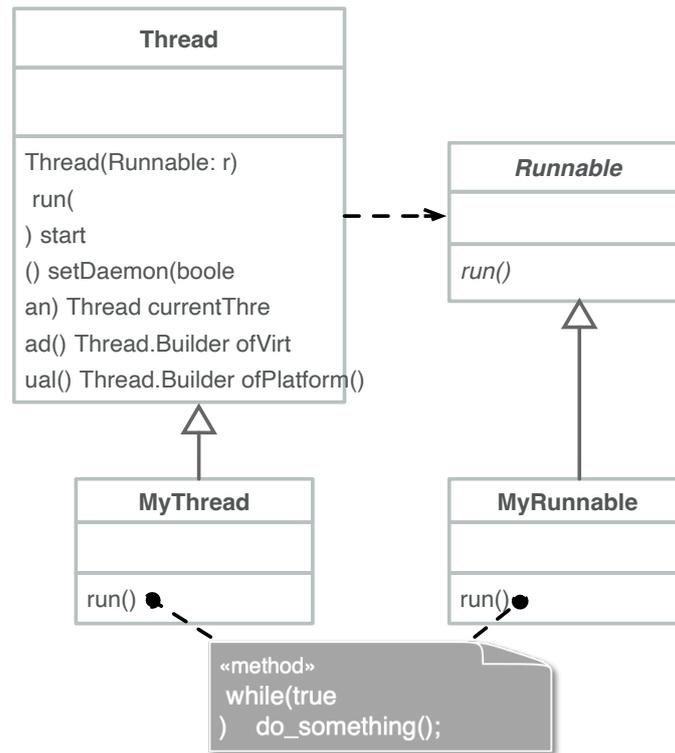
In Java findet der wechselseitige Ausschluss nur zwischen solchen Methoden statt, die explizit als `synchronized` deklariert wurden.

Monitore sind nur ein Modell (Alternativen: *Semaphores*, *Message Passing*), das die Kommunikation und Synchronisation von Threads ermöglicht. Es ist das Standardmodell in Java und wird von der Java Virtual Machine (JVM) unterstützt.

Kommunikation zwischen Threads mit Hilfe von Monitoren

- Durch Lesen und Schreiben von Daten, die in gemeinsamen Objekten gekapselt sind, die durch Monitore geschützt werden.
- Jedes Objekt ist implizit von der Klasse **Object** abgeleitet, welche eine gegenseitige Ausschlussperre definiert.
- Methoden in einer Klasse können als **synchronized** gekennzeichnet werden. Die Methode wird erst dann ausgeführt, wenn die Sperre vorliegt. Bis dahin wird gewartet. Dieser Prozess geschieht automatisch.
- Die Sperre kann auch über eine **synchronized** Anweisung erworben werden, die das Objekt benennt.
- Ein Thread kann auf eine einzelne (anonyme) Bedingungsvariable warten und diese benachrichtigen.

Nebenfäufigkeit in Java



6

- Threads werden in Java über die vordefinierte Klasse `java.lang.Thread` bereitgestellt.
- Alternativ kann das Interface:

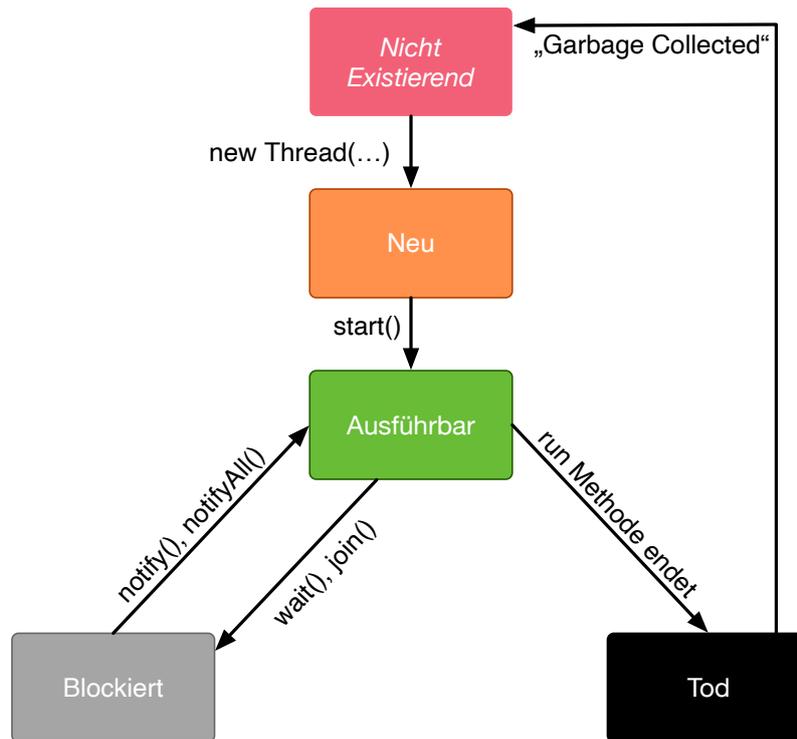
```
public interface Runnable { void run(); }
```

implementiert werden und an ein `Thread`-Objekt übergeben werden.
- Threads beginnen ihre Ausführung erst, wenn die `start`-Methode in der `Thread`-Klasse aufgerufen wird. Die `Thread.start`-Methode ruft die `run`-Methode auf. Ein Aufruf der `run`-Methode direkt führt nicht zu einer parallelen Ausführung.
- Der aktuelle Thread kann mittels der statischen Methode `Thread.currentThread()` ermittelt werden.
- Ein Thread wird beendet, wenn die Ausführung seiner `run`-Methode entweder normal oder als Ergebnis einer unbehandelten Ausnahme endet.
- Java unterscheidet *User-Threads* und *Daemon-Threads*.
Daemon-Threads sind Threads, die allgemeine Dienste bereitstellen und normalerweise nie beendet werden. Wenn alle Benutzer-Threads beendet sind, werden die *Daemon-Threads* von der JVM beendet, und das Hauptprogramm wird beendet.
Die Methode `setDaemon` muss aufgerufen werden, bevor der Thread gestartet wird.

Inter-Thread-Kommunikation bzw. Koordination

- Ein Thread kann (mit oder ohne Zeitüberschreitung) auf die Beendigung eines anderen Threads (des Ziels) warten, indem er die `join`-Methode für das Thread-Objekt des Ziels aufruft.
- Mit der Methode `isAlive` kann ein Thread feststellen, ob der Ziel-Thread beendet wurde.

Java Thread States



synchronized-Methoden und synchronized-Blöcke

- Jedem Objekt ist eine gegenseitige Ausschlussperre zugeordnet. Auf die Sperre kann von der Anwendung nicht explizit zugegriffen werden. Dies geschieht implizit, wenn:
 - eine Methode den Methodenmodifikator **synchronized** verwendet
 - Blocksynchronisierung mit dem Schlüsselwort **synchronized** verwendet wird
- Wenn eine Methode als synchronisiert gekennzeichnet ist, kann der Zugriff auf die Methode nur erfolgen, wenn das System die Sperre erhalten hat.
- Daher haben synchronisierte Methoden einen sich gegenseitig ausschließenden Zugriff auf die vom Objekt gekapselten Daten, **wenn auf diese Daten nur von anderen synchronisierten Methoden zugegriffen wird.**
- Nicht-synchronisierte Methoden benötigen keine Sperre und können daher *jederzeit* aufgerufen werden.

Beispiel: Synchronisierte Methode

```
public class SynchronizedCounter {  
  
    private int count = 0;  
  
    public synchronized void increment() {  
        count++;  
    }  
  
    public synchronized int getCount() {  
        return count;  
    }  
}
```

```
public class SharedLong {  
  
    private long theData; // reading and writing longs is not atomic  
  
    public SharedLong(long initialValue) {  
        theData = initialValue;  
    }  
  
    public synchronized long read() { return theData; }  
  
    public synchronized void write(long newValue) { theData = newValue; }  
  
    public synchronized void incrementBy(long by) {  
        theData = theData + by;  
    }  
}  
  
SharedLong myData = new SharedLong(42);
```

```
public class SynchronizedCounter {  
  
    private int count = 0;  
  
    public void increment() {  
        synchronized(this) {  
            count++;  
        }  
    }  
  
    public int getCount() {  
        synchronized(this) {  
            return count;  
        }  
    }  
}
```

Warnung

Wenn **synchronized** in seiner ganzen Allgemeinheit verwendet wird, kann er einen der Vorteile von klassischen Monitoren untergraben: Die Kapselung von Synchronisationseinschränkungen, die mit einem Objekt verbunden sind, an einer einzigen Stelle im Programm!

Dies liegt daran, dass es nicht möglich ist, die mit einem bestimmten Objekt verbundene Synchronisation zu verstehen, indem man sich nur das Objekt selbst ansieht. Andere Objekte können bzgl. des Objekts eine **synchronized**-Block verwenden.

Komplexe Rückgabewerte

```
public class SharedCoordinate {  
  
    private int x, y;  
  
    public SharedCoordinate(int initX, int initY) {  
        this.x = initX; this.y = initY;  
    }  
  
    public synchronized void write(int newX, int newY) {  
        this.x = newX; this.y = newY;  
    }  
  
    /* ⚠️ */ public /* synchronized irrelevant */ int readX() { return x; } /* ⚠️ */  
    /* ⚠️ */ public /* synchronized irrelevant */ int readY() { return y; } /* ⚠️ */  
  
    public synchronized SharedCoordinate read() {  
        return new SharedCoordinate(x, y);  
    }  
}
```

11

Die beiden Methoden: `readX` und `readY` sind nicht synchronisiert, da das Lesen von `int`-Werten atomar ist. Allerdings erlauben sie das Auslesen eines inkonsistenten Zustands! Es ist denkbar, dass direkt nach einem `readX` der entsprechende Thread unterbrochen wird und ein anderer Thread die Werte von `x` und `y` verändert. Wird dann der ursprüngliche Thread fortgesetzt, und ruft `readY` auf, so erhält er den neuen Wert von `y` und hat somit ein paar `x, y` vorliegen, dass in dieser Form nie existiert hat.

Ein konsistenter Zustand kann nur durch die Methode `read` ermittelt werden, die die Werte von `x` und `y` in einem Schritt ausliest und als Paar zurückgibt.

Kann sichergestellt werden, dass ein auslesender Thread die Instanz in einem `synchronized` Block benennt, dann kann die Auslesung eines konsistenten Zustands auch bei mehreren Methodenaufrufen hintereinander sichergestellt werden.

```
SharedCoordinate point = new SharedCoordinate(0,0);  
synchronized (point1) {  
    var x = point1.readX();  
    var y = point1.readY();  
}  
// do something with x and y
```

Diese „Lösung“ muss jedoch als sehr kritisch betrachtet werden, da die Wahrscheinlichkeit von Programmierfehlern *sehr hoch* ist und es dann entweder zur *Race Conditions* oder zu *Deadlocks* kommen kann.

Bedingte Synchronisation

Zum Zwecke der bedingten Synchronisation können in Java die Methoden **wait**, **notify** und **notifyAll** verwendet werden. Diese Methoden erlauben es auf bestimmte Bedingungen zu warten und andere Threads zu benachrichtigen, wenn sich die Bedingung geändert hat.

- Diese Methoden können nur innerhalb von Methoden verwendet werden, die die Objektsperre halten; andernfalls wird eine **IllegalMonitorStateException** ausgelöst.

1

- Die **wait**-Methode blockiert immer den aufrufenden Thread und gibt die mit dem Objekt verbundene Sperre frei.

2

- Die **notify**-Methode weckt *einen* wartenden Thread auf. Welcher Thread aufgeweckt wird, ist nicht spezifiziert.

notify gibt die Sperre nicht frei; daher muss der aufgeweckte Thread warten, bis er die Sperre erhalten kann, bevor er fortfahren kann.

- Um alle wartenden Threads aufzuwecken, muss die Methode **notifyAll** verwendet werden.

Warten die Threads aufgrund unterschiedlicher Bedingungen, so ist immer **notifyAll** zu verwenden.

- Wenn kein Thread wartet, dann haben **notify** und **notifyAll** keine Wirkung.

3

Wichtig

Wenn ein Thread aufgeweckt wird, kann er nicht davon ausgehen, dass seine Bedingung erfüllt ist!
Die Bedingung ist immer in einer Schleife zu prüfen und der Thread muss sich ggf. wieder in den Wartezustand versetzen.

4

Beispiel: Bedingte Synchronisation mit *Condition Variables*

Ein *BoundedBuffer* hat z.B. traditionell zwei Bedingungsvariablen: *BufferNotFull* und *BufferNotEmpty*.

Wenn ein Thread auf eine Bedingung wartet, kann kein anderer Thread auf die andere Bedingung warten.

Mit den bisher vorgestellten Primitiven ist eine direkte Modellierung dieses Szenarios so nicht möglich. Stattdessen müssen immer alle Threads aufgeweckt werden, um sicherzustellen, dass auch der intendierte Thread aufgeweckt wird. Deswegen ist auch das Überprüfen der Bedingung in einer Schleife notwendig.

1

```
public class BoundedBuffer {
    private final int buffer[];
    private int first;
    private int last;
    private int numberInBuffer = 0;
    private final int size;

    public BoundedBuffer(int length) {
        size = length;
        buffer = new int[size];
        last = 0;
        first = 0;
    };
    ...
}
```

```
public synchronized void put(int item) throws InterruptedException {
    while (numberInBuffer == size)
        wait();
    last = (last + 1) % size;
    numberInBuffer++;
    buffer[last] = item;
    notifyAll();
};
```

```
public synchronized int get() throws InterruptedException {
    while (numberInBuffer == 0)
        wait();
    first = (first + 1) % size;
    numberInBuffer--;
    notifyAll();
    return buffer[first];
}
}
```

Fehlersituation, die bei der Verwendung von **notify** (statt **notifyAll**) auftreten könnte.

```
BoundedBuffer bb = new BoundedBuffer(1);
Thread g1,g2 = new Thread(() => { bb.get(); } );
Thread p1,p2 = new Thread(() => { bb.put(new Object()); } );
```

```
g1.start(); g2.start(); p1.start(); p2.start();
```

	Aktionen	(Änderung des) Zustand(s) des Buffers	Auf die Sperre (Lock) wartend	An der Bedingung wartend
1	g1:bb.get() g2:bb.get(), p1:bb.put(), p2:bb.put()	empty	{g2,p1,p2}	{g1}
2	g2:bb.get()	empty	{p1,p2}	{g1,g2}
3	p1:bb.put()	empty → not empty	{p2,g1}	{g2}
4	p2:bb.put()	not empty	{g1}	{g2,p2}
5	g1:bb.get()	not empty → empty	{g2}	{p2}
6	g2:bb.get()	empty	∅	{g2,p2}

13

In Schritt 5 wurde von der VM - aufgrund des Aufrufs von `notify` durch `g1` - der Thread `g2` aufgeweckt - anstatt des Threads `p2`. Der aufgeweckte Thread `g2` prüft die Bedingung (Schritt 6) und stellt fest, dass der Buffer leer ist. Er geht wieder in den Wartezustand. Jetzt warten sowohl ein Thread, der ein Wert schreiben möchte als auch ein Thread, der einen Wert lesen möchte.

1. FORTGESCHRITTENE SYNCHRONISATIONSMECHANISMEN, -PRIMITIVE UND -KONZEPTE.

Java API für nebenläufige Programmierung

java.util.concurrent:

Bietet verschiedene Klassen zur Unterstützung gängiger nebenläufiger Programmierparadigmen, z. B. Unterstützung für *BoundedBuffers* oder Thread-Pools.

java.util.concurrent.atomic:

Bietet Unterstützung für sperrfreie (*lock-free*), thread-sichere Programmierung auf einfachen Variablen — wie zum Beispiel atomaren Integern — an.

java.util.concurrent.locks:

Bietet verschiedene Sperralgorithmen an, die die Java-Sprachmechanismen ergänzen, z. B. Schreib-Lese-Sperren und Bedingungsvariablen. Dies ermöglicht zum Beispiel: „Hand-over-Hand“ oder „Chain Locking“.

Beispiel: Bedingte Synchronisation mit *ReentrantLocks*.

Ein *BoundedBuffer* hat z.B. traditionell zwei Bedingungsvariablen: *BufferNotFull* und *BufferNotEmpty*.

```
public class BoundedBuffer<T> {  
  
    private final T buffer[];  
    private int first;  
    private int last;  
    private int numberInBuffer;  
    private final int size;  
  
    private final Lock lock = new ReentrantLock();  
    private final Condition notFull = lock.newCondition();  
    private final Condition notEmpty = lock.newCondition();
```

```
public BoundedBuffer(int length) { /* Normaler Constructor. */  
    size = length;  
    buffer = (T[]) new Object[size];  
    last = 0;  
    first = 0;  
    numberInBuffer = 0;  
}
```

```
public void put(T item) throws InterruptedException {  
    lock.lock();  
    try {  
  
        while (numberInBuffer == size) { notFull.await(); }  
        last = (last + 1) % size;  
        numberInBuffer++;  
        buffer[last] = item;  
        notEmpty.signal();  
  
    } finally {  
        lock.unlock();  
    }  
}
```

```
public T get() ... {  
    lock.lock();  
    try {  
  
        while (numberInBuffer == 0) { notEmpty.await(); }  
        first = (first + 1) % size;  
        numberInBuffer--;  
        notFull.signal();  
        return buffer[first];  
  
    } finally {  
        lock.unlock();  
    }  
}
```


Thread Prioritäten

- Obwohl den Java-Threads Prioritäten zugewiesen werden können (**setPriority**), dienen sie dem zugrunde liegenden Scheduler nur als Richtschnur für die Ressourcenzuweisung.
- Sobald ein Thread läuft, kann er die Prozessorressourcen explizit aufgeben, indem er die **yield**-Methode aufruft.
- **yield** setzt den Thread an das Ende der Warteschlange für seine Prioritätsstufe.
- Die Scheduling- und Prioritätsmodelle von Java sind jedoch schwach:
 - Es gibt keine Garantie dafür, dass immer der Thread mit der höchsten Priorität ausgeführt wird, der lauffähig ist.
 - Threads mit gleicher Priorität können in Zeitscheiben unterteilt sein oder auch nicht.
 - Bei der Verwendung nativer Threads können unterschiedliche Java-Prioritäten auf dieselbe Betriebssystempriorität abgebildet werden.

Best Practices

synchronized Code sollte so kurz wie möglich gehalten werden.

Verschachtelte Monitorkaufrufe sollten vermieden werden, da die äußere Sperre nicht freigegeben wird, wenn der innere Monitor wartet. Dies kann leicht zum Auftreten eines Deadlocks führen.

Wenn zwei (oder mehr) Threads auf die gleichen Ressourcen in unterschiedlicher Reihenfolge zugreifen, kann es zu einem Deadlock kommen.

Zu Beachten

Ressourcen immer in der gleichen Reihenfolge sperren, um Deadlocks zu vermeiden.

2. THREAD SAFETY

Prof. Dr. Michael Eichberg

▣ *Threadsicherheit*

Thread Safety - Voraussetzung

Damit eine Klasse thread-sicher ist, muss sie sich in einer single-threaded Umgebung korrekt verhalten.

D.h. wenn eine Klasse korrekt implementiert ist, dann sollte keine Abfolge von Operationen (Lesen oder Schreiben von öffentlichen Feldern und Aufrufen von öffentlichen Methoden) auf Objekten dieser Klasse in der Lage sein:

- das Objekt in einen ungültigen Zustand versetzen,
- das Objekt in einem ungültigen Zustand zu beobachten oder
- eine der Invarianten, Vorbedingungen oder Nachbedingungen der Klasse verletzen.

1

Die Klasse muss das korrekte Verhalten auch dann aufweisen, wenn auf sie von mehreren Threads aus zugegriffen wird.

- Unabhängig vom *Scheduling* oder der Verschachtelung der Ausführung dieser Threads durch die Laufzeitumgebung,
- Ohne zusätzliche Synchronisierung auf Seiten des aufrufenden Codes.

2

Dies hat zur Folge, dass Operationen auf einem thread-sicheren Objekt für alle Threads so erscheinen als ob die Operationen in einer festen, global konsistenten Reihenfolge erfolgen würden.

Thread Safety Level

Immutable 🚩 *Unveränderlich:*

Die Objekte sind konstant und können nicht geändert werden.

Thread-sicher: Die Objekte sind veränderbar, unterstützen aber nebenläufigen Zugriff, da die Methoden entsprechend synchronisiert sind.

Bedingt Thread-sicher:

All solche Objekte bei denen jede einzelne Operation thread-sicher ist, aber bestimmte Sequenzen von Operationen eine externe Synchronisierung erfordern können.

Thread-kompatibel:

Alle Objekte die keinerlei Synchronisierung aufweisen. Der Aufrufer kann die Synchronisierung jedoch ggf. extern übernehmen.

Thread-hostile „Thread-schädlich“:

Objekte, die nicht thread-sicher sind und auch nicht thread-sicher gemacht werden können, da sie zum Beispiel globalen Zustand manipulieren.

Virtueller Puffer

Implementieren Sie einen virtuellen Puffer, der Tasks (Instanzen von `java.lang.Runnable`) entgegennimmt und nach einer bestimmten Zeit ausführt. Der Puffer darf währenddessen nicht blockieren bzw. gesperrt sein.

Nutzen Sie ggf. virtuelle Threads, um auf ein explizites Puffern zu verzichten. Ein virtueller Thread kann zum Beispiel mit: `Thread.ofVirtual()` erzeugt werden. Danach kann an die Methode `start` ein `Runnable` Objekt übergeben werden.

Verzögern Sie die Ausführung (`Thread.sleep()`) im Schnitt um 100ms mit einer Standardabweichung von 20ms. (Nutzen Sie `Random.nextGaussian(mean, stddev)`)

Starten Sie 100000 virtuelle Threads. Wie lange dauert die Ausführung? Wie lange dauert die Ausführung bei 100000 platform (*native*) Threads.

Nutzen Sie ggf. die Vorlage.

```
import java.util.ArrayList;
import java.util.List;
import java.util.Random;

public class VirtualBuffer {

    private final Random random = new Random();

    private Thread runDelayed(int id, Runnable task) {
        // TODO
    }

    public static void main(String[] args) throws Exception {
        var start = System.nanoTime();
        VirtualBuffer buffer = new VirtualBuffer();
        List<Thread> threads = new ArrayList<>();
        for (int i = 0; i < 100000; i++) {
            final var no = i;
            var thread = buffer.runDelayed(
                i,
                () -> System.out.println("i'm no.: " + no));
            threads.add(thread);
        }
        System.out.println("finished starting all threads");
        for (Thread thread : threads) {
            thread.join();
        }
        var runtime = (System.nanoTime() - start)/1_000_000;
        System.out.println(
            "all threads finished after: " + runtime + "ms"
        );
    }
}
```

Virtueller Puffer

Implementieren Sie einen virtuellen Puffer, der Tasks (Instanzen von `java.lang.Runnable`) entgegennimmt und nach einer bestimmten Zeit ausführt. Der Puffer darf währenddessen nicht blockieren bzw. gesperrt sein.

Nutzen Sie ggf. virtuelle Threads, um auf ein explizites Puffern zu verzichten. Ein virtueller Thread kann zum Beispiel mit: `Thread.ofVirtual()` erzeugt werden. Danach kann an die Methode `start` ein `Runnable` Objekt übergeben werden.

Verzögern Sie die Ausführung (`Thread.sleep()`) im Schnitt um 100ms mit einer Standardabweichung von 20ms. (Nutzen Sie `Random.nextGaussian(mean, stddev)`)

Starten Sie 100000 virtuelle Threads. Wie lange dauert die Ausführung? Wie lange dauert die Ausführung bei 100000 platform (*native*) Threads.

Nutzen Sie ggf. die Vorlage.

Thread-sichere Programmierung

Implementieren Sie eine Klasse **ThreadsafeArray** zum Speichern von nicht-**null** Objekten (**java.lang.Object**) an ausgewählten Indizes — vergleichbar mit einem normalen Array. Im Vergleich zu einem normalen Array sollen die Aufrufer jedoch ggf. blockiert werden, wenn die Zelle belegt ist. Die Klasse soll folgende Methoden bereitstellen:

get(int index): Liefert den Wert an der Position **index** zurück. Der aufrufende Thread wird ggf. blockiert, bis ein Wert an der Position **index** gespeichert wurde. (Die **get**-Methode entfernt den Wert nicht aus dem Array.)

set(int index, Object value):

Speichert den Wert **value** an der Position **index**. Falls an der Position **index** bereits ein Wert gespeichert wurde, wird der aufrufende Thread blockiert, bis der Wert an der Position **index** gelöscht wurde.

delete(int index):

Löscht ggf. den Wert an der Position **index** wenn ein Wert vorhanden ist. Andernfalls wird der Thread blockiert, bis es einen Wert gibt, der gelöscht werden kann.

- a. Implementieren Sie die Klasse **ThreadsafeArray** nur unter Verwendung der Standardprimitive: **synchronized**, **wait**, **notify** und **notifyAll**. Nutzen Sie die Vorlage.
- b. Können Sie sowohl **notify** als auch **notifyAll** verwenden?
- c. Implementieren Sie die Klasse **ThreadsafeArray** unter Verwendung von **ReentrantLocks** und **Conditions**. Nutzen Sie die Vorlage.
- d. Welche Vorteile hat die Verwendung von **ReentrantLocks**?