Threads

Threads

- **Threads** (an abbreviation of threads of control) are how we can get more than one thing to happen at once in a program.
- A thread (a.k.a. lightweight process) is associated with a particular process (a.k.a. heavyweight process, a retronym). A heavyweight process may have several threads.
- Physical threads are directly created and scheduled by the operating system. Virtual threads (aka green threads or user threads) are created and scheduled by the language runtime - which usually maps them to a smaller pool of physical threads (but not necessarily).
- Threads share the text, data and heap segments. Each thread has its own stack and status. Thus a thread has a minimum of internal state and a minimum of allocated resources.

How can you be in two place at once when you're not anywhere at all?
–Firesign Theater.

Creating Threads in Java (1)

Threads are part of the core Java language. There are two ways to create a new thread of execution.

- One is to the extend the class `java.lang.Thread`. This subclass should override the `run` method of class Thread. An instance of the subclass can then be allocated and started.

```
class Student extends Thread {
    public void run() {
        //listen, talk, daydream, fidget
        ...
    }
}
...
Student myStudent = new Student();
myStudent.start();
...
```

Creating Threads in Java (2)

- The second way to create a thread is to declare a class that implements the `Runnable` interface. That class then implements the `run` method. An instance of the class can then be allocated, passed as an argument when creating a Thread object, which can then be started.

```
class Student implements Runnable{
    public void run() {
        //listen, talk, daydream, fidget
        ...
    }
}
...
Student myStudent = new Student();
Thread myThread = new Thread(myStudent).start();
...
```

Basic Thread Examples in Java

- Example 1: Shows how to create a thread by subclassing `Thread` class: ThreadExample.java
- Example 2: Shows how to create a thread by implementing the `Runnable interface`: RunnableExample.java
- Example 3: Create a thread quagmire... MaxThreads.java

In Java, each thread is an object!

Threads: use cases

- To model nondeterminism found in the real world. E.g. rowing a boat with a modest leak in the hull!
- To improve performance by incorporating parallelism in an application.
  - We need to compute something. If we can break that into smaller, mostly independent parts, then multiple threads will help speed up the process. We can sort in parallel. We can search in parallel. We can compile a large code base in parallel.
  - A server can serve multiple clients by spinning off a thread for each client. A server can talk to multiple database servers simultaneously.
  - A client (such as a web browser) can use multiple threads to improve responsiveness for a client. For example, multiple threads can be downloading multiple files from separate servers (or even the same server if it is multithreaded)
- To capitalize on asynchronous behavior. E.g. waiting for I/O to complete, interacting with elements of a GUI application.

Threads for Parallelism

- The previous example showed the use of threads to simulate nondeterministic behavior. Let's see another use of threads to create parallelism.
- **Example 1**: A sequential program that simulates a widget factory: WidgetMaker.java
- **Example 2**: A parallel program to make widgets faster (using multiple threads): FasterWidgetMaker.java

- **In-class Exercise**
  Determine the number of cores on your system.
  Run *FasterWidgetMaker* for 4, 8, 16, 32, and 64 threads.
  Discuss the speedup you observe. Is it what you expected? Less or more?

Physical vs Virtual Threads

- Threads that are managed by the language runtime are known as virtual (or green or user) threads. Typically the runtime gets a smaller collection of physical threads from the OS and then manages the scheduling of the virtual threads on the physical threads.
- Java supports physical threads as well as virtual threads (*production ready only from version 21*). Virtual threads are similar to goroutines in Go.
- Example 1: Shows the difference in speed between creating virtual threads vs physical threads: VirtualThreadsDemo.java
- Example 2: Shows how we can create a large number of virtual threads: MaxVirtualThreads.java

Relevant Java Classes/Interfaces

- See documentation for basic classes: `java.lang.Thread`, `java.lang.ThreadGroup` and `java.lang.Runnable` interface.
- See the `java.lang.Object` class for synchronization methods.
- For automatic management of threads, see: `Executor` interface from `java.util.concurrent` package.

Controlling Threads

- `start()` and `stop()` [deprecated]
- `suspend()`[deprecated], `resume()`[deprecated], `sleep()`
- `join()`: causes the caller to block until the thread dies or with an argument
- `interrupt()`: wake up a thread that is sleeping or blocked on a long I/O operation (in millisecs) causes a caller to wait to see if a thread has died
- Example: threads/InterruptTest.java

A Thread's Life

A thread continues to execute until one of the following thing happens.

- it returns from its target `run()` method.
- it's interrupted by an uncaught exception.
- it's `stop()` method is called.

What happens if the `run()` method never terminates, and the application that started the thread never calls the `stop()` method?

The thread remains alive even after the application has finished!
(so the Java interpreter has to keep on running...)

Daemon Threads

- Useful for simple, periodic tasks in an application.
- The `setDaemon()` method marks a thread as a daemon thread that should be killed and discarded when no other application threads remain.
- Code snippet:

```
class Devil extends Thread {
    Devil() {
        setDaemon( true);
        start();
    }
    public void run() {
        //perform evil tasks in the background
        ...
    }
}
```

- Virtual threads are created as daemon threads.

Thread Synchronization (1)

- Java threads are preemptible. Java threads may or may not be time-sliced. We should not make any timing assumptions.
- Threads have priorities that can be changed (increased or decreased).
- This implies that multiple threads will have race conditions (read/write conflicts based on time of access) when they run.
- We have to resolve these conflicts with proper design and implementation.
- Example of a race condition: Account.java, TestAccount.java

Thread Synchronization (2)

- Java has `synchronized` keyword for guaranteeing mutually exclusive access to a method or a block of code. Only one thread can be active among all synchronized methods and synchronized blocks of code in a class. For example:

```
synchronized void update() { ... }
```

- Every Java object has an implicit monitor associated with it to implement the synchronized keyword. Inner class has a separate monitor than the containing outer class.
- Access to a block of code (accessing shared data) can also be synchronized using the built-in monitor in an arbitrary or related object.

```
Object lockObject = new Object();

// The object lockObject can be used in several classes,
// enabling synchronization among methods from multiple classes.

// assume that count is a static variable shared among multiple
    objects
synchronized(lockObject) {
    count = count + 1;
}
```

- Java allows Rentrant Synchronization, that is, a thread can reacquire a lock it already owns. For example, a synchronized method can call another synchronized method.

Synchronization Example 1

- Example of a race condition: Account.java, TestAccount.java
- Thread safe version using `synchronized` keyword: SynchronizedAccount.java

Thread Synchronization (3)

- The `wait()` and `notify()` methods (of the `Object` class) allow a thread to give up its hold on a lock at an arbitrary point, and then wait for another thread to give it back before continuing.
- Another thread must call `notify()` for the waiting thread to wakeup. If there are other threads around, then there is no guarantee that the waiting thread gets the lock next. Starvation is a possibility. We can use an overloaded version of `wait()` that has a timeout.
- The method `notifyAll()` wakes up all waiting threads instead of just one waiting thread.

Example with wait()/notify()

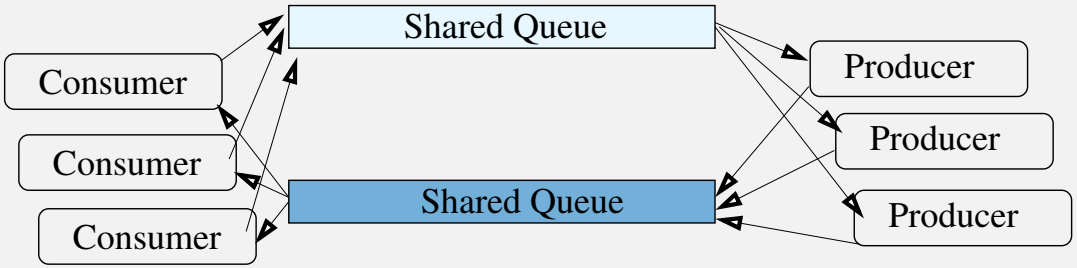
```
class MyThing {
  synchronized void waiterMethod() {
    // do something
    // Then we need to wait for the notifier to do something
    // The wait() gives up the lock, puts calling thread to sleep
    wait();
    // continue where we left off
  }

  synchronized void notifierMethod() {
    // do something
    // notify the waiter that we've done it
    notify();
    //do more things
  }

  synchronized void relatedMethod() {
    // do some related stuff
  }
}
```

Synchronization Example 2: Producer/Consumer Problem

- A *producer* thread produces objects and places them into a queue, while a *consumer* thread removes objects and consumes them.
- Often, the queue has a maximum depth.
- The producer and consumer don't operate at the same speed. We can also have multiple producers/consumers as well as multiple shared queues!



- Example: Suppose the producer creates messages every second but the consumer reads and displays only every two seconds. *How long will it take for the queue to fill up? What will happen when it does?*
- Example: SharedQueue.java, Producer.java, Consumer.java, PC.java
- The Producer/Consumer or a Thread Pool pattern is a widely used one for multi-threaded applications as well as in servers (as well as in more complex clients).

Synchronization Example 3: Ping Pong

- Example: PingPong.java. Will this work correctly? Why or why not?
- Example: SynchronizedPingPong.java. This solves the problem using `wait()` and `motify()` methods.
- Are the threads really simulating ping pong? We need them to exchange an object over the network!

Thread Pool: Executor (1)

- Tasks are logical units of work. Threads are a mechanism by which tasks can run asynchronously.
- Thread Pool: A number of threads are created to perform a number of tasks, which are organized in a queue. Typically, there are many more tasks than threads.
- Java provides a thead pool via the `Executor` interface in the `java.util.concurrent` package.

```
public interface Executor {
    void execute (Runnable command);
}
```

Thread Pool: Executor (2)

- We can use one of the factory methods in `Executors` in the package `java.util.concurrent` to create a thread pool. For example, the following creates a fixed size pool with `NTHREADS` threads with an unbounded queue size.

```
private static final Executor pool =
        Executors.newFixedThreadPool(NTHREADS);
```

- Replace `Executor` with `ExecutorService` to get more control on the thread pool.

```
private static final ExecutorService pool =
        Executors.newFixedThreadPool(NTHREADS);
```

- Example: ExecutorExample.java

Synchronization Wrappers (1)

- Many of the data structures in `java.util` package aren't synchronized (Why?). We can convert them into synchronized versions using wrappers.
- The synchronization wrappers add automatic synchronization (thread-safety) to an arbitrary collection. Each of the six core collection interfaces – Collection, Set, List, Map, SortedSet, and SortedMap – has one static factory method.

```
public static <T> Collection<T> synchronizedCollection(Collection<T> c);
public static <T> Set<T> synchronizedSet(Set<T> s);
public static <T> List<T> synchronizedList(List<T> list);
public static <K,V> Map<K,V> synchronizedMap(Map<K,V> m);
public static <T> SortedSet<T> synchronizedSortedSet(SortedSet<T> s);
public static <K,V> SortedMap<K,V> synchronizedSortedMap(SortedMap<K,V>
    m);
```

- Create the synchronized collection with the following trick.

```
List<Type> list = Collections.synchronizedList(new ArrayList<Type>());
```

A collection created in this fashion is every bit as thread-safe as a normally synchronized collection, such as a `Vector`.

Synchronization Wrappers (2)

- In the face of concurrent access, it is imperative that the user manually synchronize on the returned collection when iterating over it. The reason is that iteration is accomplished via multiple calls into the collection, which must be composed into a single atomic operation.

```
Collection<Type> c = Collections.synchronizedCollection(myCollection);
synchronized(c) {
        for (Type e: c)
                    process(e);
}
```

- If an explicit iterator is used, the iterator method must be called from within the synchronized block. Failure to follow this advice may result in nondeterministic behavior.
- One minor downside of using wrapper implementations is that we do not have the ability to execute any noninterface operations of a wrapped implementation.

For more details, see:
http://docs.oracle.com/javase/tutorial/collections/implementations/wrapper.html

Processes in Java

- We can create native processes using `ProcessBuilder` class and manage them with the `Process` class. This is the equivalent of *fork-exec* from C/Linux.

```
ProcessBuilder builder =
        new ProcessBuilder("myCommand", "myArg");
Process p = builder.start();
p.waitFor(); //if we want to wait for it to finish
```

- Use the methods `getInputStream()`, `getOutputStream()` and `getErrorStream()` to manage streams to the child process.
- Example: ProcessExample.java, MaxProcesses.java

Exercises

1. Write a multi-threaded Java program that computes n random numbers in the range $[0 \ldots 1]$ using *nthreads* threads, where n and *nthreads* are supplied on the command line. What kind of speedup do we get with multiple threads?
2. Rewrite the above program using the `Executor` interface. Do we get the same speedup?
3. Rewrite the first program using the virtual threads? Do we get better speedup? Why or why not?
4. Rewrite the ExecutorExample.java to use threads directly instead of using the `Executor`. Make sure to write code to limit the number of threads that are running.
5. The `Vector` class provided in Java library is synchronized. We could have used that instead of `SharedQueue` in the producer consumer code. What is the advantage of writing our own `SharedQueue` class?
6. Rewrite the SharedQueue.java such that it is generic. Rerun the producer/consumer example with our generic queue.

References

- Javadocs for `java.lang.Thread`, `java.lang.Runnable`, `java.util.concurrent` and related packages
- Brian Goetz, Tim Peierls, Joshua Bloch and Joseph Bowbeer: *Java Concurrency in Practice*
- `https://www.baeldung.com/java-concurrency`