

Build a web scraper based on virtual threads!



X #jfall

@David_Vlijmincx

www.davidvlijmincx.com | Oracle ACE



@ronveen

www.linkedin.com/in/ron-veen | 20+ YOE

Outline



- Set up
- Threads and virtual threads
- Thread performance (How fast can you scrape)
- Structured concurrency
- Scoped values
- Show your Scraper (If you want)

What is a web scraper



- A bot
- Downloads web pages to find new web pages
- GOTO 1

What is a web scraper



```
Crawled 2000 web page(s)Total execution time: 1605ms  
Throughput: 1246.1059190031153 pages/sec  
  
Process finished with exit code 0
```

Set up



- SSID: HANDS-ON
- WIFI password: jfall2023

What You need:

- https://github.com/davidtos/virtual_thread_workshop
- IDE
- Java 21

(Platform) Threads

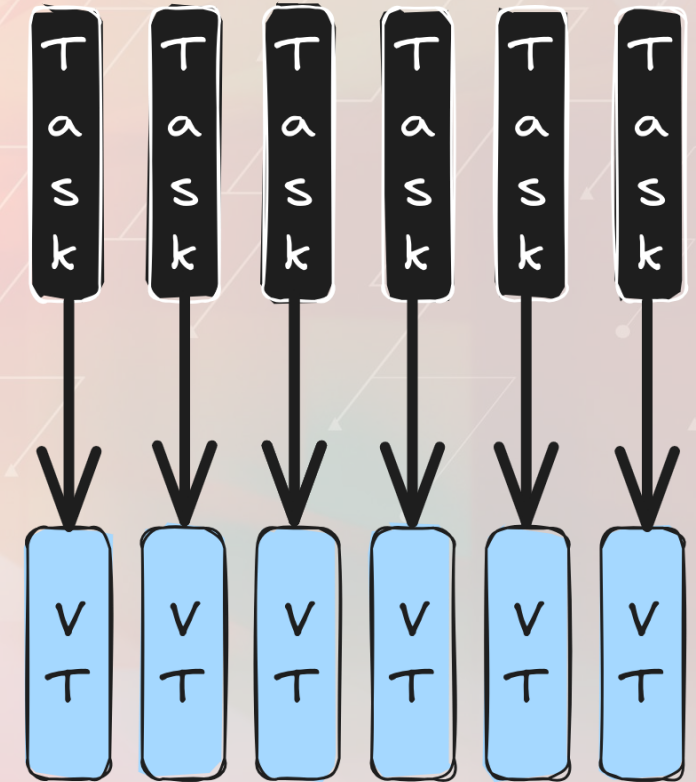


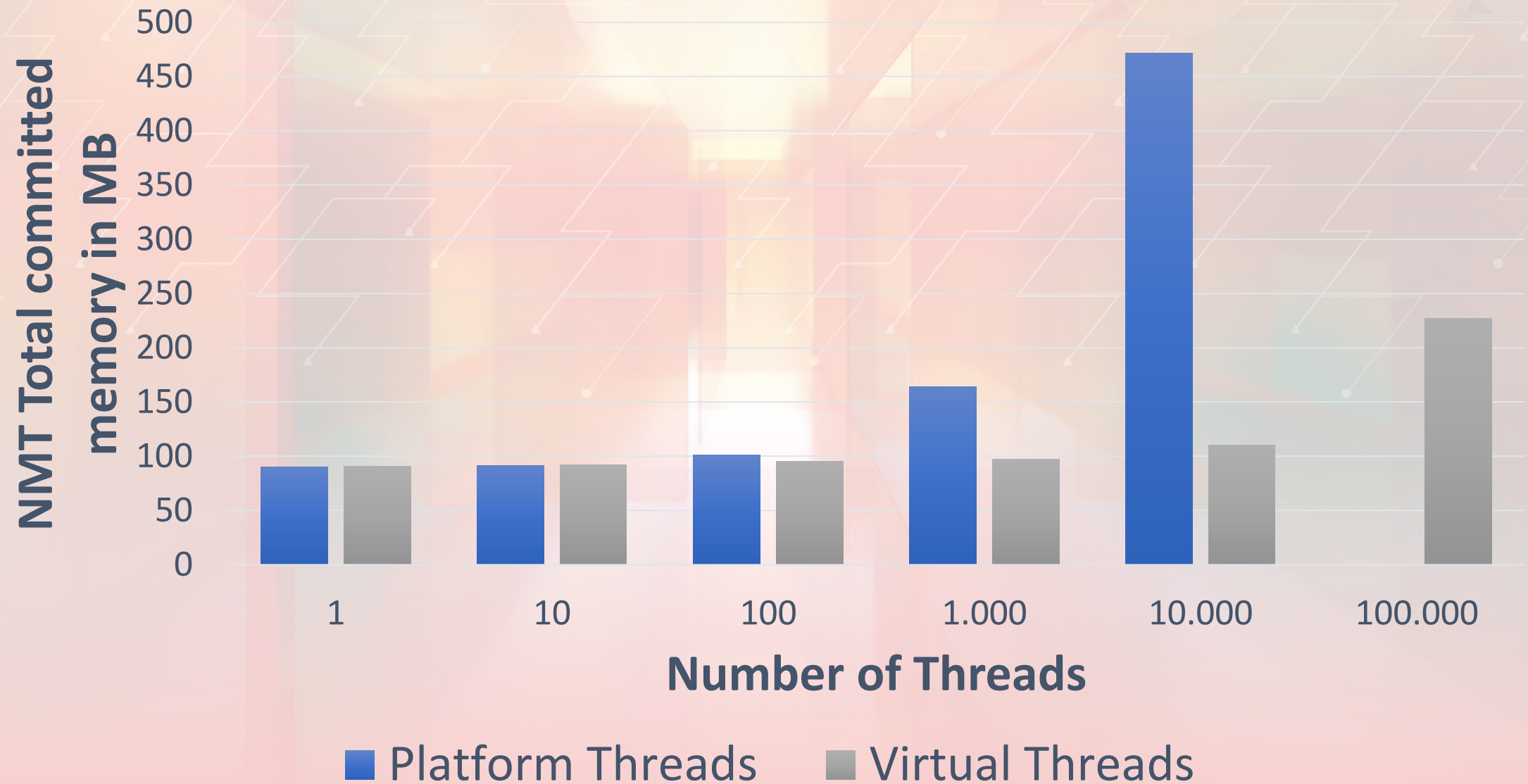
```
var thread = new Thread(task);  
thread.start();
```

```
var executor = Executors.newFixedThreadPool(  
    Runtime.getRuntime().availableProcessors()  
)
```

Virtual Threads

- Alternative implementation of Threads
- The OS does not know about them
 - Concept inside the JVM
 - Stack frames live on the heap
 - Resizable stacks
- You don't have to allocate a lot of memory at the start
- Cheap to create





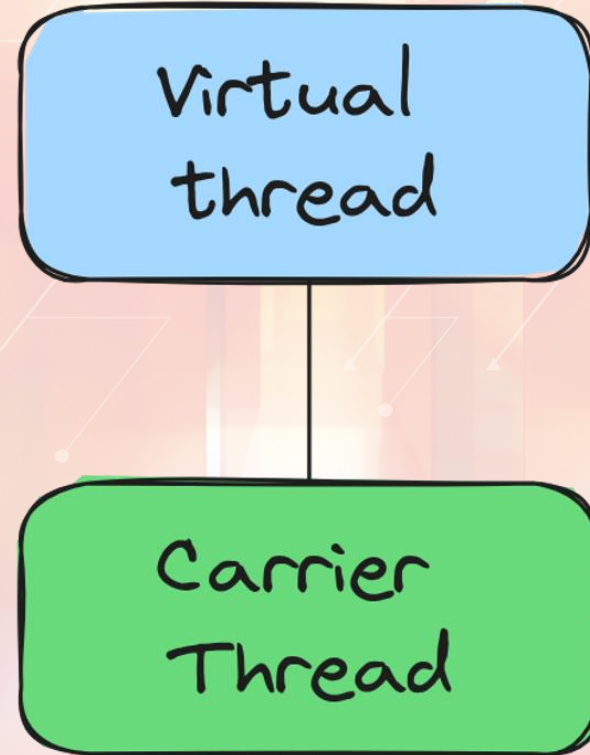
Virtual Threads



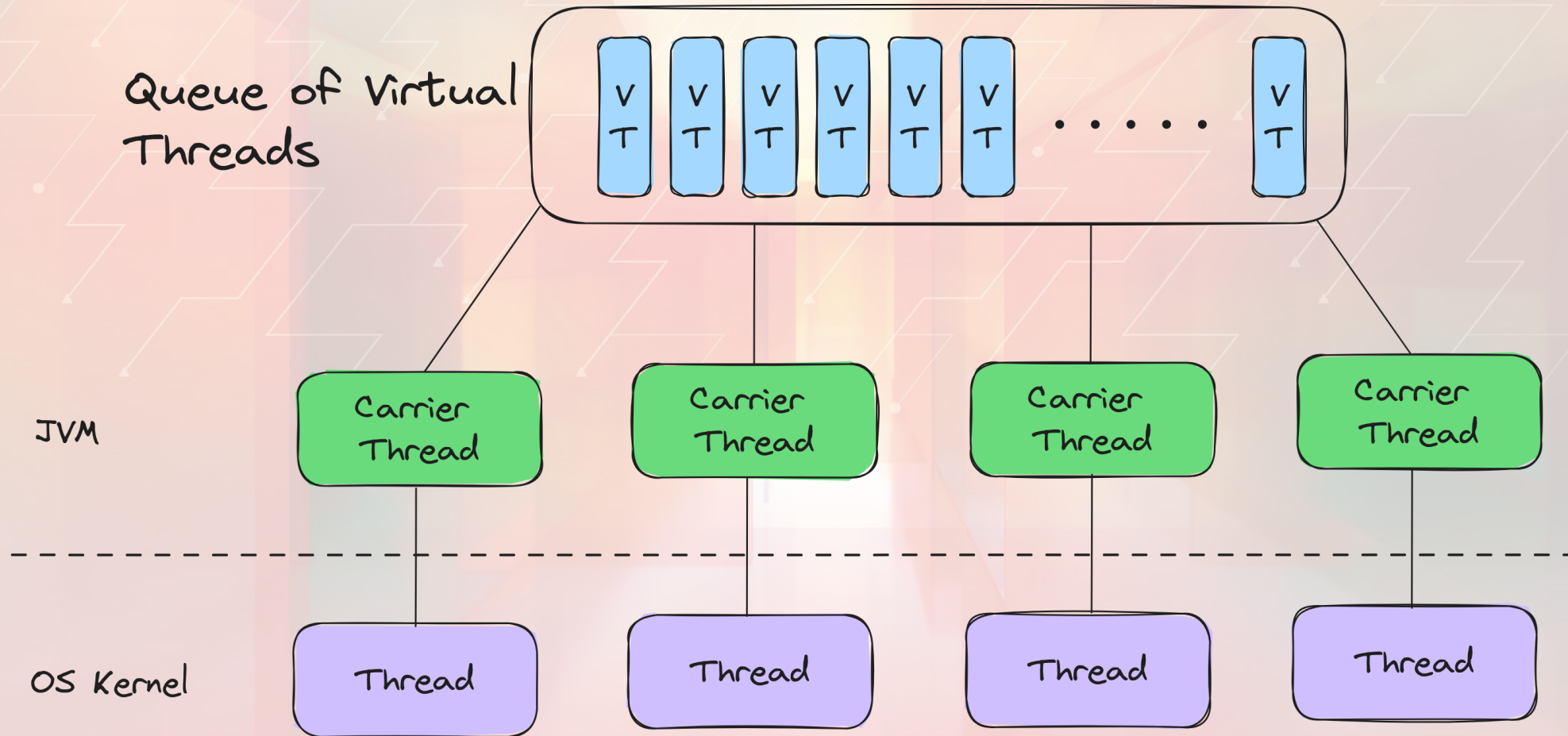
```
Thread.startVirtualThread(RunnableTask);
```

```
Executors.newVirtualThreadPerTaskExecutor()
```

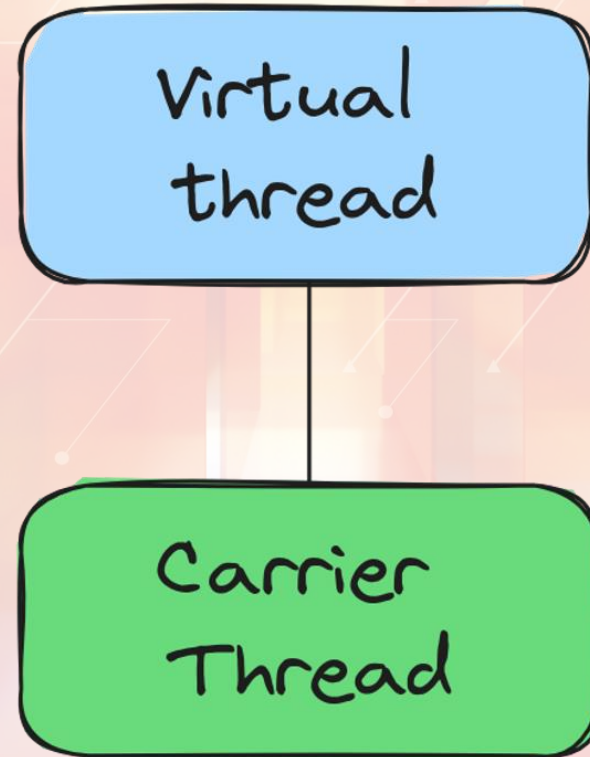
How does it work



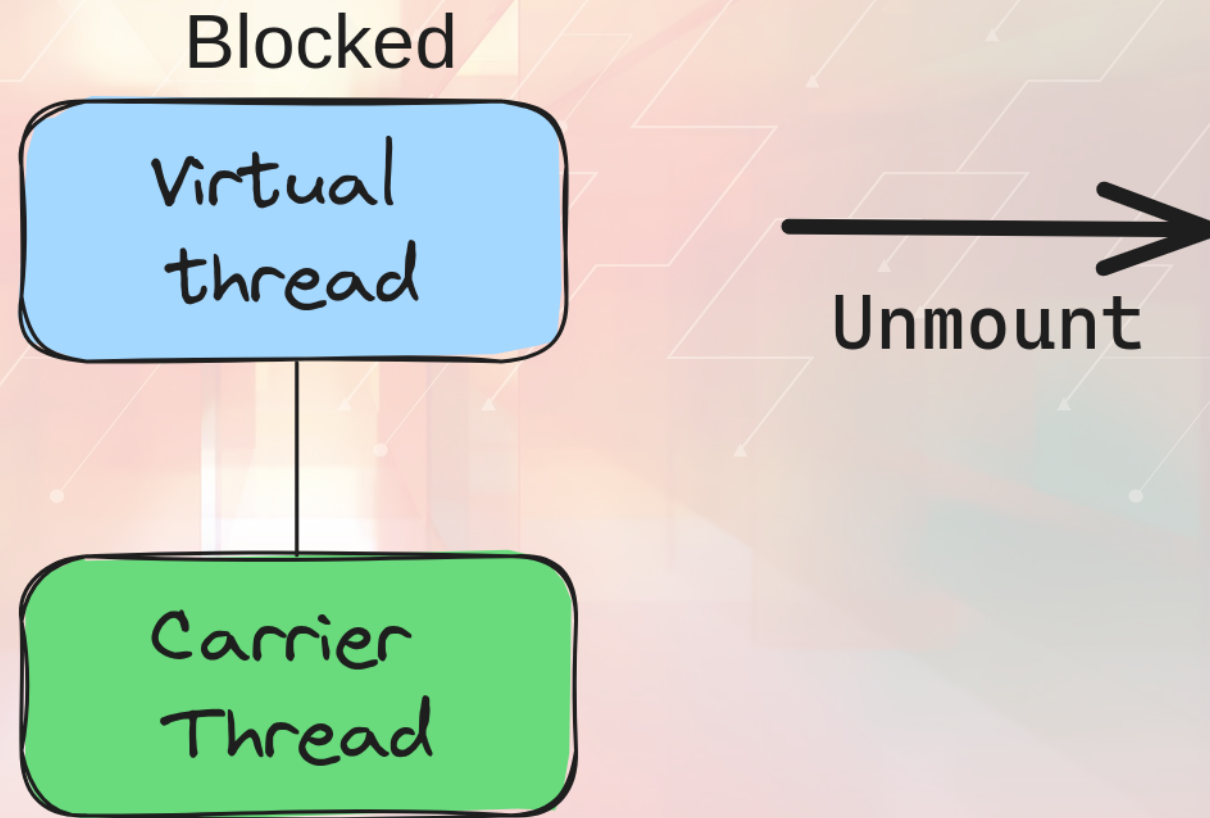
How does it work



Mounting and unmounting

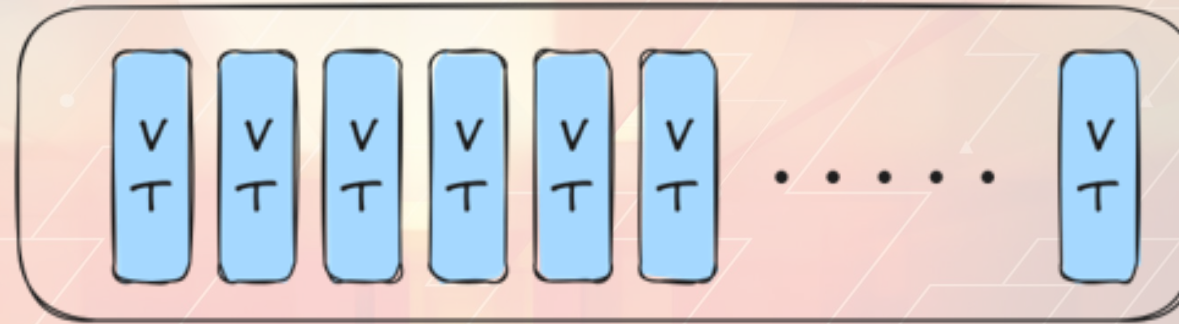


Mounting and unmounting



Mounting and unmounting

Queue of Virtual
Threads

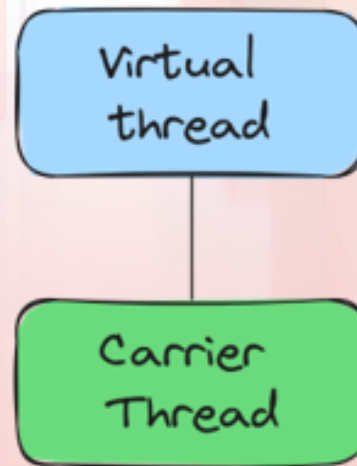


Mount

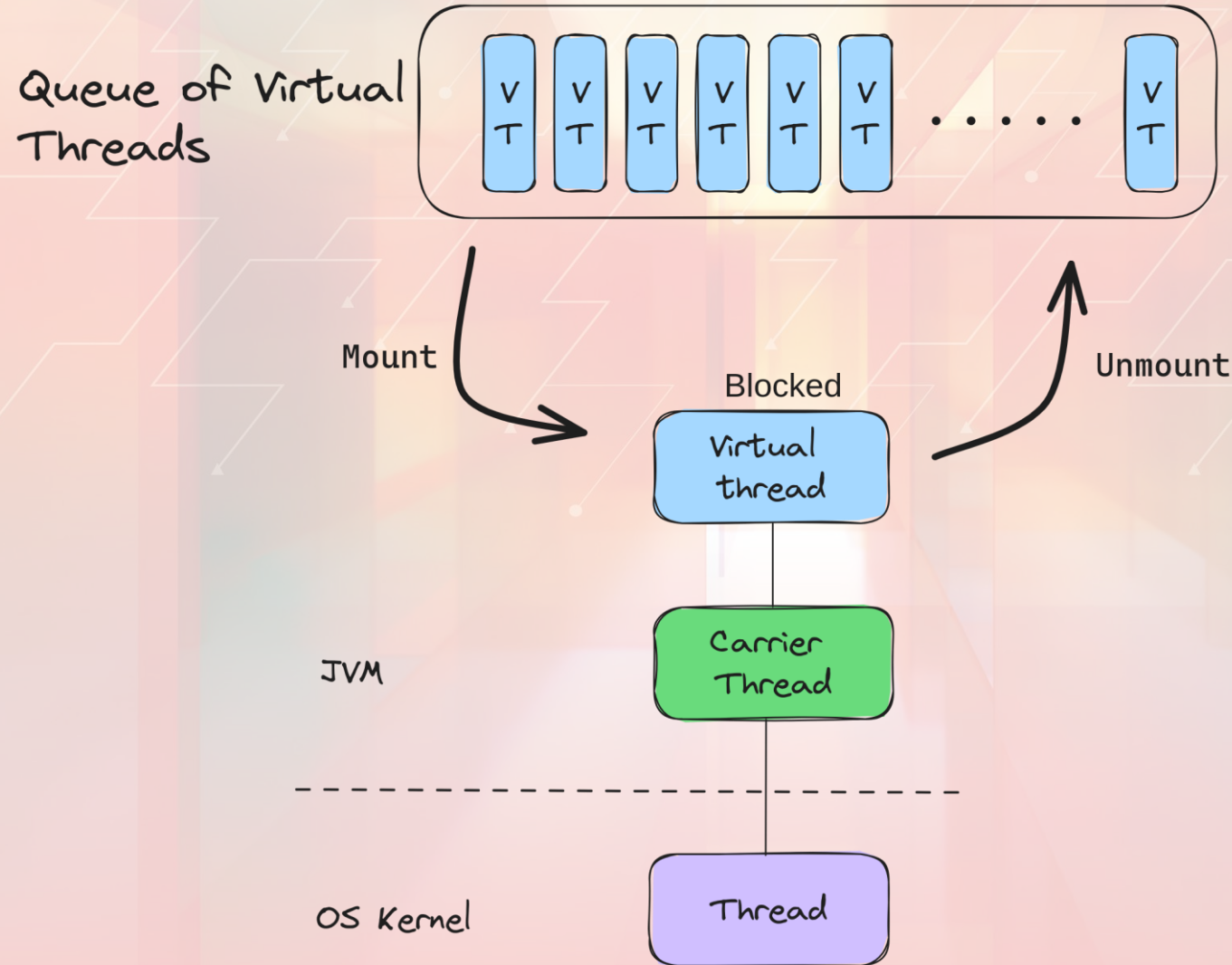


Virtual
thread

Carrier
Thread



Mounting and unmounting



See pinned virtual threads



Run option:

`-Djdk.tracePinnedThreads=short`

System property:

`jdk.tracePinnedThreads=full`

Not every method is ready



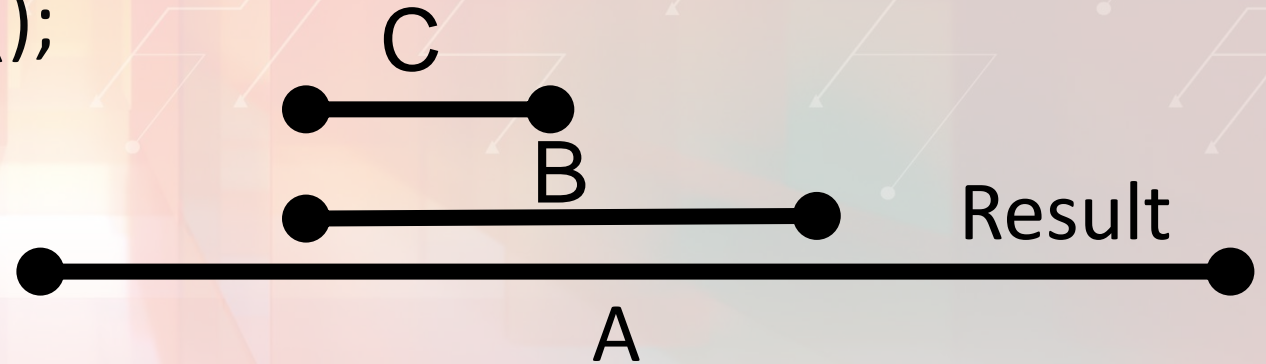
- Synchronized Block
- `object.wait()`
- Calls to native code (JNI)
- Certain filesystem operations

Structured concurrency



Implicit relation

```
static public String methodA() throws ... {  
    Future<String> foo = methodB();  
    Future<String> bar = methodC();  
  
    return foo.get() + bar.get();  
}
```



Structured concurrency



```
try (var scope = new StructuredTaskScope.ShutdownOnFailure()) {}  
  
try (... = new StructuredTaskScope.ShutdownOnSuccess<>()) {}
```


Creating a Scope



```
class Scrape implements Runnable {  
    final static ScopedValue<HttpClient> CLIENT =  
        ScopedValue.newInstance();  
    ....  
}
```

Creating a Scope




```
ScopedValue.runWhere(Scrape.CLIENT,  
    client,  
    new Scrape(queue, visited))
```

Someone want to show their
Scraper?

Thanks for your attention

Please rate my session in the J-Fall app



 #jfall