

이상준

(현) 카카오 엔터테인먼트

(전) 나이스평가정보

(전) 우리에프아이에스

(전) 롯데정보통신

(전) 플레이스5



T.O.C

1. **Motivation**
2. 기존 자바 **Thread** 문제점
3. 리액티브 프로그래밍 & **Spring Webflux**
4. **What is Virtual Thread & 구조**
5. **Performance Comparison**
6. **Conclusion**

Spring 기반 신규 프로젝트

Spring Webflux 쓸 때가 되었나?

Kotlin Coroutine으로 할까?

오!!!! Virtual Thread 정식 출시?! 그럼 Webflux 안써도 되는거 아냐?

하던데로 Spring MVC + Virtual Thread 로 할까?

일단 성능 비교 ㅋㅋ!

```
Text('Section Title',  
  style: TextStyle(  
    color: Colors.yellow[200],  
  ),  
),  
),  
s.star,  
r: Colors.yellow[500],  
Text('23'),
```

devfest



기존 자바 Thread 문제점

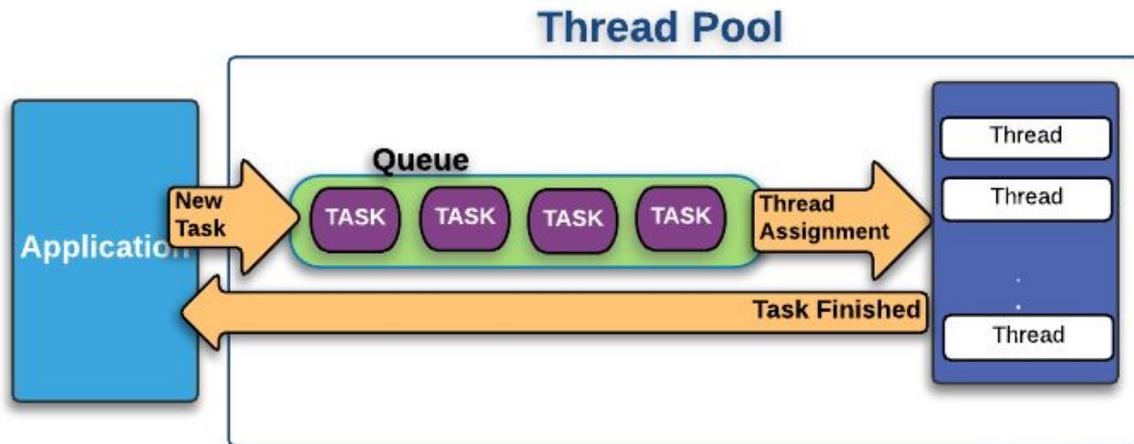
기존 자바의 Thread

기존 자바의 스레드는 실제 OS의 Thread와 1:1 구조

OS Thread는 갯수가 제한적이고 생성하고 유지하는 비용이 비쌘

생성과 유지비가 싸니 pool을 만들어서 재활용 (ex. db connection pool)

Tomcat default max thread size : 200



LOGICBIG.COM

Thread Per Request 모델

- Apache Httpd, Apache Tomcat
- 요청 1개당 1개의 Thread 사용
- Throughput을 늘리려면 Thread가 많아야
- OS Thread 무작정 늘릴 수 없음.

```
mbio@gdg-devfest-1:~$ uname -a
Linux gdg-devfest-1 6.2.0-1019-gcp #21~22.04.1-Ubuntu SMP Thu Nov 16 18:18:34 UTC 2023 x86_64 x86_64 x86_64 GNU/Linux
mbio@gdg-devfest-1:~$ grep -c processor /proc/cpuinfo
4
mbio@gdg-devfest-1:~$ cat /proc/meminfo | grep MemTotal
MemTotal:      16365492 kB
mbio@gdg-devfest-1:~$ cat /proc/sys/kernel/thread-max
127790
mbio@gdg-devfest-1:~$
```

GCP n2d-standard-4 (4코어, 16GB, 10G) 장비의 thread-max 사이즈

C10K Problem

그 유명한 C10K Problem (1999년 Dan Kegel)

Nonblocking I/O, Asynchronous I/O

요청을 처리하는 과정을 뜯어보면 **Blocking I/O**에서 **wait**에 대부분 시간을 소요

Ex. 먹는시간 > 대기시간



해결 방법

1. 장비 업그레이드

- 너무 비싸고, 가성비가 안나옴
- **OS Thread**는 동시에 **core** 수 만큼만 처리가능함
- **OS Thread**를 마구 늘리면 오히려 **context switching** 비용만 비싸짐

2. Event Driven

- 요청(=이벤트) 과 **I/O**처리 분리
- **I/O Multiplexing** 기반 싱글스레드 이벤트 루프
- **Blocking I/O**를 기다리는 시간 없이 요청 처리

Event Driven - I/O Multiplexing

커널레벨 프로그램

- **select(), poll(), kqueue(), epoll()**

크로스 플랫폼 라이브러리

- **libuv (bsd계열 kqueue, linux계열 epoll, SunOS poll ...)**

Python - FastAPI(uvloop)

Javascript - Node.js, nginx

JAVA - NIO, Netty

Redis

```
Text('Section Title',  
  style: TextStyle(  
    color: Colors.yellow[200],  
  ),  
),  
),  
s.star,  
r: Colors.yellow[500],  
Text('23'),
```

devfest

 Google Developer Groups
Incheon/Songdo

리액티브 프로그래밍 & Spring Webflux

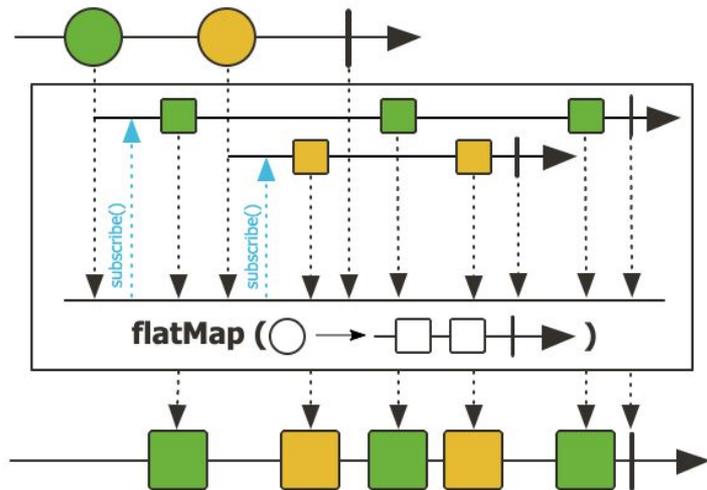
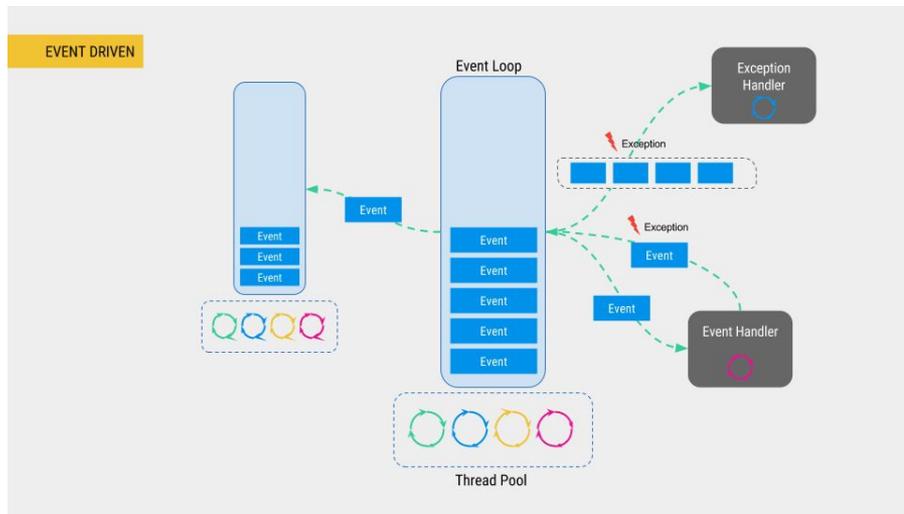
Spring Webflux

Event-Driven, Event loop

Spring 기반 웹 프레임워크도 C10K Problem 이슈 해결!

Reactive Programming 패러다임을 따르는 Reactive Streams의 등장

Reactive Streams : Event Driven 을 구현하기 위한 API



리액티브 프로그래밍

In **computing**, **reactive programming** is a **declarative programming paradigm** concerned with **data streams** and the propagation of change. With this paradigm, it's possible to

Programming paradigms

- Action

-> 위키피디아: [데이터 스트림과 변경사항의 전파에 중점을 둔 선언적 프로그래밍 패러다임](#)

자바진영의 리액티브 프로그래밍 구현체 **RxJava, Reactor**

Reactor기반의 스프링 웹프로젝트가 **Spring Webflux**

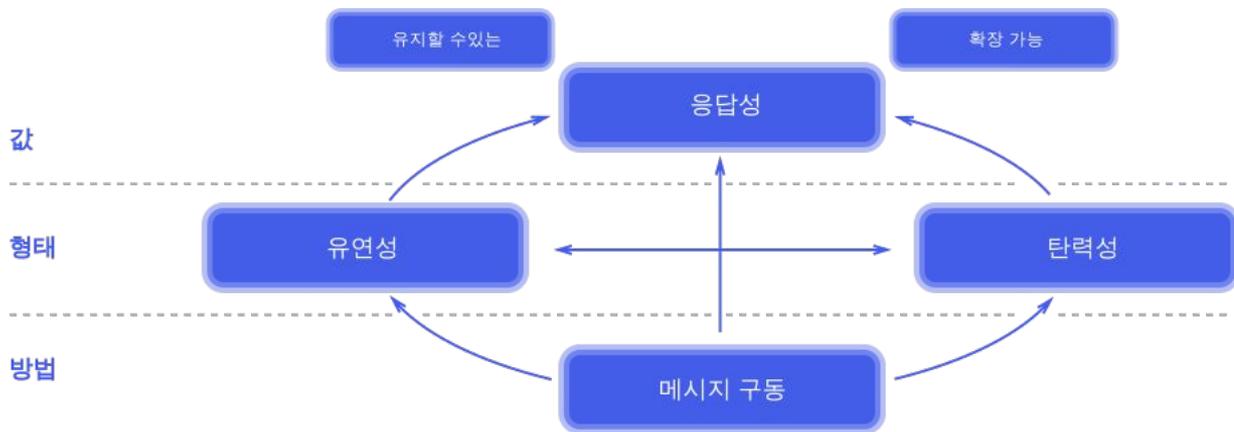
그외 **Akka, Vert.x, Quassar ...**

Reactive Programming

데이터 스트림과 변경사항의 전파에 중점을 둔 선언적 프로그래밍 패러다임

명령형 -> 선언형 / 순차형 -> 동시성

리액티브 선언문



Reactive Programming

반응형 프로그래밍의 대표 - 스프레드 시트

- 선언적으로 함수를 선언

- **A1, B1**의 값이 변경되면 변경사항이 전파됨

- **C1**의 값이 자동으로 변경 -> 전파 -> **D1** 값 변경

C1 fx =SUM(A1+B1)				
	A	B	C	D
1	2	3	5	8
2				

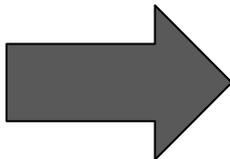
D1 fx =SUM(C1+3)				
	A	B	C	D
1	2	3	5	8
2				

선언형 프로그래밍

원하는 결과를 프로그래밍

언제 어떤 방식으로 결과가 나올지는 추상화

```
🍏 # ~ kubectl run nginx --image=nginx
🍏 # ~ kubectl create deployment nginx --image=nginx
🍏 # ~ kubectl expose deployment nginx --port=80
🍏 # ~ kubectl edit deployment nginx
🍏 # ~ kubectl set image deployment nginx nginx=nginx:1.25
🍏 # ~ kubectl delete -f nginx.yaml
```



```
23%
19 apiVersion: apps/v1
18 kind: Deployment
17 metadata:
16   name: nginx-deployment
15   labels:
14     app: nginx
13 spec:
12   selector:
11     matchLabels:
10       app: webapp
9     template:
8       metadata:
7         labels:
6           app: webapp
5       spec:
4         containers:
3           - name: nginx
2             image: nginx:1.25
1             ports:
20
```

변경사항 전파 (이벤트, 메세징)

Pull -> Push

동기적인 요청과 응답대기에서 벗어남

데이터의 변경과 관련된 **이벤트 발행 -> 이벤트 구독**

앞에서본 **I/O Multiplexing**

Flutter Provider, React conext api

아키 레이어로 가면 **MQ**기반 이벤트 드라이븐

Flutter Provider / React Context API

```
class Person with ChangeNotifier {
  Person(this.name, this.age);
  final String name;
  int age;
  void increaseAge() {
    this.age++;
    notifyListeners();
  }
}

void main() {
  runApp(ChangeNotifierProvider(create: (_) => Person(name: "mbio", age: 32), child: MyApp(),));
}

class MyHomePage extends StatelessWidget {
  const MyHomePage({Key key}) : super(key: key);
  @override
  Widget build(BuildContext context) {
    return Scaffold(
      appBar: AppBar(
        title: const Text('Provider Class'),
      ),
      body: Center(
        child: Text(
          '...',
          name: '${Provider.of<Person>(context).name};!',
          age: '${Provider.of<Person>(context).age}',
        ),
      ),
      floatingActionButton: FloatingActionButton(
        onPressed: () => Provider.of<Person>(context, listen: false).increaseAge(),
      ),
    );
  }
}
```

```
export const PersonContext = createContext({
  name: '',
  age: 0,
  increaseAge: () => {},
});

const PersonProvider = ({ children }) => {
  const [name, setName] = useState("mbio");
  const [age, setAge] = useState(32);
  const increaseAge = () => setAge(age => age + 1);
  return (<PersonContext.Provider value={{ name, age, increaseAge }}>{children}</PersonContext.Provider>);
};

const MyHomePage = () => {
  const { name, age, increaseAge } = useContext(PersonContext);

  return (
    <div>
      <h1>Provider Class</h1>
      Name: <b>{name}</b>
      Age: <b>{age}</b>
      <button onClick={increaseAge}>Increase Age</button>
    </div>
  );
};

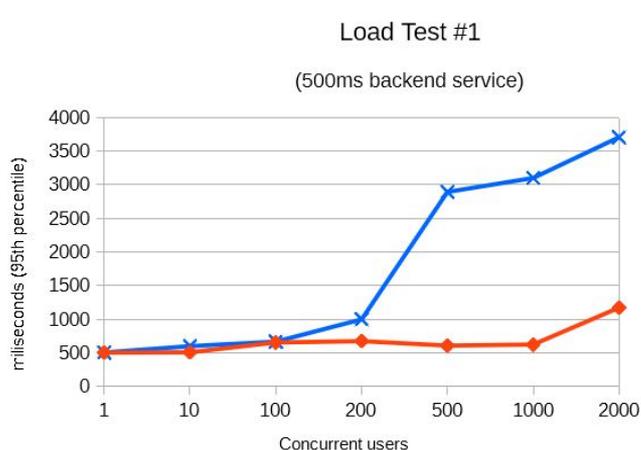
function App() {
  return (
    <PersonProvider>
      <MyHomePage />
    </PersonProvider>
  );
}
```

What is Webflux Problem

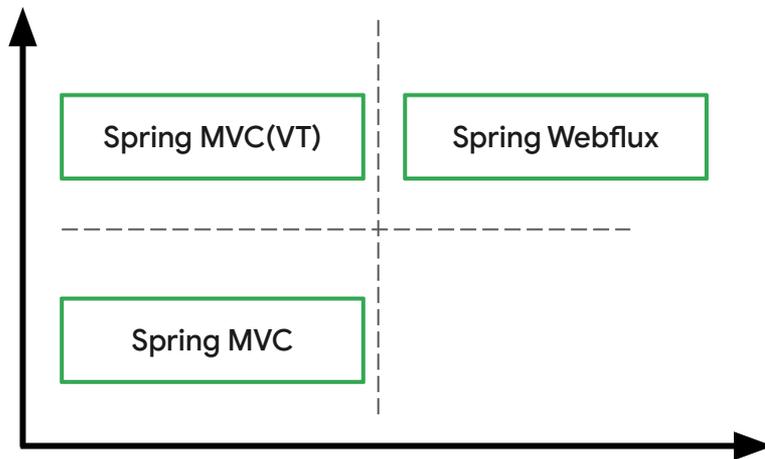
- **C10K** 해결, 리액티브 프로그래밍 스타일도 맘에 드는데 왜?
- 코드 어려움
- **Thread**를 중심으로 하는 **Java**의 디자인과 어울리지 않음
- **Exception Stack Trace, Debugger** 등이 전부 스레드 기반
- 1개의 요청을 처리하는 1개의 스레드를 추적하는 개념으로 트레이싱 했었는데 스레드가 계속 바뀌니 추적이 안됨

There is no silver bullet

- **reactive streams**가 등장하고 **Webflux** 같은 프레임워크가 생긴건 사실 자바가 비동기 논블록킹 지원을 잘 안해줘서 (**kotlin coroutine**도 마찬가지)
- **Blocking** 시점에 기다리지 않게 스케줄링을 하는게 쟁점인데 개발자가 직접 해야함
- **Framework** 같은걸 왜 쓰는데? 비즈니스 로직에 집중



Throughput



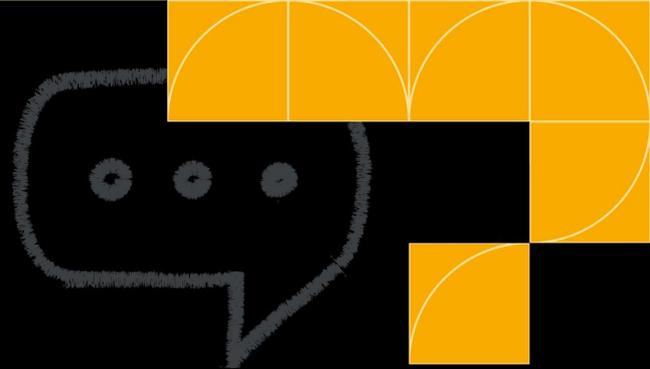
코드 난이도

(=인건비)


```
Text('Simple Statement or URL',  
style: TextStyle(  
  color: Colors.yellow[200],  
),  
),  
),
```

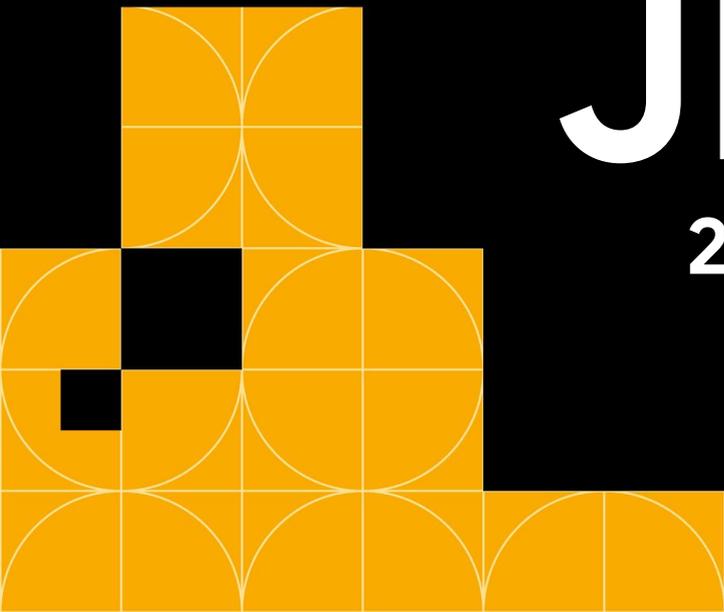
devfest

```
s.star,  
r: Colors.yellow[500],  
Text('23'),
```



JDK 21

2023.09.19 LTS



Virtual Thread가 뭐예요?

JDK 21에 정식 출시된 **경량(lightweight) 쓰레드**

Project Loom의 결과물 (**thread**보다 작은 단위의 **task** 지원)

JVM이 동작하는 **OS Thread**를 사용하지 않음

Virtual Machine에서 자체적으로 스케줄링하여

수백만개의 쓰레드를 사용할 수 있게 해줌

User Thread vs Kernel Thread

User Thread

- **Lightweight thread**, Green thread, Virtual Thread, Fiber Thread

Kernel Thread

- **OS Thread**, Platform Thread, Carrier Thread

가상 쓰레드는 새로운 개념이 아님

Golang Goroutine = “A goroutine is a **lightweight thread** managed by the Go runtime. “

Haskell GHC, Erlang, Elixir ... High Throughput으로 유명한 언어들은 이전부터 제공

자바의 Thread

Green Thread (싱글코어 시절에 설계)

- JVM 1.1 ~ 1.3
- Java Threads: Native Thread = M:1

Platform Thread

- JVM 1.3~
- Java Thread: Native Thread = 1:1

Virtual Thread

- JVM 21~
- Java Threads: Native Threads = M:N (M > N)

Spring Boot 3.2.0 - 2023.11.23 Release

JDK 21 Virtual Thread 공식 지원

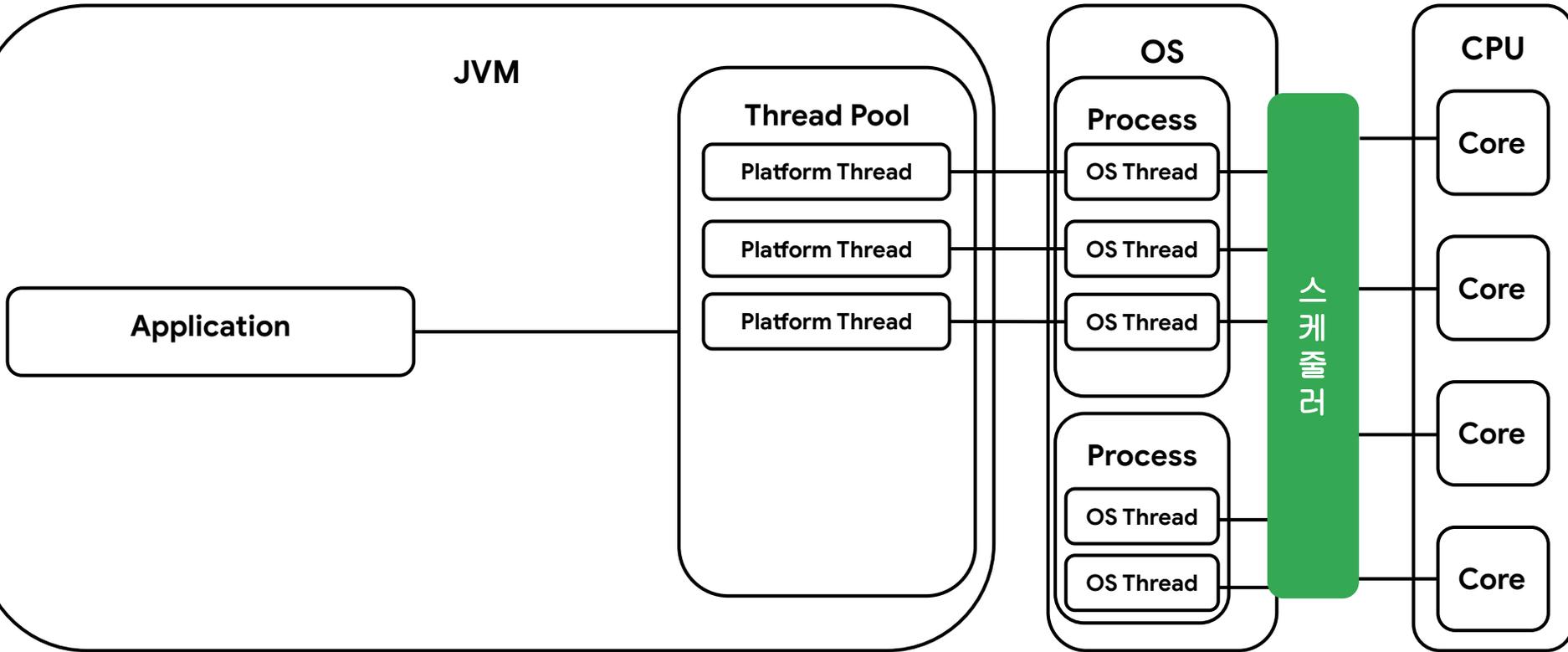
사용방법

- `spring.threads.virtual.enabled=true`

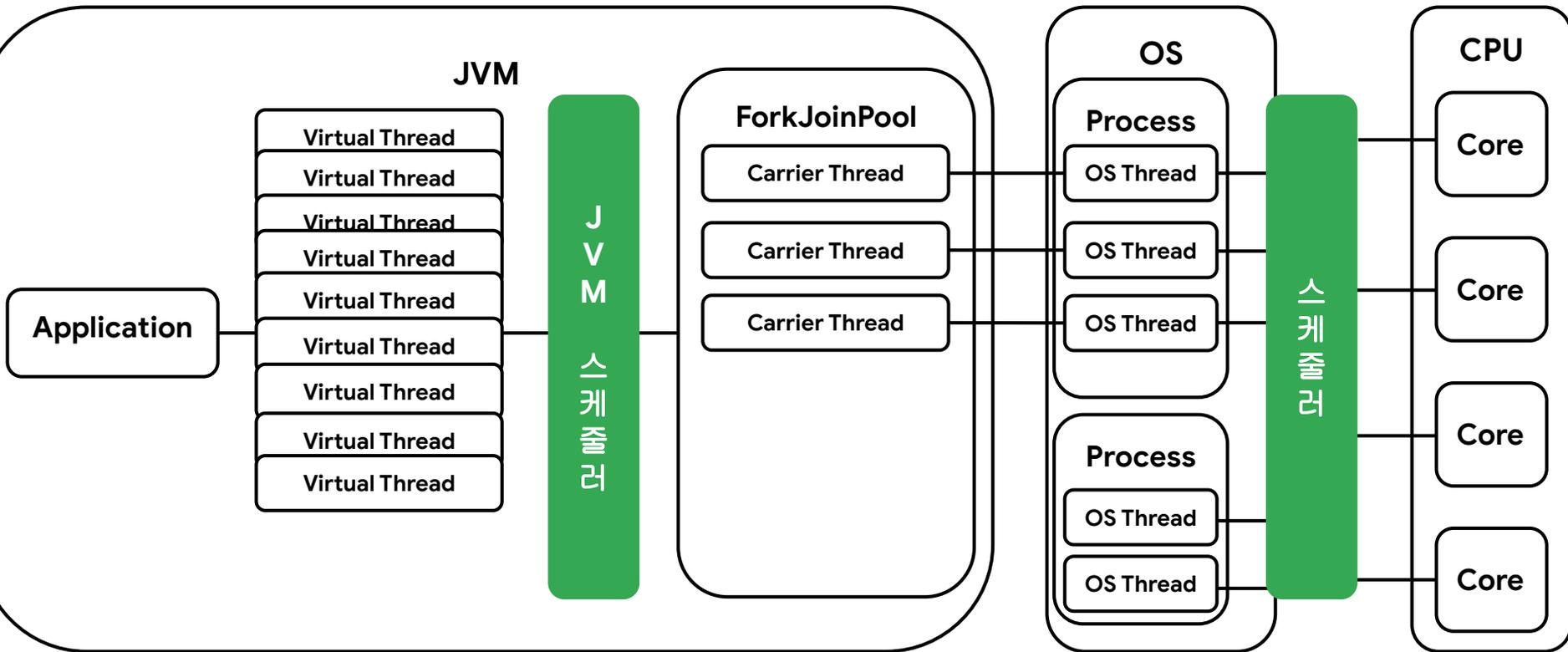
Thread.currentThread().toString()

- **Spring MVC**
Thread[#43,http-nio-8080-exec-1,5,main]
- **Spring MVC- Virtual Thread**
VirtualThread[#58,tomcat-handler-0]/runnable@ForkJoinPool-1-worker-1
- **Spring Webflux**
Thread[#68,reactor-http-nio-3,5,main]

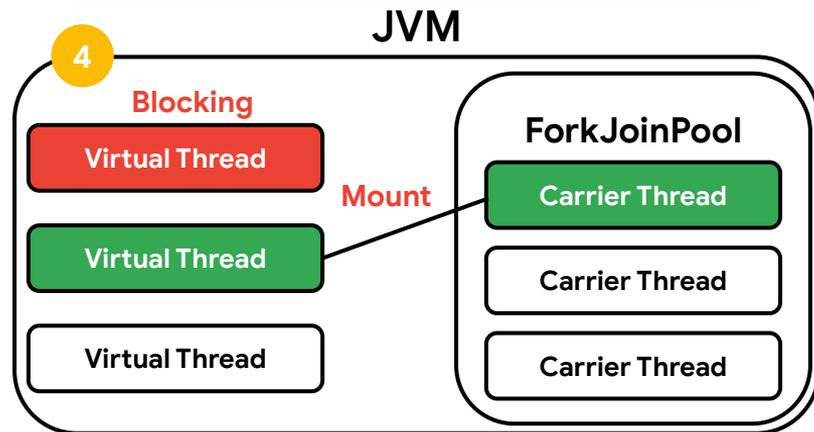
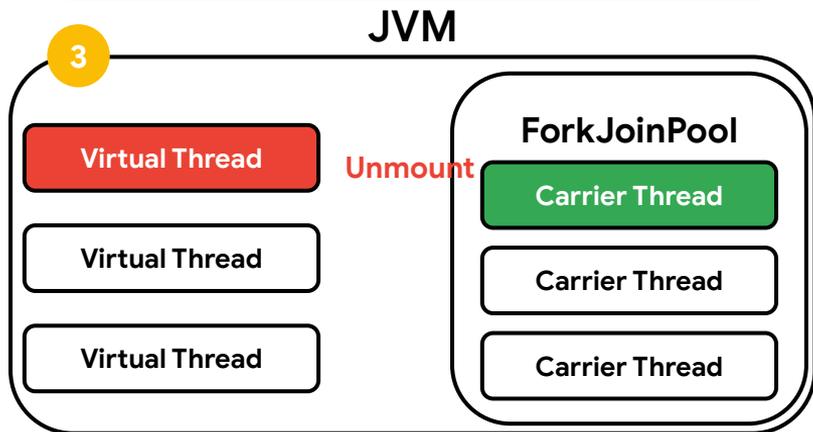
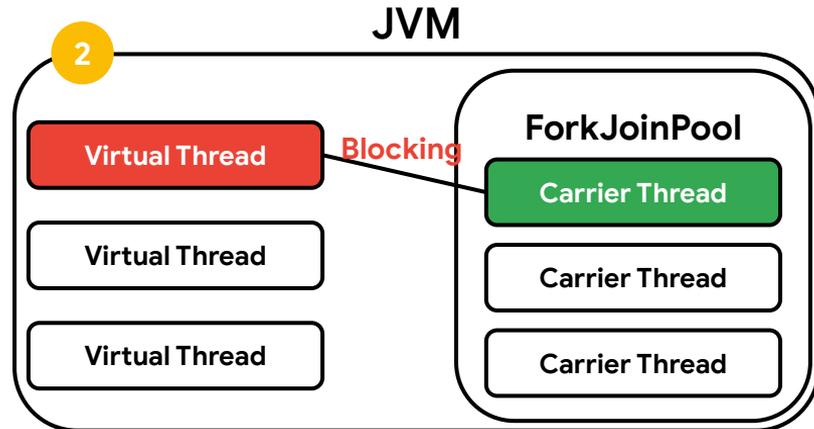
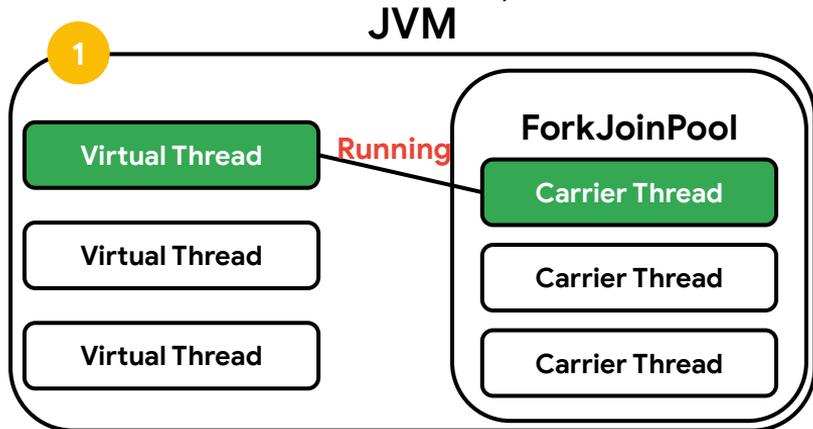
Virtual Thread (Fiber Thread)



Virtual Thread (Fiber Thread)



Virtual Thread (Fiber Thread)



Carrier Thread

기존 **OS Thread**와 **1:1**로 매핑되던 **Platform Thread**

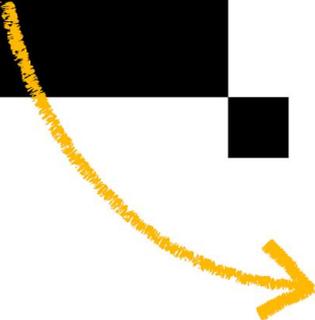
Virtual Thread에서는 **OS**와 매칭되는 쓰레드의 이름이 **Carrier Thread**로 변경

ForkJoinPool 안에서 동작하는 **Worker Thread**가 **Carrier Thread**

- **CPU** 코어 갯수 만큼 생성
- **Available Processor** 만큼
- **Like, Webflux**


```
Text('Section Title',  
style: TextStyle(  
  color: Colors.yellow[200],  
),  
),  
),  
s.star,  
r: Colors.yellow[500],  
Text('23'),
```

devfest



 Google Developer Groups
Incheon/Songdo

Performance Test

성능 테스트 환경과 툴

Grafana K6

- VUser 3000, Duration 30s, ResponseTime (P90)

Java 실행 환경

- Google Cloud, n2d-standard-4 (4코어, 16GB, 7Gbps)
- MaxHeap 4GB
- Tomcat default max thread size : 200

Database 별도 서버

- Google Cloud SQL - Mysql 8.0.35
- Enterprise-개발설정
- Connection pool : 150

요약

리전	asia-northeast3(서울)
DB 버전	MySQL 8.0.35
vCPU	4 vCPU
메모리	16GB
데이터 캐시	사용 중지됨
스토리지	100GB
연결	공개 IP
백업	자동
가용성	단일 영역
point-in-time recovery	사용 설정됨
네트워크 처리량(MB/초)	1,000/1,000
?	
디스크 처리량(MB/초) ?	읽기: 48.0/240.0 쓰기: 48.0/240.0
IOPS ?	읽기: 3,000/15,000 쓰기: 3,000/15,000

Cloud SQL 인스턴스

성능 테스트 Application & API

성능 테스트 Application

- **JDK8 + Spring Boot 2.7.13 (Jetbrain 조사결과 아직 JDK8 1등)**
- **JDK21 + Spring Boot 3.2.0 + Virtual Thread**
- **JDK21 + Spring Boot 3.2.0 + Platform Thread**
- **JDK21 + Spring Boot 3.2.0 + webflux**

성능 테스트 API

- 단순 문자열 반환
- **1초 blocking API**
- **1초 blocking + 고속 select**
- **1초 응답 쿼리 (slow query)**

성능 테스트 API

Spring MVC

```
@RestController
@RequiredArgsConstructor
@Slf4j
public class HelloController {
    private final JdbcTemplate jdbcTemplate;

    new *
    @GetMapping("/{thread-name}")
    public String getThreadName() {
        return Thread.currentThread().toString();
    }

    new *
    @GetMapping("/{block}")
    public String block() throws InterruptedException {
        Thread.sleep(1000);
        return Thread.currentThread().toString();
    }

    new *
    @GetMapping("/{block/query}")
    public String blockingAndQuery() throws InterruptedException {
        Thread.sleep(1000);
        jdbcTemplate.queryForList( sql: "select now()");
        return Thread.currentThread().toString();
    }

    new *
    @GetMapping("/{slow-query}")
    public String slowQuery() {
        jdbcTemplate.queryForList( sql: "select sleep(1);");
        return Thread.currentThread().toString();
    }
}
```

Spring Webflux

```
@RestController
@RequiredArgsConstructor
@Slf4j
public class HelloController {
    private final DatabaseClient databaseClient;

    new *
    @GetMapping("/{thread-name}")
    public Mono<String> getThreadName() {
        return Mono.just(Thread.currentThread().toString());
    }

    new *
    @GetMapping("/{block}")
    public Mono<String> block() {
        return Mono.delay(Duration.ofSeconds(1)).then(Mono.just(Thread.currentThread().toString()));
    }

    new *
    @GetMapping("/{block/query}")
    public Mono<String> blockingAndQuery() {
        return Mono.delay(Duration.ofSeconds(1))
            .then(databaseClient.sql("select now();")
                .map(row -> Thread.currentThread().toString()).first());
    }

    new *
    @GetMapping("/{slow-query}")
    public Mono<String> query() {
        return databaseClient.sql("select sleep(1);")
            .map(row -> Thread.currentThread().toString()).first();
    }
}
```

단순 문자열 반환

MVC-VT 회차	TPS	Response Time
1	10198.4	527 ms
2	11211.9	510 ms
3	10836.8	522 ms

MVC-PT 회차	TPS	Response Time
1	10906.2	517.17 ms
2	10940.9	512.64 ms
3	11067.2	509.18 ms

Webflux 회차	TPS	Response Time
1	11416.2	500 ms
2	11334.3	512.04 ms
3	11291.4	513.28 ms

JDK8 회차	TPS	Response Time
1	11220.2	500.46 ms
2	11040.9	515.19 ms
3	11046.7	520 ms

1초 blocking

MVC-VT 회차	TPS	Response Time
1	2855.4	1018 ms
2	2858.5	1003 ms
3	2853.1	1012 ms

MVC-PT 회차	TPS	Response Time
1	197.0	15007 ms
2	197.2	15010 ms
3	197.2	15132 ms

Webflux 회차	TPS	Response Time
1	2863.2	1003 ms
2	2870.8	1002 ms
3	2853.4	1005 ms

JDK8 회차	TPS	Response Time
1	198.0	15046 ms
2	197.9	15213 ms
3	197.8	15123 ms

1초 blocking + 고속 select

MVC-VT 회차	TPS	Response Time
1	1987.5	1873 ms
2	2013.5	1701 ms
3	1948.1	1772 ms

MVC-PT 회차	TPS	Response Time
1	196.2	15153 ms
2	195.9	15013 ms
3	195.2	15211 ms

Webflux 회차	TPS	Response Time
1	2851.2	1014 ms
2	2847	1013 ms
3	2887.5	1005 ms

JDK8 회차	TPS	Response Time
1	196.8	15144 ms
2	197.6	15034 ms
3	198.68	15135 ms

Slow Query (응답 1초)

MVC-VT 회차	TPS	Response Time
1	4	262초
2	-	-
3	-	-

MVC-PT 회차	TPS	Response Time
1	148.2	20.06 s
2	149.1	20.05 s
3	149.0	20.04 s

Webflux 회차	TPS	Response Time
1	147.9	20.03 s
2	148.7	20.06 s
3	148.5	20.03 s

JDK8 회차	TPS	Response Time
1	148.2	20.76 s
2	148.8	21.03 s
3	148.4	20.25 s

Synchronized - Carrier Thread Blocking

Synchronized 사용 시 Virtual Thread에 연결된 Carrier Thread가 Blocking.

이런 경우를 pinned

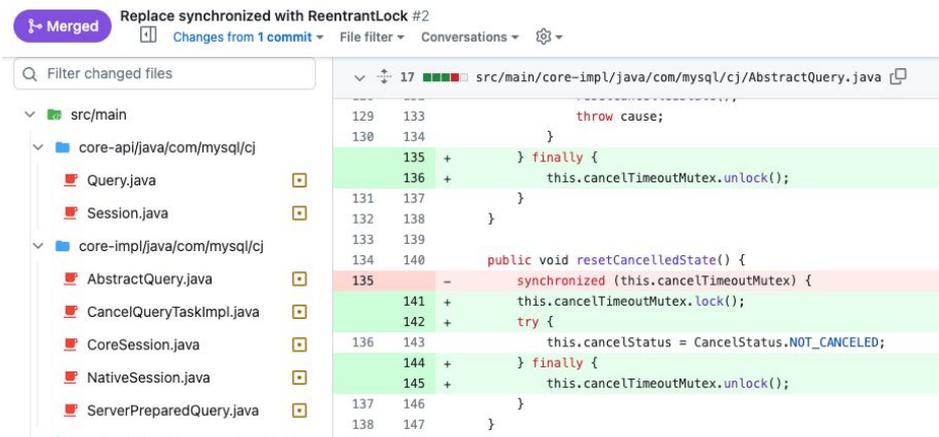
JEP 444: Virtual Threads

There are two scenarios in which a virtual thread cannot be unmounted during blocking operations because it is *pinned* to its carrier:

1. When it executes code inside a synchronized block or method, or
2. When it executes a native method or a foreign function.

Synchronized - Carrier Thread Blocking

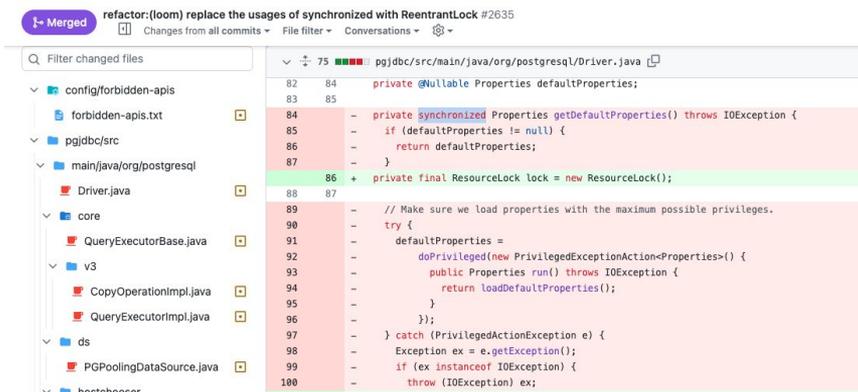
라이브러리들에서 **virtual thread** 대응을 위해 작업중인것으로 보임



Replace synchronized with ReentrantLock #2

```
src/main/core-impl/java/com/mysql/cj/AbstractQuery.java
129 133         throw cause;
130 134     }
135 +     } finally {
136 +         this.cancelTimeoutMutex.unlock();
137     }
138 139 }
139 140
140 141 public void resetCancelledState() {
141 142     synchronized (this.cancelTimeoutMutex) {
142 143         this.cancelTimeoutMutex.lock();
143 144     }
144 145     try {
145 146         this.cancelStatus = CancelStatus.NOT_CANCELED;
146 147     } finally {
147 148         this.cancelTimeoutMutex.unlock();
148 149     }
149 150 }
150 151 }
151 152 }
```

Mysql - mysql-connector-j



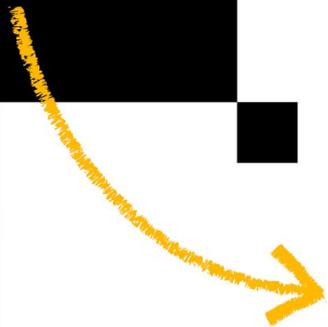
refactor:(loom) replace the usages of synchronized with ReentrantLock #2635

```
pgjdbc/src/main/java/org/postgresql/Driver.java
82 84     private @Nullable Properties defaultProperties;
83 85
84 -     private synchronized Properties getDefaultProperties() throws IOException {
85 -         if (defaultProperties != null) {
86 -             return defaultProperties;
87 -         }
88 +     private final ResourceLock lock = new ResourceLock();
89
90 // Make sure we load properties with the maximum possible privileges.
91 try {
92     defaultProperties =
93         doPrivileged(new PrivilegedExceptionAction<Properties>() {
94             public Properties run() throws IOException {
95                 return loadDefaultProperties();
96             }
97         });
98 } catch (PrivilegedActionException e) {
99     Exception ex = e.getException();
100     if (ex instanceof IOException) {
101         throw (IOException) ex;
102     }
103 }
```

Postgres - pgjdbc

```
ext(  
  'Section Title',  
  style: TextStyle(  
    color: Colors.yellow[200],  
  ),  
),  
),  
s.star,  
r: Colors.yellow[500],  
Text('23'),
```

devfest



Google Developer Groups

Incheon/Songdo

Conclusion

오해 하면 안되는 것

Continuation 구현

Structured Concurrency 전혀 다른 얘기

Spring MVC의 Thread Per Request 모델 동일

1개 요청을 1개의 task(virtual thread)로 처리하면서 발생하는 blocking을 기다리는건 동일

300ms API를 3번 호출하면? -> 당연히 900ms

API를 동시호출해서 처리하고 싶으면? 하던데로 CompletableFuture

기존의 Platform Thread 방식이 대체되는것 X

결론 - 개인적인 견해

DB 커넥션 관련 이슈만 해결되면 **Spring MVC + Virtual Thread** 사용

Virtual Thread와 **Webflux**의 처리량 차이로 인해 이슈가 생길정도라면?

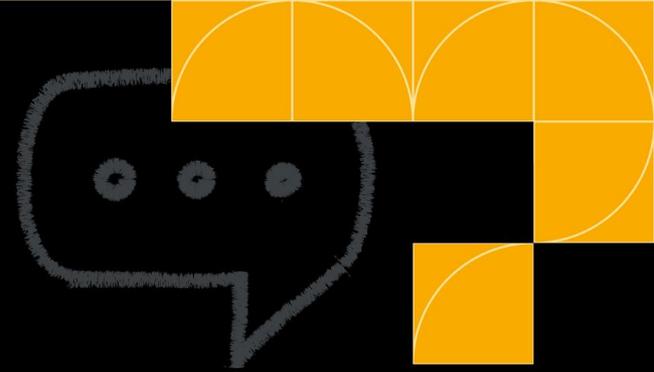
Webflux 부터 도입X -> 스케일아웃, 샤딩, 캐싱, **CQRS**, 업무 프로세스 재설계 등..

그래도 부족하면 **Rust? Elixir?**

```
Text(
  'Simple Statement or URL',
  style: TextStyle(
    color: Colors.yellow[200],
  ),
),
),
```

devfest

```
s.star,
r: Colors.yellow[500],
Text('23'),
```



Q & A

