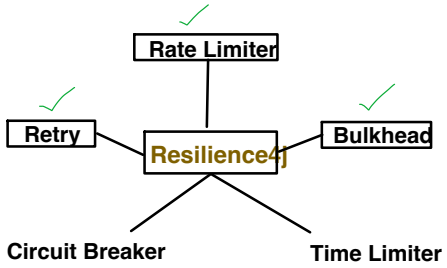
Circuit Breaker: Fault-Tolerant Microservice (Part-4)

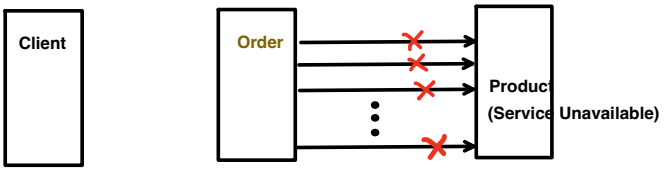
Friday, 1 August 2025 10:07 PM

To build Fault tolerant microservices: Resilience4j provides below mechanisms



Circuit Breaker:

- This pattern prevents an application to make repeated calls to a downstream service that is likely to fail.



Product Service is down, no matter how many times Order service will invoke Product service either by retrying the same call or different call, It will fail.

Disadvantages:

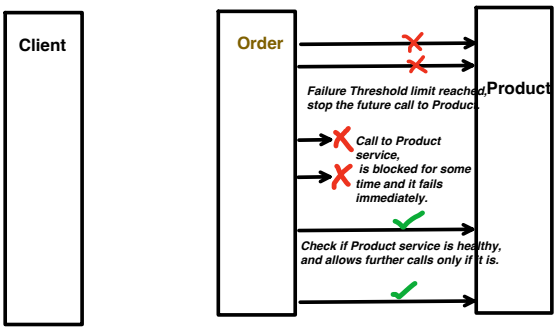
- It unnecessarily adds load to Product service and because of that Product service might take longer time to recover.
- Order service is unnecessarily wasting its resource (latency and thread blocking) by making call that is likely going to fail.

So, what's the solution?

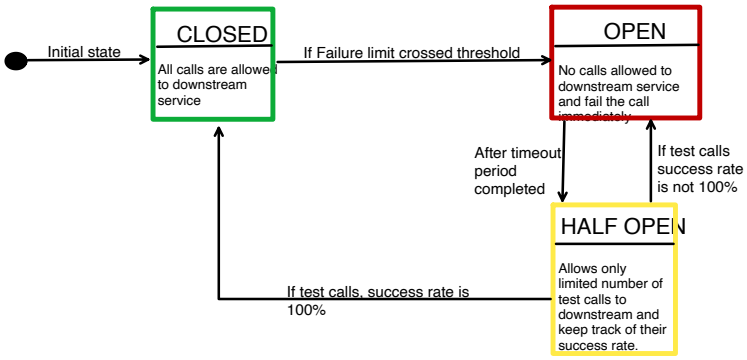
How does Order service know, when to stop the downstream Product service call and when to restart?

Answer is "Circuit Breaker". Again, lets revisit its definition.

It prevents an application to make repeated calls to a downstream service that is likely to fail.



States of Circuit Breaker:



Pom.xml dependency

```xml
<dependency>
    <groupId>io.github.resilience4j</groupId>
    <artifactId>resilience4j-spring-boot3</artifactId>
    <version>2.1.0</version>
</dependency>
```

```java
@RestController
@RequestMapping("/orders")
public class OrderController {

    @Autowired
    OrderService orderService;

    @GetMapping("/{id}")
    public void callProductAPI(@PathVariable String id) {
        orderService.invokeProductAPI(id);

    }
}
```

```java
@FeignClient(name = "product-service")
public interface ProductClient {

    @GetMapping(value = "/products/{id}")
    String getProductById(@PathVariable("id") String id);
}
```

```java
@Component
public class OrderService {

    @Autowired
    ProductClient productClient;

    @CircuitBreaker(name = "productService", fallbackMethod = "fallback")
    public void invokeProductAPI(String id) {
        productClient.getProductById(id);
    }

    public void fallback(Throwable ex) {
        System.out.println( "not able to invoke product service");
    }
}
```

Fallback method is invoked for each failure attempt.

application.properties

```properties
server.port=8081
spring.application.name=order-service
eureka.client.service-url.defaultZone=http://localhost:8761/eureka

#product service - circuit breaker configurations
resilience4j.circuitbreaker.instances.productService.sliding-window-type=COUNT_BASED
resilience4j.circuitbreaker.instances.productService.sliding-window-size=10
resilience4j.circuitbreaker.instances.productService.minimum-number-of-calls=5
resilience4j.circuitbreaker.instances.productService.failure-rate-threshold=50
resilience4j.circuitbreaker.instances.productService.wait-duration-in-open-state=10s
resilience4j.circuitbreaker.instances.productService.permitted-number-of-calls-in-half-open-state=3
resilience4j.circuitbreaker.instances.productService.automatic-transition-from-open-to-half-open-enabled=true
```

minimum-number-of
in a window, before i

failure-rate-threshold
once this threshold %

wait-duration-in-ope
state, before it move
(automatically, if au
open-enabled is set

permitted-number-o
calls allowed in HAL
downstream is healt

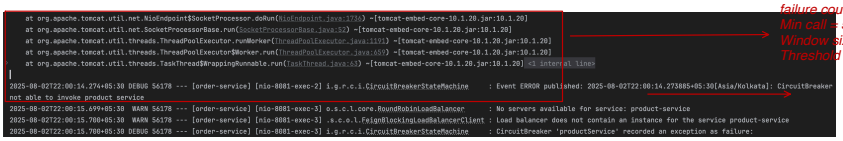
sliding-window-type=COUNT_BASED
sliding-window-size=10
Tracks N (in this case 10) number of recent calls.

sliding-window-type=TIME_BASED
sliding-window-size=10s
Tracks calls made in a last N time duration(in this case 10sec)

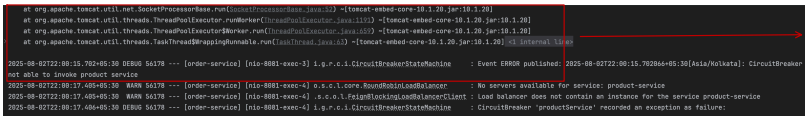
By default, the Circuit Breaker records all **RuntimeExceptions** and **Errors** as failures.
But if we want specific exception to be recorded and ignored, we can also configured it like below:

```properties
resilience4j.circuitbreaker.instances.productService.record-exceptions=java.io.IOException,org.springframework.web.client.Ht
resilience4j.circuitbreaker.instances.productService.ignore-exceptions=java.lang.IllegalArgumentException
```

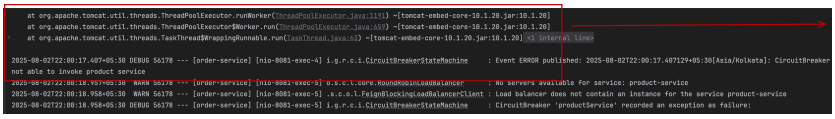
Output:



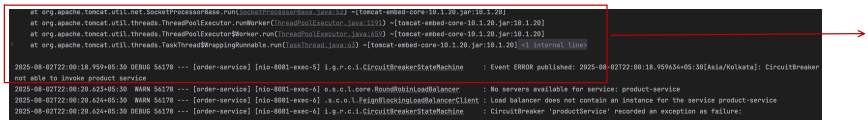
1st call to downstream : Failed
failure count: 1
Min call = 5
Window size = 10
Threshold = 50% of window size = 5 failure is the threshold



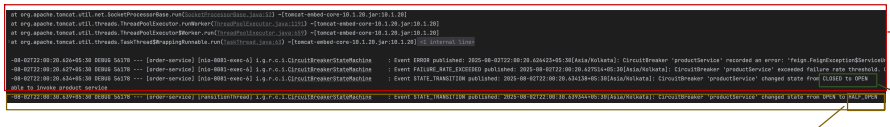
2nd call to downstream : Failed
failure count: 2
Min call = 5
Window size = 10
Threshold = 50% of window size = 5 failure is the threshold



3rd call to downstream : Failed
failure count: 3
Min call = 5
Window size = 10
Threshold = 50% of window size = 5 failure is the threshold



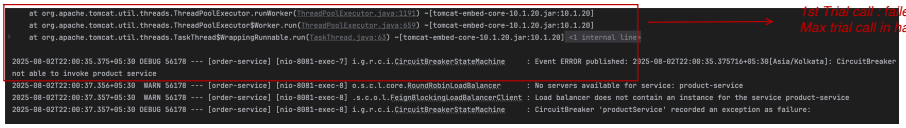
4th call to downstream : Failed
failure count: 4
Min call = 5
Window size = 10
Threshold = 50% of window size = 5 failure is the threshold



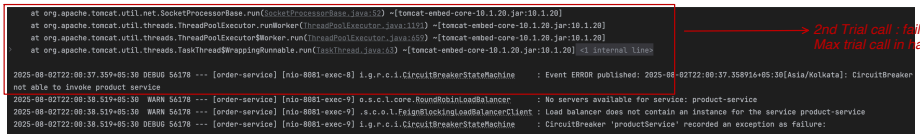
5th call to downstream : Failed
failure count: 5
Min call = 5
Window size = 10
Threshold = 50% of window size = 5 failure is the threshold

Now min call value is reached and also failure Threshold limit is touched, so state changed from CLOSED to OPEN

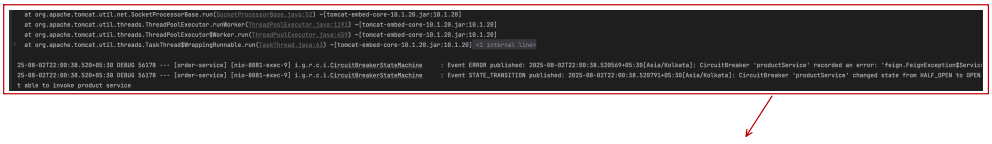
After 10sec of wait timeout,
state changed form OPEN to HALF_OPEN



1st Trial call : failed
Max trial call in half_open state = 3



2nd Trial call : failed
Max trial call in half_open state = 3



3rd Trial call : failed
Max trial call in half_open state = 3

Since max trial call limit reached, and success is not 100%, so status changed from HALF_OPEN to OPEN.

AOP intercept the call and pass it to "*CircuitBreakerStateMachine.java*" class.
Which has complete logic of changing one state to another, whenever there is a failure.

Sample method from framework class:

```
Transitions to open state when thresholds have been exceeded.
Params: result – the Result
private void checkIfThresholdsExceeded(Result result) {
    if (Result.hasExceededThresholds(result) && isClosed.compareAndSet( expectedValue: true,  newValue: false)) {
        publishCircuitThresholdsExceededEvent(result, circuitBreakerMetrics);
        transitionToOpenState();
    }
}
```

After every error, it checks if state need to be changed or not.
Like here its checking, if threshold limit reached and state is CLOSED, then transit to OPEN state.

Likewise similar method is present for different state with specific transition logic.

One question, might be coming to you:

- Okay, for every failure call, we are checking if state need to be changed or not, make sense.
- But once it moved to OPEN state, then how automatically say after 10sec it move to HALF_OPEN state?

It uses, ScheduledThreadPoolExecutor

OpenState method from *CircuitBreakerStateMachine.java* framework class:

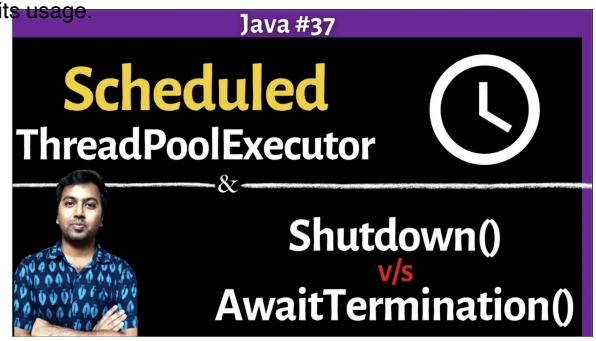
```
OpenState(final int attempts, final long waitDurationInMillis, final Instant retryAfterWaitDuration,
          CircuitBreakerMetrics circuitBreakerMetrics) {
    this.attempts = attempts;
    this.retryAfterWaitDuration = retryAfterWaitDuration;
    this.circuitBreakerMetrics = circuitBreakerMetrics;

    if (circuitBreakerConfig.isAutomaticTransitionFromOpenToHalfOpenEnabled()) {
        ScheduledExecutorService scheduledExecutorService = schedulerFactory.getScheduler();
        transitionToHalfOpenFuture = scheduledExecutorService
                .schedule(this::toHalfOpenState, waitDurationInMillis, TimeUnit.MILLISECONDS);
    } else {
        transitionToHalfOpenFuture = null;
    }
    isOpen = new AtomicBoolean( initialValue: true);
}
```

It passes the request to ScheduledThreadPoolExecutor.
1st parameter is the Task i.e. "toHalfOpenState" to transit the state from OPEN to HALF_OPEN.

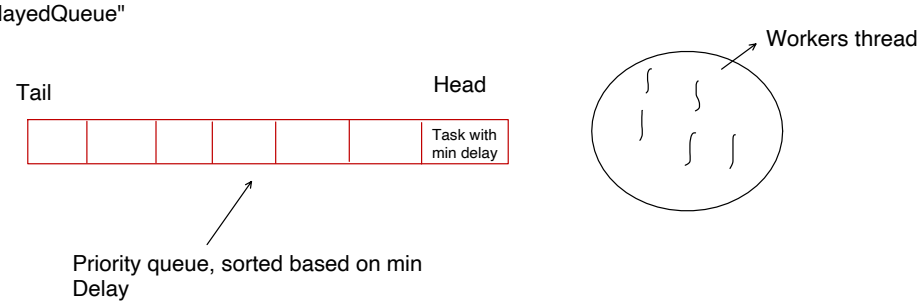
2nd parameter is the delay : like 10 or 20 or 30

3rd parameter is Time Unit: second or minutes or millisecond

I have already covered, how to use ScheduledThreadPoolExecutor in Java playlist. Pls check it out, if there is any doubt with its usage.



So internally **ScheduledThreadPoolExecutor** uses the concept of "DelayedQueue"



- Each available Worker thread, will look at the head of the Delayed Queue.
  - If task delay is not yet expired, then thread waits(blocks) for that specific remaining delay period.
  - Once the delay is over, OS wakes up the thread.
  - One thread, pick the task from the Head and start executing it.
  - Other thread start with the next task in the delayed queue, if task is not yet ready, then thread will wait(block) again.