

ORACLE®

**Lesson 3-5:
Parallel Streams
(And When Not To Use Them)**

Serial And Parallel Streams

- Collection stream sources
 - `stream()`
 - `parallelStream()`
- Stream can be made parallel or sequential at any point
 - `parallel()`
 - `sequential()`
- The last call wins
 - Whole stream is either sequential or parallel
- Calling `concat()` with a sequential and parallel stream will produce a parallel stream

Parallel Streams

- Implemented internally using the fork-join framework
- Will default to as many threads for the pool as the OS reports processors
 - Which may not be what you want

```
System.setProperty(  
"java.util.concurrent.ForkJoinPool.common.parallelism",  
"32767");
```

- Remember, parallel streams always need more work to process
 - But they might finish it more quickly

Parallel Stream Considerations

- `findFirst()` and `findAny()`
 - `findAny()` is non-deterministic, so better for parallel stream performance
 - Use `findFirst()` if a deterministic result is required
- `forEach()` and `forEachOrdered()`
 - `forEach()` is non-deterministic for a parallel stream and ordered data
 - Use `forEachOrdered()` if a deterministic result is required

When To Use Parallel Streams

No Simple Answer

- Data set size is important, as is the type of data structure
 - ArrayList: **GOOD**
 - HashSet, TreeSet: **OK**
 - LinkedList: **BAD**
- Operations are also important
 - Certain operations decompose to parallel tasks better than others
 - `filter()` and `map()` are excellent
 - `sorted()` and `distinct()` do not decompose well

When To Use Parallel Streams

Quantative Considerations

- N = size of the data set
- Q = Cost per element through the Stream pipeline
- $N \times Q$ = Total cost of pipeline operations
- The bigger $N \times Q$ is the better a parallel stream will perform
- It is easier to know N than Q , but Q can be estimated
- If in doubt, profile

Section 5

Summary

- Streams can be processed sequentially or in parallel
 - The whole stream is processed sequentially or in parallel
 - In most cases how the stream is defined will not affect the result
 - `findFirst()`, `findAny()`, `forEach()`, `forEachOrdered()` do
- Don't assume that a parallel stream will return a result faster
 - Many factors affect performance

ORACLE®