**Capítulo DEZESSETE**  
**Espiando, Mapeando, Reduzindo e Coletando**

**Objetivos do Exame**

Desenvolver código para extrair dados de um objeto usando os métodos peek() e map(), incluindo as versões primitivas do método map().

Salvar resultados em uma coleção usando o método collect e agrupar/particionar dados usando a classe Collectors.

Uso dos métodos merge() e flatMap() da API de Stream.

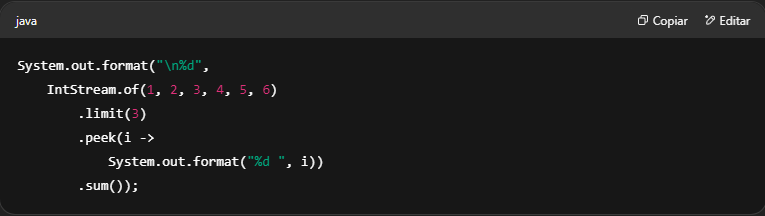
**peek()**

peek() é um método simples:

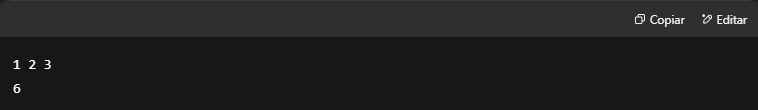


Ele apenas executa o Consumer fornecido e retorna um novo stream com os mesmos elementos do original.

Na maioria das vezes, esse método é usado com System.out.println() para fins de depuração (para ver o que há no stream):



A saída:

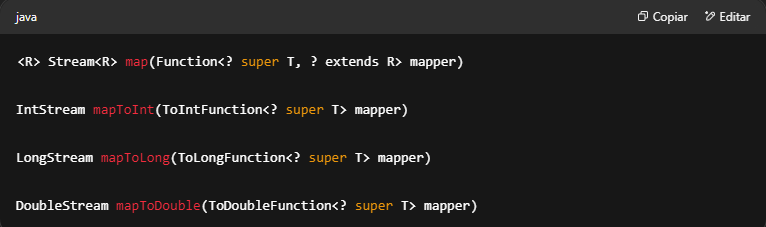


Observe que peek() é uma operação intermediária. No exemplo, não podemos usar algo como forEach() para imprimir os valores retornados por limit(), porque forEach() é uma operação terminal (e não poderíamos mais chamar sum()).

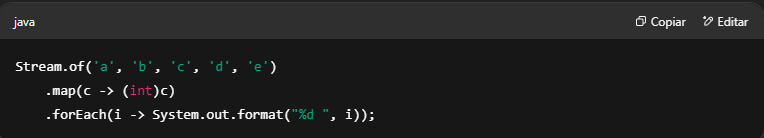
É importante enfatizar que peek() se destina a ver os elementos de um stream em um ponto específico do pipeline, sendo considerada uma má prática alterar o stream de qualquer forma. Se quiser fazer isso, use o método a seguir.

**map()**

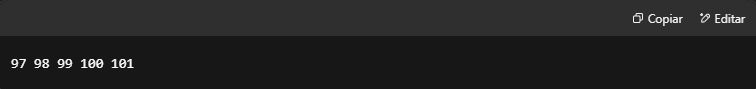
map() é usado para transformar o valor ou o tipo dos elementos de um stream:



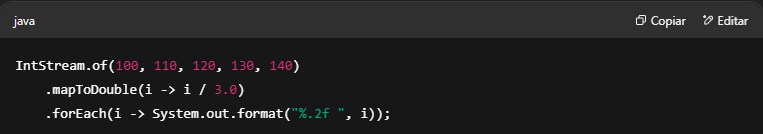
Como pode ver, map() recebe uma Function para converter os elementos de um stream do tipo T para o tipo R, retornando um stream desse tipo R:



A saída:



Há versões para transformar em tipos primitivos, por exemplo:

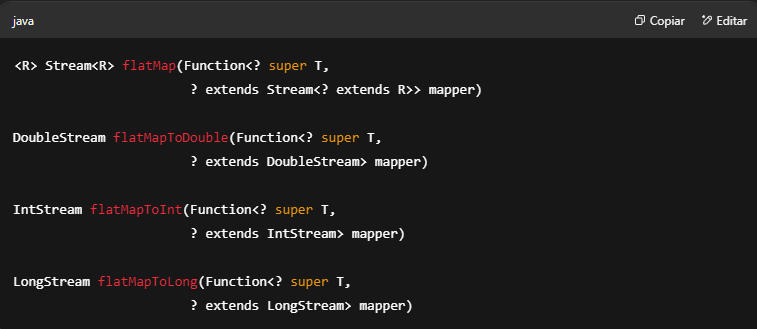


Irá produzir:



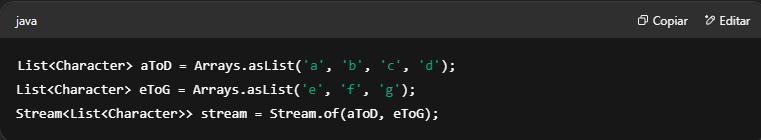
**flatMap()**

flatMap() é usado para “achatar” (ou combinar) os elementos de um stream em um (novo) stream:

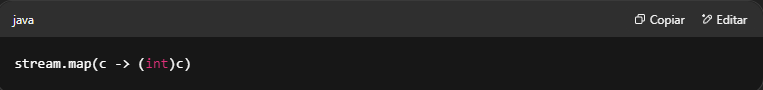


A partir de sua assinatura (e das assinaturas das versões primitivas), podemos ver que, em contraste com map() (que retorna um único valor), flatMap() deve retornar um Stream. Se flatMap() retornar null, um stream vazio será retornado em vez disso.

Vamos ver como isso funciona. Se temos um stream de listas de caracteres:



E queremos converter todos os caracteres para sua representação int, não podemos mais usar map():

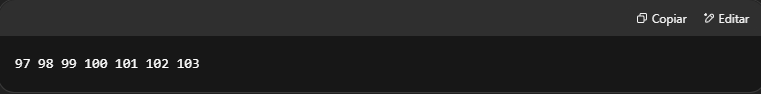


Porque (como cada elemento do stream é passado para map) c representa um objeto do tipo List<Character>, não Character.

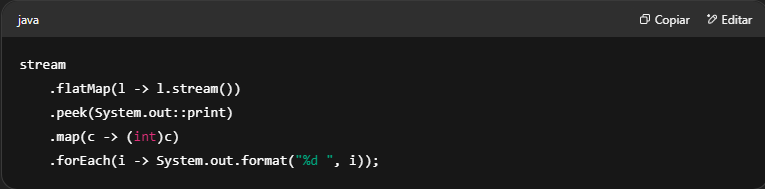
O que precisamos fazer é obter os elementos das listas em um único stream e então converter cada caractere para um int. Felizmente, a parte da “combinação” é exatamente o que flatMap() faz:



Então este código pode produzir:



Usar peek() após flatMap() pode esclarecer como os elementos são processados:



Como se pode ver na saída, o stream retornado por flatMap() é passado pelo pipeline, como se estivéssemos trabalhando com um stream de elementos únicos e não com um stream de listas de elementos:



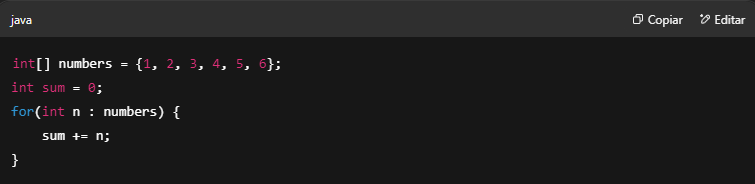
Dessa forma, com flatMap() você pode converter um Stream<List<Object>> em um Stream<Object>. No entanto, o conceito importante é que esse método retorna um stream e não um único elemento (como map() faz).

**Redução**

Uma **redução** é uma operação que pega muitos elementos e os combina (ou os reduz) em um único valor ou objeto, e isso é feito aplicando uma operação várias vezes.

Alguns exemplos de reduções são somar N elementos, encontrar o maior elemento entre N números ou contar elementos.

Como no exemplo a seguir, onde usando um laço for, reduzimos um array de números à sua soma:



Claro que fazer reduções com streams ao invés de laços tem mais benefícios, como facilitar a paralelização e melhorar a legibilidade.

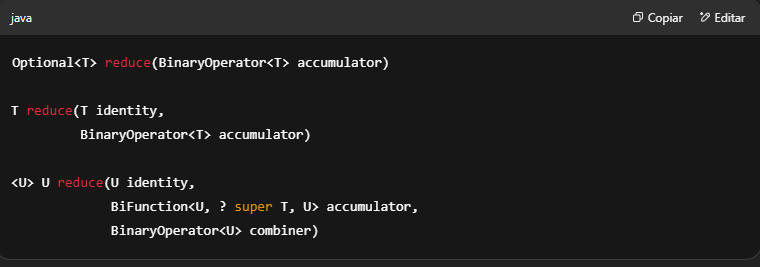
A interface Stream possui dois métodos para redução:

* reduce()
* collect()

Podemos implementar reduções com ambos os métodos, mas collect() nos ajuda a implementar um tipo de redução chamado **redução mutável**, onde um recipiente (como uma Collection) é usado para acumular o resultado da operação.

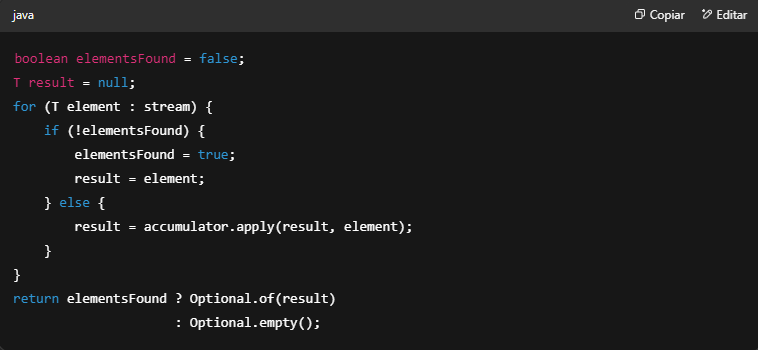
**reduce()**

Este método possui três versões:



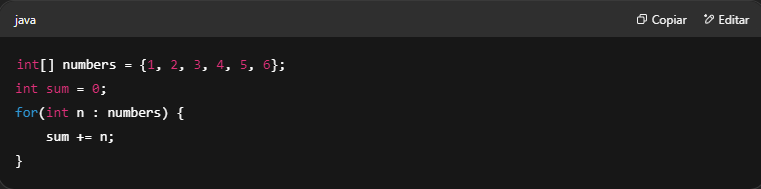
Lembre-se de que um BinaryOperator<T> é equivalente a um BiFunction<T, T, T>, onde os dois argumentos e o tipo de retorno são todos do mesmo tipo.

Comecemos com a versão que recebe um argumento. Ela é equivalente a:

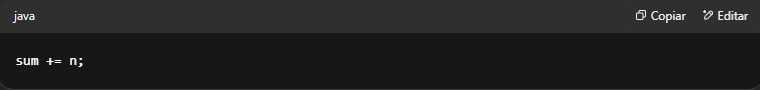


Este código simplesmente aplica uma função a cada elemento, acumulando o resultado e retornando um Optional encapsulando esse resultado, ou um Optional vazio se não houver elementos.

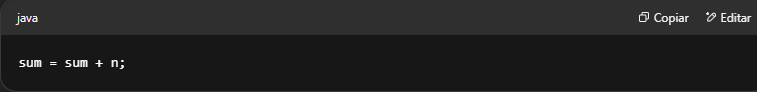
Vejamos um exemplo concreto. Acabamos de ver como uma soma é uma operação de redução:



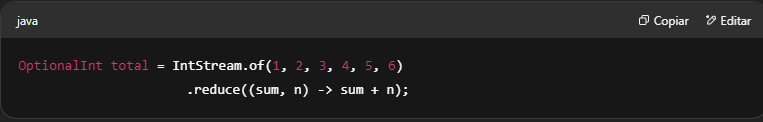
Aqui, a operação acumuladora é:



Ou:



O que se traduz para:

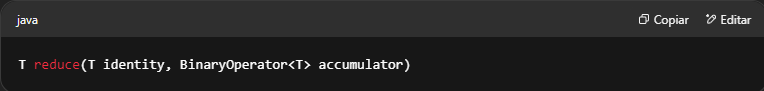


(Observe como a versão primitiva de stream usa a versão primitiva de Optional.)

Isso é o que acontece passo a passo:

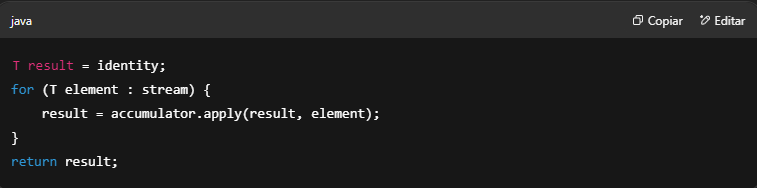
1. Uma variável interna que acumula o resultado é inicializada com o primeiro elemento do stream (1).
2. Esse acumulador e o segundo elemento do stream (2) são passados como argumentos para o BinaryOperator representado pela expressão lambda (sum, n) -> sum + n.
3. O resultado (3) é atribuído ao acumulador.
4. O acumulador (3) e o terceiro elemento do stream (3) são passados como argumentos para o BinaryOperator.
5. O resultado (6) é atribuído ao acumulador.
6. Os passos 4 e 5 são repetidos para os próximos elementos do stream até que não haja mais elementos.

No entanto, e se você precisar ter um valor inicial? Para casos assim, temos a versão que recebe dois argumentos:



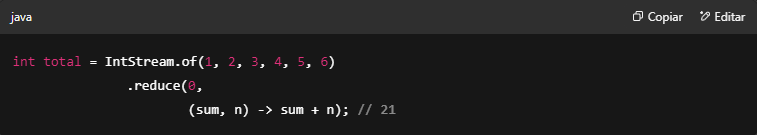
O primeiro argumento é o valor inicial, e ele é chamado de **identidade** porque, estritamente falando, esse valor deve ser uma identidade para a função acumuladora, ou seja, para cada valor v, accumulator.apply(identity, v) deve ser igual a v.

Esta versão de reduce() é equivalente a:

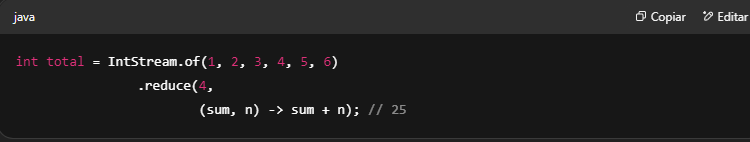


Observe que esta versão **não** retorna um objeto Optional, porque se o stream estiver vazio, o valor de identidade será retornado.

Por exemplo, o exemplo da soma pode ser reescrito como:



Ou usando um valor inicial diferente:

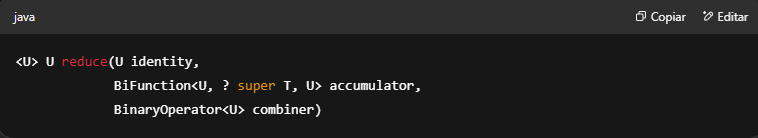


Contudo, observe que no último exemplo, o primeiro valor **não** pode ser considerado uma identidade (como no primeiro exemplo), já que, por exemplo, 4 + 1 não é igual a 4.

Isso pode causar alguns problemas ao trabalhar com **streams paralelos**, os quais revisaremos no próximo capítulo.

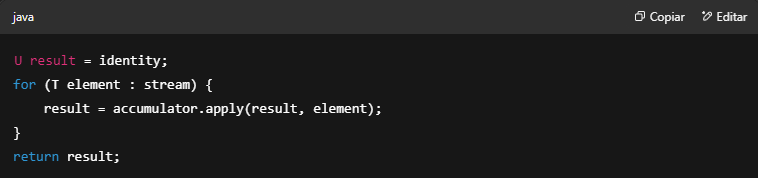
Agora, observe que com essas versões, você pega elementos do tipo T e retorna um valor reduzido **também do tipo T**.

Contudo, se quiser retornar um valor reduzido de um **tipo diferente**, você deve usar a versão de três argumentos de reduce():



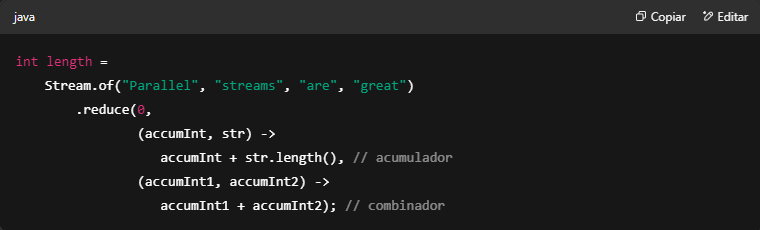
(Observe o uso dos tipos T e U.)

Esta versão é equivalente a:

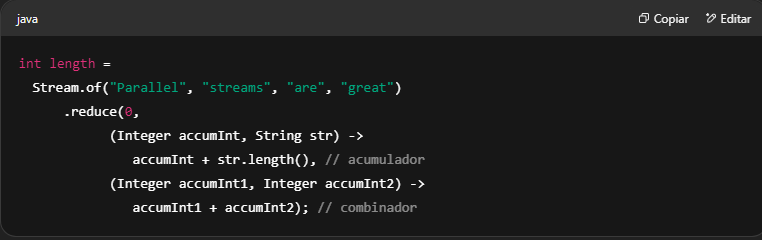


Considere, por exemplo, que queremos obter a **soma dos comprimentos** de todas as strings de um stream. Assim, estamos recebendo strings (tipo T) e queremos um resultado inteiro (tipo U).

Nesse caso, usamos reduce() assim:



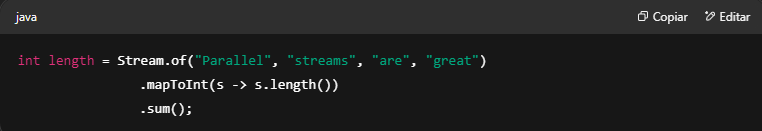
Podemos tornar mais claro adicionando os tipos dos argumentos:



Como a função acumuladora adiciona uma etapa de mapeamento (transformação) à função de acumulação, essa versão do reduce() pode ser escrita como uma combinação de map() com as outras versões do método reduce() (você talvez conheça isso como o padrão *map-reduce*):



Ou simplesmente:



Porque, de fato, as operações de cálculo que aprendemos no capítulo anterior são implementadas como operações de redução nos bastidores:

* average
* count
* max
* min
* sum

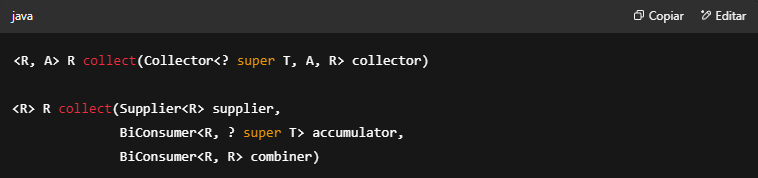
Além disso, observe que se o valor retornado for do **mesmo tipo**, a função combinadora não é mais necessária (afinal, ela acaba sendo igual à função acumuladora). Assim, nesse caso, é melhor usar a versão de dois argumentos.

Recomenda-se usar o método reduce() com três argumentos quando:

* Trabalhar com **streams paralelos** (mais sobre isso no próximo capítulo).
* Ter uma única função como mapeadora e acumuladora for mais eficiente do que ter funções separadas para mapeamento e redução.

**collect()**

Este método possui duas versões:



A primeira versão usa coletores predefinidos da classe Collectors, enquanto a segunda permite que você crie seus próprios coletores. Streams primitivos (como IntStream) possuem apenas essa última versão de collect().

Lembre-se de que collect() realiza uma **redução mutável** nos elementos de um stream, o que significa que ele usa um objeto mutável para acumular, como uma Collection ou um StringBuilder. Em contraste, reduce() combina dois elementos para produzir um novo e representa uma **redução imutável**.

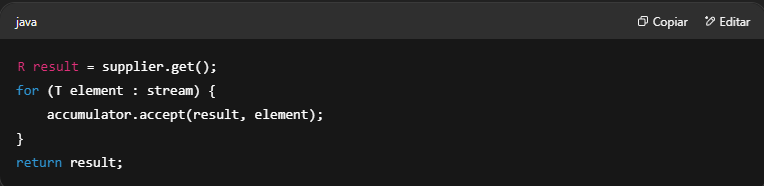
No entanto, comecemos com a versão que recebe três argumentos, pois ela é semelhante à versão de reduce() que também recebe três argumentos.

Como pode ser visto pela sua assinatura, primeiro, ela recebe um Supplier que retorna o objeto que será usado como recipiente (acumulador) e retornado ao final.

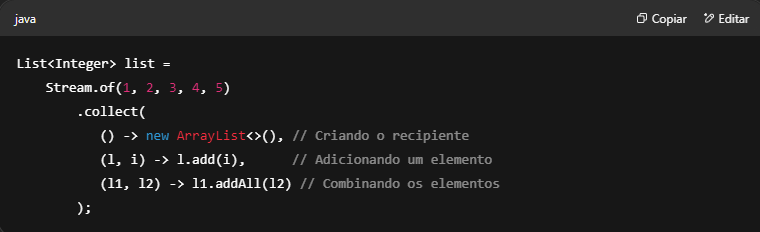
O segundo parâmetro é uma função acumuladora, que recebe o recipiente e o elemento a ser adicionado a ele.

O terceiro parâmetro é a função de combinação (*combiner*), que mescla os resultados intermediários no resultado final (útil ao trabalhar com streams paralelos).

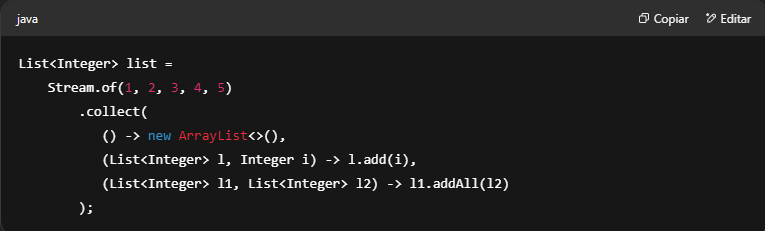
Esta versão de collect() é equivalente a:



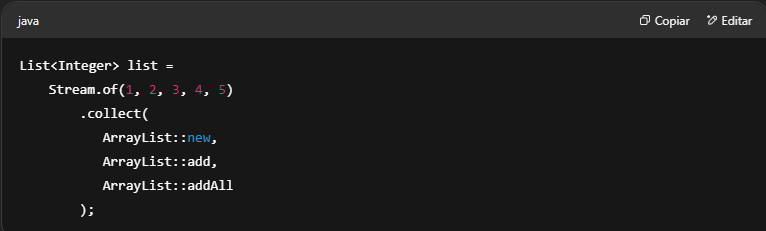
Por exemplo, se quisermos “reduzir” ou “coletar” todos os elementos de um stream em uma List, podemos fazer isso da seguinte forma:



Podemos tornar mais claro ao adicionar os tipos dos argumentos:



Ou também podemos usar referências de método:



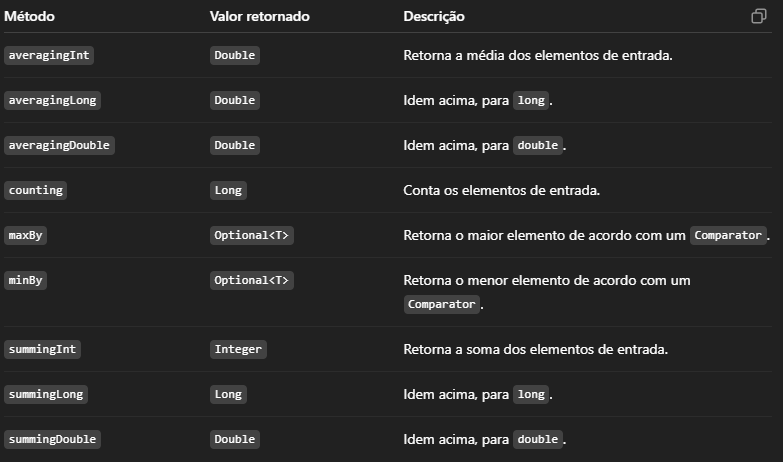
**Collectors**

A versão anterior de collect() é útil para aprender como os coletores funcionam, mas, na prática, é melhor usar a outra versão.

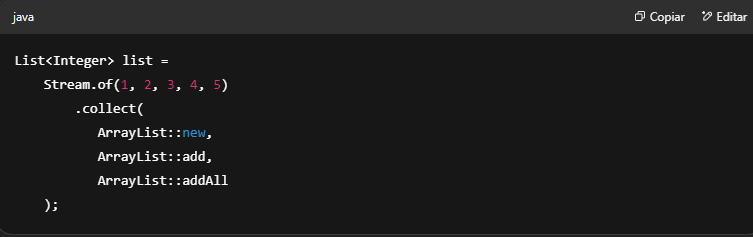
Alguns coletores comuns da classe Collectors são:

| **Método de collect()** | **Valor retornado** | **Descrição** |
| --- | --- | --- |
| toList | List | Acumula os elementos em uma List. |
| toSet | Set | Acumula os elementos em um Set. |
| toCollection | Collection | Acumula os elementos em uma implementação de Collection. |
| toMap | Map | Acumula os elementos em um Map. |
| joining | String | Concatena os elementos em uma String. |

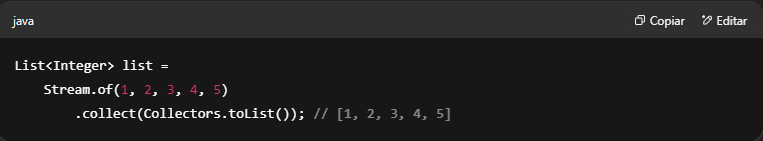
Como métodos de cálculo podem ser implementados como reduções, a classe Collectors também os fornece como coletores:



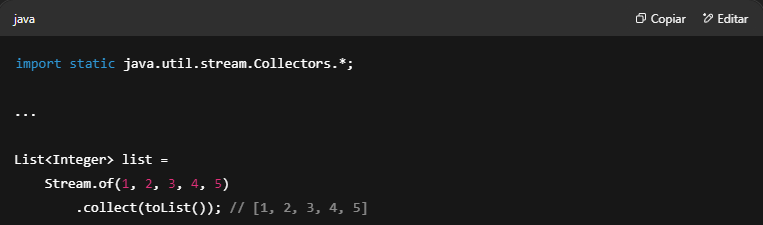
Dessa forma, podemos reescrever o exemplo anterior:



Como:



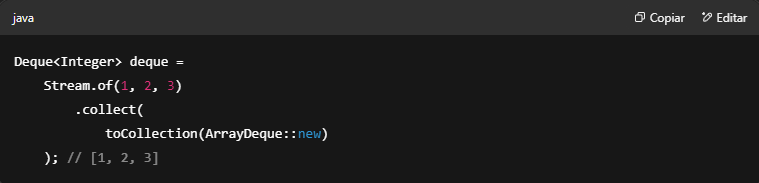
Como todos esses métodos são estáticos, podemos usar *imports estáticos*:



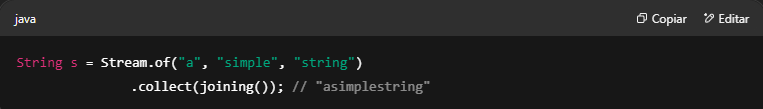
Se quisermos coletar os elementos em um Set:



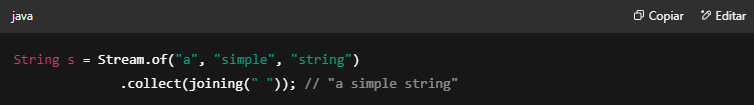
Se quisermos criar outra implementação de Collection:



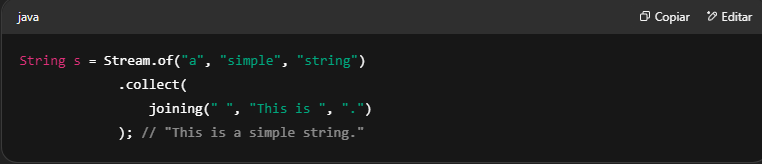
Se estivermos trabalhando com streams de String, podemos unir todos os elementos em uma única String com:



Também podemos passar um separador:



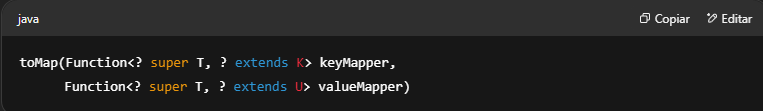
E um prefixo e um sufixo:



**toMap()**

No caso de mapas, as coisas ficam um pouco mais complicadas, porque, dependendo das nossas necessidades, temos **três opções**.

Na primeira opção, toMap() recebe dois argumentos (não estou mostrando os tipos de retorno porque são difíceis de ler e não trazem muito valor aqui):

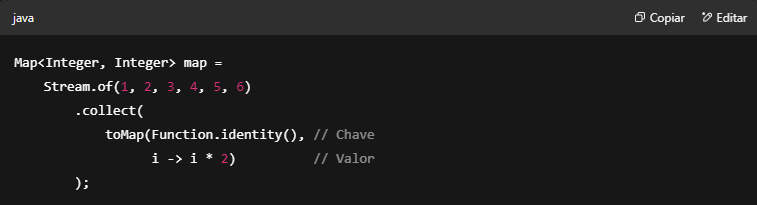


Ambas as Functions recebem um elemento do stream como argumento e retornam a chave ou o valor de uma entrada do Map, por exemplo:



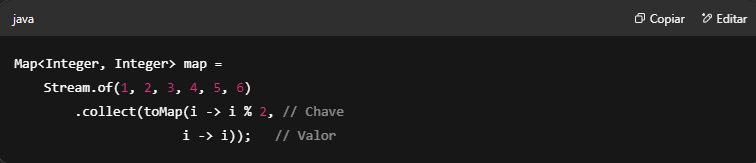
Aqui, estamos usando o elemento (como 1) como chave, e o elemento multiplicado por dois como valor (como 2).

Também podemos escrever i -> i como Function.identity():

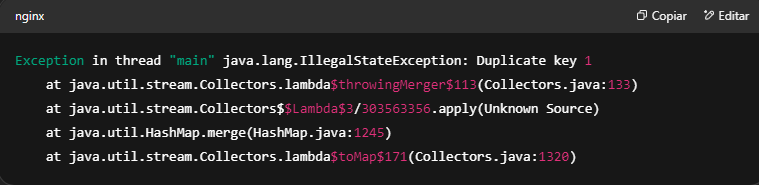


java.util.function.Function.identity() retorna uma função que sempre retorna seu argumento de entrada, ou seja, é equivalente a t -> t.

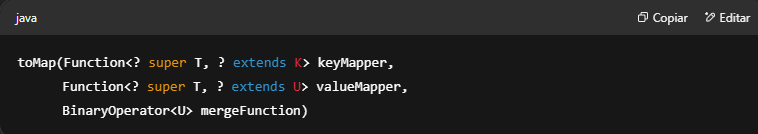
Mas o que acontece se mais de um elemento for mapeado para a mesma chave, como em:



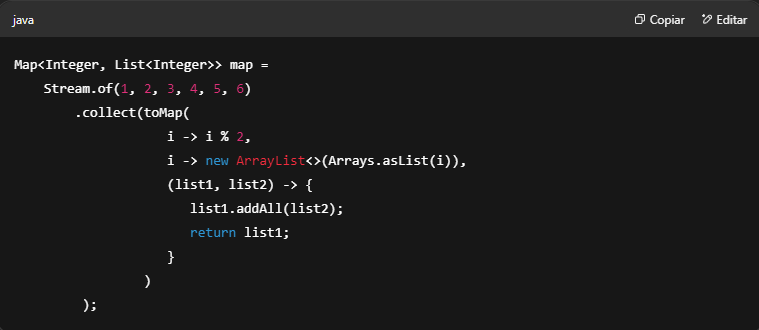
O Java não saberá o que fazer, então uma exceção será lançada:



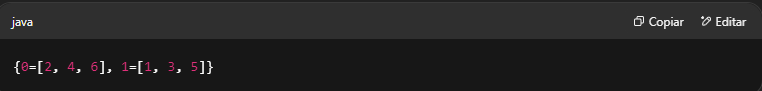
Para esses casos, usamos a versão que recebe três argumentos:



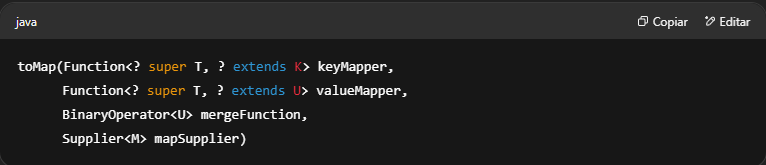
O terceiro argumento é uma função que define o que fazer quando há uma chave duplicada. Por exemplo, podemos criar uma List para adicionar os valores:



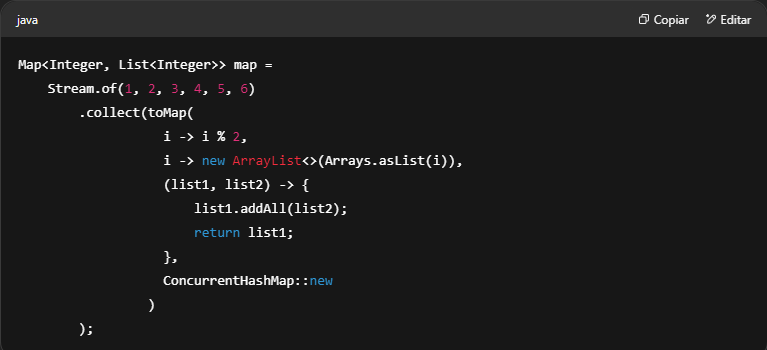
Isso retornará o seguinte mapa:



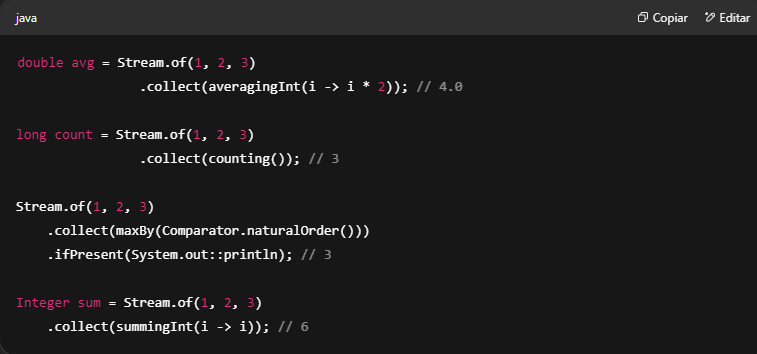
A terceira versão de toMap() recebe todos esses argumentos **mais um** que retorna um novo Map vazio no qual os resultados serão inseridos:



Assim, podemos trocar a implementação padrão (HashMap) por ConcurrentHashMap, por exemplo:



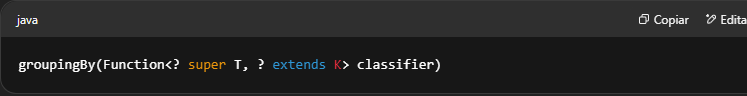
Sobre os **métodos de cálculo**, eles são fáceis de usar. Exceto por counting(), eles recebem uma Function para produzir um valor ao qual aplicar a operação, ou (no caso de maxBy e minBy) recebem um Comparator para produzir o resultado:



**groupingBy()**

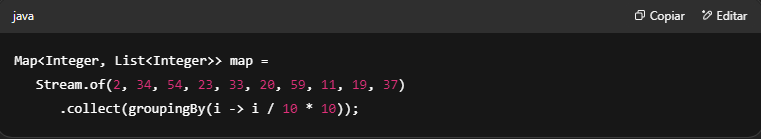
A classe Collectors fornece duas funções para agrupar os elementos de um stream em uma lista, em um estilo semelhante ao GROUP BY do SQL.

O primeiro método é groupingBy() e possui três versões. Esta é a primeira:



Ela recebe uma Function que classifica elementos do tipo T, agrupa-os em uma lista e retorna o resultado em um Map onde as chaves (do tipo K) são os valores retornados pela função.

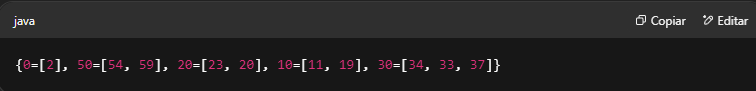
Por exemplo, se quisermos agrupar um stream de números pela faixa a que pertencem (dezenas, vintenas, etc.), podemos fazer algo assim:



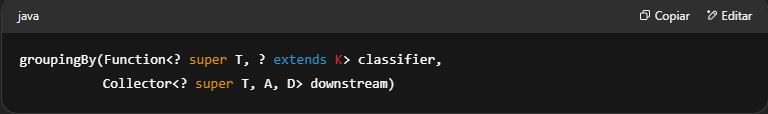
No momento em que você compara este código com a maneira tradicional de fazer isso (com um laço for), é que percebe o poder dos streams:



De qualquer maneira, ambos retornarão o seguinte mapa:

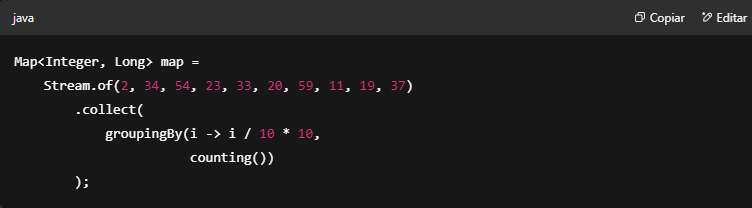


A segunda versão recebe um **coletor descendente** como argumento adicional:



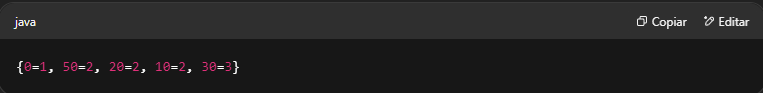
Um **coletor descendente** é um coletor que é aplicado aos resultados de outro coletor.

Podemos usar qualquer coletor aqui, por exemplo, para contar os elementos em cada grupo do exemplo anterior:

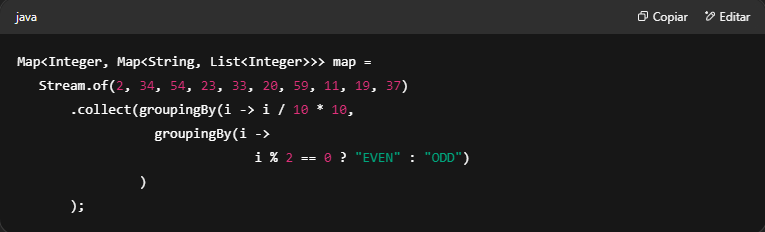


(Observe como o tipo dos valores do Map muda para refletir o tipo retornado pelo coletor descendente, counting().)

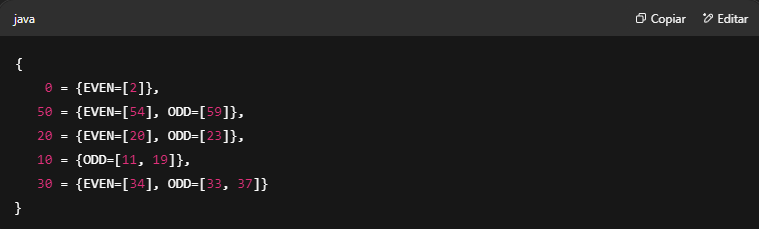
Isso retornará o seguinte mapa:



Podemos até usar outro groupingBy() para classificar os elementos em um segundo nível. Por exemplo, em vez de contar, podemos classificar ainda mais os elementos em pares ou ímpares:



Isso retornará o seguinte mapa (com um pouco de formatação):



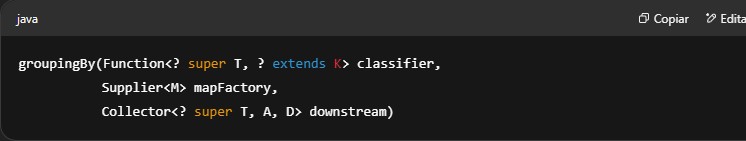
A chave do mapa de nível superior é um Integer, pois o primeiro groupingBy() retorna um Integer.

O tipo dos valores do mapa de nível superior mudou (novamente) para refletir o tipo retornado pelo coletor descendente, groupingBy().

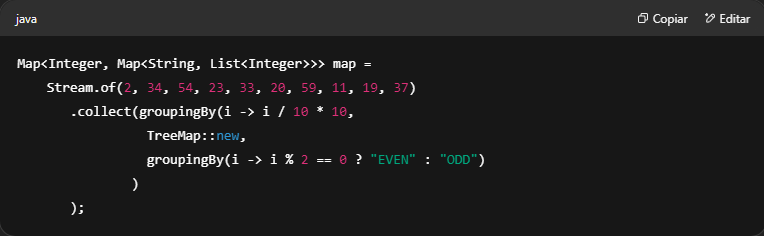
Neste caso, uma String é retornada, portanto, esse será o tipo das chaves do mapa de segundo nível, e como estamos trabalhando com um stream de Integer, os valores são do tipo List<Integer>.

Vendo a saída desses exemplos, você pode estar se perguntando: **há uma maneira de obter o resultado ordenado?**

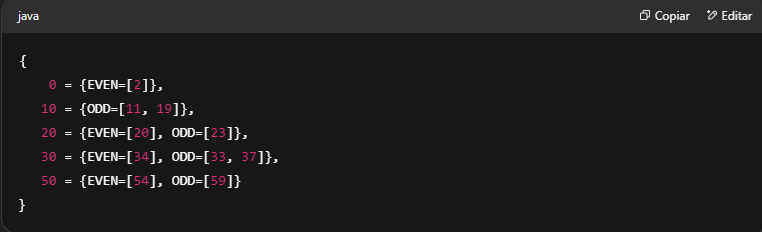
Bem, TreeMap é a única implementação de Map que é ordenada. Felizmente, a terceira versão de groupingBy() adiciona um argumento Supplier que nos permite escolher o tipo do Map resultante:



Dessa forma, se passarmos uma instância de TreeMap:



Isso retornará o seguinte mapa:

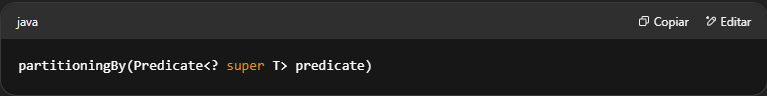


**partitioningBy()**

O segundo método para agrupamento é partitioningBy().

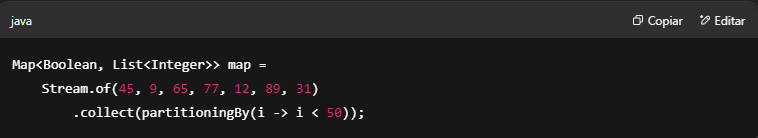
A diferença em relação a groupingBy() é que partitioningBy() retornará um Map com Boolean como tipo de chave, o que significa que há **apenas dois grupos**, um para true e outro para false.

Há duas versões deste método. A primeira é:



Ela particiona os elementos de acordo com um Predicate e os organiza em um Map<Boolean, List<T>>.

Por exemplo, se quisermos particionar um stream de números entre os que são menores que 50 e os que não são, podemos fazer assim:



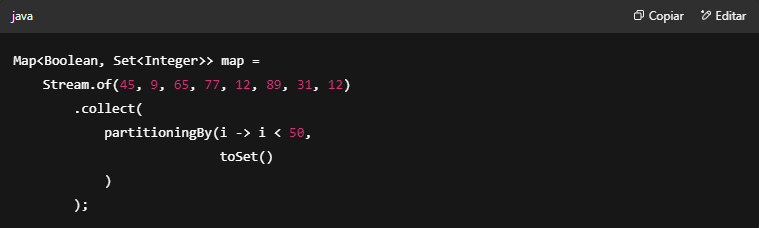
Isso retornará o seguinte mapa:



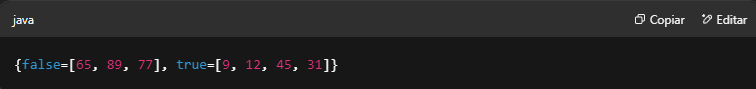
Como você pode ver, por causa do Predicate, o mapa sempre terá **dois elementos**.

E, assim como em groupingBy(), este método possui uma segunda versão que recebe um **coletor descendente**.

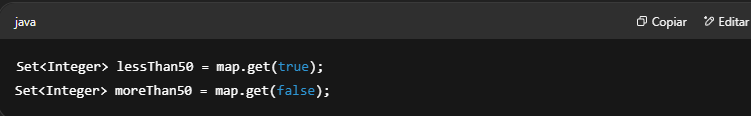
Por exemplo, se quisermos remover duplicatas, basta coletar os elementos em um Set desta forma:



Isso produzirá o seguinte Map:



No entanto, diferente de groupingBy(), não há uma versão que nos permita mudar o tipo do Map retornado. Mas isso não importa — você só tem duas chaves que pode obter com:

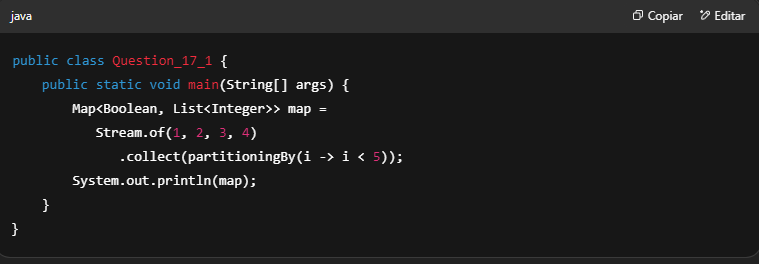


**Pontos-Chave (Key Points)**

* peek() executa o Consumer fornecido e retorna um novo stream com os **mesmos elementos** do stream original. Na maioria das vezes, esse método é usado para fins de **depuração**.
* map() é usado para transformar o **valor ou o tipo** dos elementos de um stream por meio de uma Function fornecida.
* flatMap() é usado para “**achatar**” (ou combinar) os elementos de um stream em **um único** (novo) stream. Em contraste com map() (que retorna um único valor), flatMap() **deve retornar um Stream**.
* Uma **redução** é uma operação que pega muitos elementos e os **combina** (ou reduz) em um único valor ou objeto.
* reduce() realiza uma redução nos elementos de um stream usando uma função acumuladora, um valor de identidade opcional e também uma função combinadora opcional.
* collect() implementa um tipo de redução chamado **redução mutável**, em que um recipiente (como uma Collection) é usado para acumular o resultado da operação.
* A classe Collectors fornece métodos estáticos como toList() e toMap() para criar uma coleção ou um mapa a partir de um stream, além de métodos de cálculo como averagingInt().
* Collectors.groupingBy() agrupa os elementos de um stream usando uma Function fornecida como classificador. Ele também pode receber um **coletor descendente** para criar outro nível de classificação.
* Você também pode **agrupar** (ou particionar) os elementos em um stream com base em uma condição (Predicate) usando o método Collectors.partitioningBy().

**Autoavaliação (Self Test)**

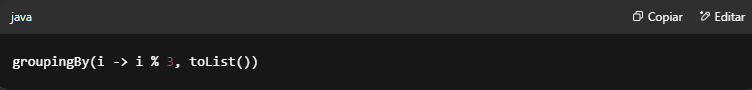
1. Dado:



Qual é o resultado?

**A.** {true=[1, 2, 3, 4]}  
**B.** {false=[], true=[1, 2, 3, 4]}  
**C.** {false=[1, 2, 3, 4]}  
**D.** {false=[1, 2, 3, 4], true=[]}

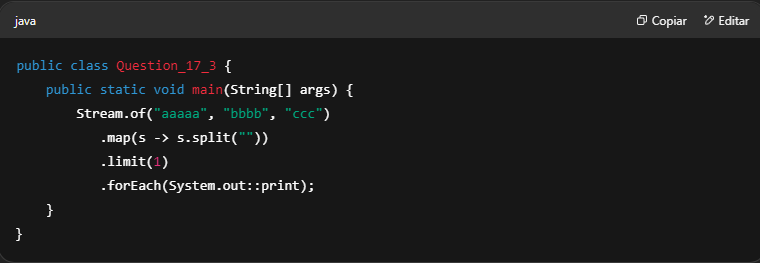
1. Dado:



Qual das seguintes opções é equivalente?

**A.** partitioningBy(i -> i % 3 == 0, toList())  
**B.** partitioningBy(i -> i % 3, toList())  
**C.** groupingBy(i -> i % 3 == 0)  
**D.** groupingBy(i -> i % 3)

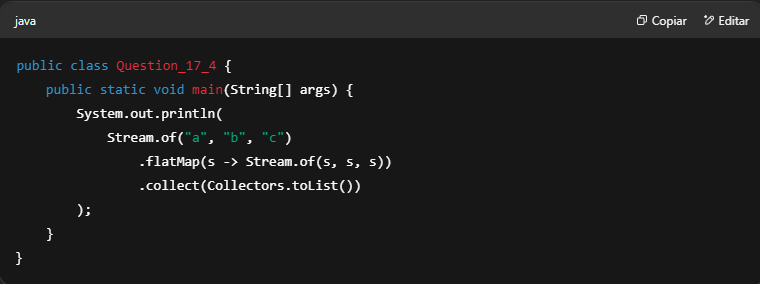
1. Dado:



Qual é o resultado?

**A.** aaaaa  
**B.** abc  
**C.** a  
**D.** Nenhuma das anteriores

1. Dado:



Qual é o resultado?

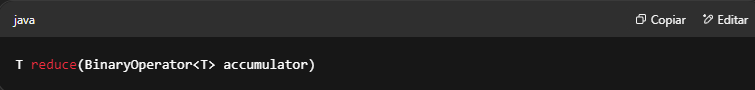
**A.** [a, a, a, b, b, b, c, c, c]  
**B.** [a, a, a]  
**C.** [a, b, c]  
**D.** Falha de compilação

1. Qual das opções a seguir é a maneira correta de implementar OptionalInt min() com uma operação de redução?

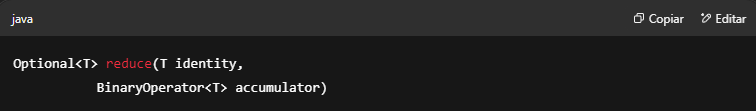
**A.** reduce((a, b) -> a > b)  
**B.** reduce(Math::min)  
**C.** reduce(Integer.MIN\_VALUE, Math:min)  
**D.** collect(Collectors.minBy())

1. Qual das seguintes é uma sobrecarga correta do método reduce()?

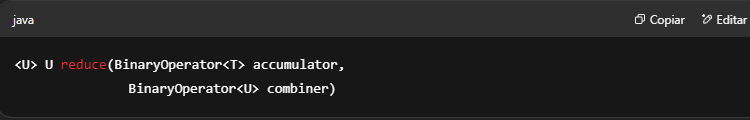
**A.**



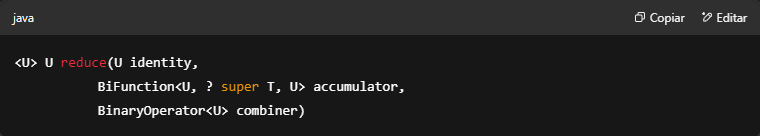
B.



C.



D.

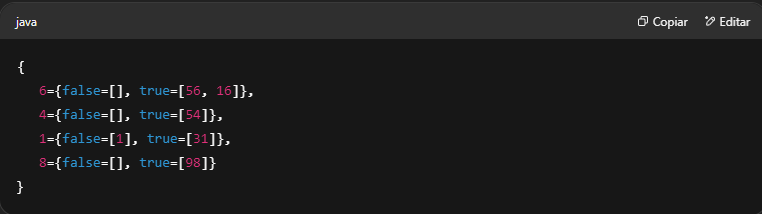


1. Dado:

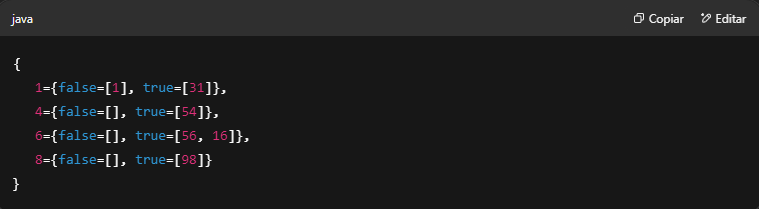


Qual é o resultado?

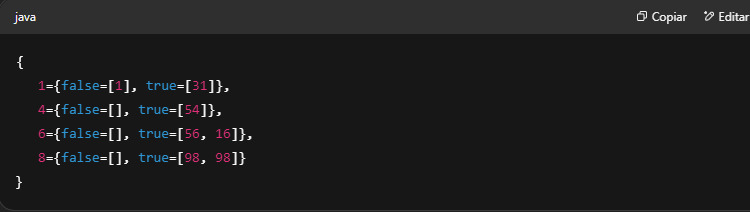
**A.**



B.



C.



D.

