

第8讲 | 对比Vector、ArrayList、LinkedList有何区别？

2018-05-22 杨晓峰



第8讲 | 对比Vector、ArrayList、LinkedList有何区别？

杨晓峰



- 00:00 / 12:45

我们在日常的工作中，能够高效地管理和操作数据是非常重要的。由于每个编程语言支持的数据结构不尽相同，比如我最早学习的C语言，需要自己实现很多基础数据结构，管理和操作会比较麻烦。相比之下，Java则要方便的多，针对通用场景的需求，Java提供了强大的集合框架，大大提高了开发者的生产力。

今天我要问你的的是有关集合框架方面的问题，对比Vector、ArrayList、LinkedList有何区别？

典型回答

这三者都是实现集合框架中的List，也就是所谓的有序集合，因此具体功能也比较近似，比如都提供按照位置进行定位、添加或者删除的操作，都提供迭代器以遍历其内容等。但因为具体的设计区别，在行为、性能、线程安全等方面，表现又有很大不同。

Vector是Java早期提供的线程安全的动态数组，如果不需要线程安全，并不建议选择，毕竟同步是有额外开销的。Vector内部是使用对象数组来保存数据，可以根据需要自动的增加容量，当数组已满时，会创建新的数组，并拷贝原有数组数据。

ArrayList是应用更加广泛的动态数组实现，它本身不是线程安全的，所以性能要好很多。与Vector近似，ArrayList也是可以根据需要调整容量，不过两者的调整逻辑有所区别，Vector在扩容时会提高1倍，而ArrayList则是增加50%。

LinkedList顾名思义是Java提供的双向链表，所以它不需要像上面两种那样调整容量，它也不是线程安全的。

考点分析

似乎从我接触Java开始，这个问题就一直是经典的面试题，前面我的回答覆盖了三者的一些基本的设计和实现。

一般来说，也可以补充一下不同容器类型适合的场景：

- Vector和ArrayList作为动态数组，其内部元素以数组形式顺序存储的，所以非常适合随机访问的场合。除了尾部插入和删除元素，往往性能会相对较差，比如我们在中间位置插入一个元素，需要移动后续所有元素。
- 而LinkedList进行节点插入、删除却要高效得多，但是随机访问性能则要比动态数组慢。

所以，在应用开发中，如果事先可以估计到，应用操作是偏向于插入、删除，还是随机访问较多，就可以针对性的进行选择。这也是面试最常见的一个考察角度，给定一个场景，选择适合的数据结构，所以对于这种典型选择一定要掌握清楚。

考察Java集合框架，我觉得有很多方面需要掌握：

- Java集合框架的设计结构，至少要有个整体印象。
- Java提供的主要容器（集合和Map）类型，了解或掌握对应的数据结构、算法，思考具体技术选择。
- 将问题扩展到性能、并发等领域。
- 集合框架的演进与发展。

作为Java专栏，我会在尽量围绕Java相关进行扩展，否则光是罗列集合部分涉及的数据结构就要占用很大篇幅。这并不代表那些不重要，数据结构和算法是基本功，往往也是必考的点，有些公司甚至以考察这些方面而非常知名（甚至是“臭名昭著”）。我这里以需要掌握典型排序算法为例，你至少需要熟知：

- 内部排序，至少掌握基础算法如归并排序、交换排序（冒泡、快排）、选择排序、插入排序等。
- 外部排序，掌握利用内存和外部存储处理超大数据集，至少要理解过程和思路。

考察算法不仅仅是如何简单实现，面试官往往会刨根问底，比如哪些是排序是不稳定的呢（快排、堆排），或者思考稳定意味着什么；对不同数据集，各种排序的最好或最差情况；从某个角度如何进一步优化（比如空间占用，假设业务场景需要最小辅助空间，这个角度堆排序就比归并优异）等，从简单的了解，到进一步的思考，面试官通常还会观察面试者处

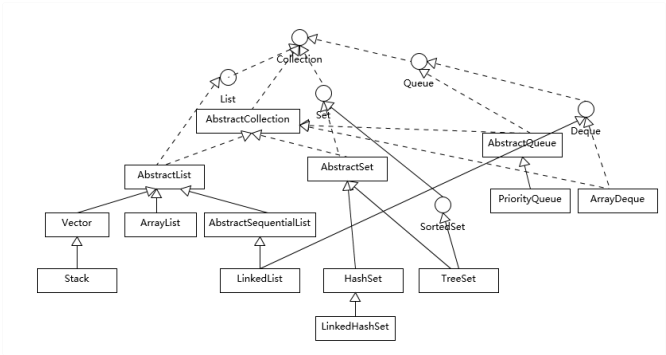
理问题和沟通时的思路。

以上只是一个方面的例子，建议学习相关书籍，如《算法导论》《编程珠玑》等，或相关[教程](#)。对于特定领域，比如推荐系统，建议咨询领域专家。单纯从面试的角度，很多朋友推荐使用一些算法网站如LeetCode等，帮助复习和准备面试，但坦白说我并没有刷过这些算法题，这也是仁者见仁智者见智的事情，招聘时我更倾向于考察面试者自身最擅长的东西，免得招到纯面试高手。

知识扩展

我们先一起来理解集合框架的整体设计，为了有个直观的印象，我画了一个简要的类图。注意，为了避免混淆，我这里没有把java.util.concurrent下面的线程安全容器添加进来；也没有列出Map容器，虽然通常概念上我们也会把Map作为集合框架的一部分，但是它本身并不是真正的集合（Collection）。

所以，我今天主要围绕狭义的集合框架，其他都会在专栏后面的内容进行讲解。



我们可以看到Java的集合框架，Collection接口是所有集合的根，然后扩展开提供了三大类集合，分别是：

- List，也就是我们前面介绍最多的有序集合，它提供了方便的访问、插入、删除等操作。
- Set，Set是不允许重复元素的，这是和List最明显的区别，也就是不存在两个对象equals返回true。我们在日常开发中有很多需要保证元素唯一性的场合。
- Queue/Deque，则是Java提供的标准队列结构的实现，除了集合的基本功能，它还支持类似先进先出（FIFO，First-in-First-Out）或者后入先出（LIFO，Last-In-First-Out）等特定行为。这里不包括BlockingQueue，因为通常是并发编程场合，所以被放置在并发包里。

每种集合的通用逻辑，都被抽象到相应的抽象类之中，比如AbstractList就集中了各种List操作的通用部分。这些集合不是完全孤立的，比如，LinkedList本身，既是List，也是Deque哦。

如果阅读过更多[源码](#)，你会发现，其实，TreeSet代码里实际默认是利用TreeMap实现的，Java类库创建了一个Dummy对象“PRESENT”作为value，然后所有插入的元素其实是以键的形式放入了TreeMap里面；同理，HashSet其实也是以HashMap为基础实现的，原来他们只是Map类的马甲！

就像前面提到过的，我们需要对各种具体集合实现，至少了解基本特征和典型使用场景，以Set的几个实现为例：

- TreeSet支持自然顺序访问，但是添加、删除、包含等操作要相对低效（log(n)时间）。
- HashSet则是利用哈希算法，理想情况下，如果哈希散列正常，可以提供常数时间的添加、删除、包含等操作，但是它不保证有序。
- LinkedHashSet，内部构建了一个记录插入顺序的双向链表，因此提供了按照插入顺序遍历的能力，与此同时，也保证了常数时间的添加、删除、包含等操作，这些操作性能略低于HashSet，因为需要维护链表的开销。
- 在遍历元素时，HashSet性能受自身容量影响，所以初始化时，除非有必要，不然不要将其背后的HashMap容量设置过大。而对于LinkedHashSet，由于其内部链表提供的方便，遍历性能只和元素多少有关系。

我今天介绍的这些集合类，都不是线程安全的，对于java.util.concurrent里面的线程安全容器，我在专栏后面会去介绍。但是，并不代表这些集合完全不能支持并发编程的场景，在Collections工具类中，提供了一系列的synchronized方法，比如

```
static <T> List<T> synchronizedList(List<T> list)
```

我们完全可以利用类似方法来实现基本的线程安全集合：

```
List list = Collections.synchronizedList(new ArrayList());
```

它的实现，基本就是将每个基本方法，比如get、set、add之类，都通过synchronized添加基本的同步支持，非常简单粗暴，但也非常实用。注意这些方法创建的线程安全集合，都符合迭代时fail-fast行为，当发生意外的开发修改时，尽早抛出ConcurrentModificationException异常，以避免不可预计的行为。

另外一个经常被考察到的问题，就是理解Java提供的默认排序算法，具体是什么排序方式以及设计思路等。

这个问题本身就是有点陷阱的意味，因为需要区分是Arrays.sort()还是Collections.sort()（底层是调用Arrays.sort()）；什么数据类型；多大的数据集（太小的数据集，复杂排序是没必要的，Java会直接进行二分插入排序）等。

- 对于原始数据类型，目前使用的是所谓双轴快速排序（Dual-Pivot QuickSort），是一种改进的快速排序算法，早期版本是相对传统的快速排序，你可以阅读[源码](#)。
- 而对于对象数据类型，目前则是使用TimSort，思想上也是一种归并和二分插入排序（binarySort）结合的优化排序算法。TimSort并不是Java的独创，简单说它的思路是查找数据集中已经排好序的分区（这里叫run），然后合并这些分区来达到排序的目的。

另外，Java 8引入了并行排序算法（直接使用parallelSort方法），这是为了充分利用现代多核处理器的计算能力，底层实现基于fork-join框架（专栏后面会对fork-join进行相对详细的介绍），当处理的数据集比较小的时候，差距不明显，甚至还表现差一点；但是，当数据集增长到数万或百万以上时，提高就非常大了，具体还是取决于处理器和系统环境。

<p>排序算法仍然在不断改进，最近双轴快速排序实现的作者提交了一个更进一步的改进，历时多年的研究，目前正在审核和验证阶段。根据作者的性能测试对比，相比于基于归并排序的实现，新改进可以提高随机数据排序速度提高10% ~ 20%，甚至在其他特征的数据集上也有几倍的提高，有兴趣的话你可以参考具体代码和介绍： http://mail.openjdk.java.net/pipermail/core-libs-dev/2018-January/051000.html。</p> <p>在Java 8之中，Java平台支持了Lambda和Stream，相应的Java集合框架也进行了大范围的增强，以支持类似为集合创建相应stream或者parallelStream的方法实现，我们可以非常方便的实现函数式代码。</p> <p>阅读Java源代码，你会发现，这些API的设计和实现比较独特，它们并不是实现在抽象类里面，而是以默认方法的形式实现在Collection这样的接口里！这是Java 8在语言层面的新特性，允许接口实现默认方法，理论上来说，我们原来实现在类似Collections这种工具类中的方法，大多可以转换到相应的接口上。针对这一点，我在面向对象主题，会专门梳理Java语言面向对象基本机制的演进。</p> <p>在Java 9中，Java标准类库提供了一系列的静态工厂方法，比如，List.of()、Set.of()，大大简化了构建小的容器实例的代码量。根据业界实践经验，我们发现相当一部分集合实例都是容量非常有限的，而且在生命周期中并不会进行修改。但是，在原有的Java类库中，我们可能不得不写成：</p> <pre>ArrayList<String> list = new ArrayList<>(); list.add("Hello"); list.add("World");</pre> <p>而利用新的容器静态工厂方法，一句代码就够了，并且保证了不可变性。</p> <pre>List<String> simpleList = List.of("Hello", "world");</pre> <p>更进一步，通过各种of静态工厂方法创建的实例，还应用了一些我们所谓的最佳实践，比如，它是不可变的，符合我们对线程安全的需求；它因为不需要考虑扩容，所以空间上更加紧凑等。</p> <p>如果我们去看of方法的源码，你还会发现一个特别有意思的地方：我们知道Java已经支持所谓的可变参数（varargs），但是官方类库还是提供了一系列特定参数长度的方法，看起来似乎非常不优雅，为什么呢？这其实是为了最优的性能，JVM在处理变长参数的时候会有明显的额外开销，如果你需要实现性能敏感的API，也可以进行参考。</p> <p>今天我从Vector、ArrayList、LinkedList开始，逐步分析其设计实现区别、适合的应用场景等，并进一步对集合框架进行了简单的归纳，介绍了集合框架从基础算法到API设计实现的各种改进，希望能对你的日常开发和API设计能够有帮助。</p> <p>一课一练</p> <p>关于今天我们讨论的题目你做到心中有数了吗？留一道思考题给你，先思考一个应用场景，比如你需要实现一个云计算任务调度系统，希望可以保证VIP客户的任务被优先处理，你可以利用哪些数据结构或者标准的集合类型呢？更进一步讲，类似场景大多是基于什么数据结构呢？</p> <p>请在留言区写写你对这个问题的思考，我会选出经过认真思考的留言，送给你一份学习鼓励金，欢迎你与我一起讨论。</p> <p>你的朋友是不是也在准备面试呢？你可以“请朋友读”，把今天的题目分享给好友，或许你能帮到他。</p>	
 <div><h1>Java 核心技术36讲</h1><p>—— Oracle 首席工程师 带你修炼 Java 内功 ——</p><p>杨晓峰 Oracle 首席工程师</p></div> 	
<p>雷霖霖的爸爸</p> <p>在这个题目下，自然就会想到优先级队列了，但还需要额外考虑vip再分级，即同等级vip的平权的问题，所以应该考虑除了直接的和vip等级相关的优先级队列优先级规则问题，还得考虑同等级多个客户互不相被单一客户大量任务阻塞的问题，数据结构确实是基础，即使这个思考题考虑的这个场景，待调度数据估计会放在redis里面吧</p> <p>作者回复</p>	<p>2018-05-22</p>
<p>赞</p> <p>孙晓刚</p>	<p>2018-05-23</p>
<p>精选第一个对于读写效率问题，我觉得表述有欠缺，或者说不能那么绝对。</p> <p>1、并不是所有的增删都会开辟新内存，没有开辟新内存的尾部增，效率也是杠杠的。</p> <p>2、尾部删除也不需要开辟新内存，只是移出最后一个对象。</p> <p>之前我也是接收了ArrayList的特性随机访问快，增删效率差。直到看到源码才知道，没那么绝对。直接导致结果就是本身适合使用ArrayList的场景会因为这个笼统的说法而选LinkedList</p> <p>作者回复</p>	<p>2018-05-22</p>
<p>嗯，我文中特意强调了不包括尾部</p>	<p>2018-05-23</p>
<p>L.B.Q.Y</p>	

请教老师个问题，Collection接口的声明是带范型的，其中定义的Object[] toArray()方法为什么不是范型方式的？有什么原因吗？ 作者回复	2018-05-23
按照Javadoc，我觉得这个方法设计目的，就是让调用者精确控制类型；里面声明了，toArray(new Object[0])等同于toArray() 公号-Java 大后端	2018-05-23
Vector、ArrayList、LinkedList均为线型的数据结构，但是从实现方式与应用场景中又存在差别。 1 底层实现方式 ArrayList内部用数组来实现，LinkedList内部采用双向链表实现，Vector内部用数组实现。 2 读写机制 ArrayList在执行插入元素是超过当前数组预定义的最大值时，数组需要扩容，扩容过程需要调用底层System.arraycopy()方法进行大量的数组复制操作；在删除元素时并不会减少数组的容量（如果需要缩小数组容量，可以调用trimToSize()方法）；在查找元素时要遍历数组，对于非null的元素采取equals的方式寻找。 LinkedList在插入元素时，须创建一个新的Entry对象，并更新相应元素的前后元素的引用；在查找元素时，需遍历链表；在删除元素时，要遍历链表，找到要删除的元素，然后从链表上将此元素删除即可。 Vector与ArrayList仅在插入元素时容量扩充机制不一致。对于Vector，默认创建一个大小为10的Object数组，并将capacityIncrement设置为0；当插入元素数组大小不够时，如果capacityIncrement大于0，则将Object数组的大小扩大为现有size+capacityIncrement；如果capacityIncrement<=0,则将Object数组的大小扩大为现有大小的2倍。 3 读写效率 ArrayList对元素的增加和删除都会引起数组的内存分配空间动态发生变化。因此，对其进行插入和删除速度较慢，但检索速度很快。 LinkedList由于基于链表方式存放数据，增加和删除元素的速度较快，但是检索速度较慢。 4 线程安全性 ArrayList、LinkedList为非线程安全；Vector是基于synchronized实现的线程安全的ArrayList。 需要注意的是：单线程应尽量使用ArrayList，Vector因为同步会有性能损耗，即使在多线程环境下，我们可以利用Collections这个类中为我们提供的synchronizedList(List list)方法返回一个线程安全的同步列表对象。 问题回答 利用PriorityBlockingQueue或Disruptor可实现基于任务优先级为调度策略的执行调度系统。 曹铮	2018-05-22
既然是Java的主题，那就用PriorityBlockingQueue吧。 如果是真实场景肯定会考虑高可用能持久化的方案。 其实我觉得应该参考银行窗口，同时三个窗口，就是三个队列，柜台就是消费者线程，某一个窗口vip优先，没有vip时也为普通客户服务。要实现，要么有个dispatcher，要么保持vip通道不许普通进入。vip柜台闲时从其他队列偷 作者回复	2018-05-22
有道理	2018-05-23
Miaozhe	2018-05-26
杨老师，问个问题，Collection接口下面已细化了List,Set和Queue子接口，未什么又定义了AbstractCollection这个抽象类？具体是什么考虑？以为我发现3个接口的子类都是集成这个抽象类。 作者回复	2018-05-28
三个都是Collection，总还是有共同行为的	2018-05-28
我奋斗去了	2018-05-22
可以使用priority queue，维护两个队列 一个VIP队列 一个普通用户队列 。当VIP队列有人的情况优先处理 作者回复	2018-05-23
为什么用两个队列，PriorityQueue不是有优先级了	2018-05-23
Leo	2018-05-22
使用优先级队列实现堆，可以根据优先级进行操作	2018-05-22
13683815260	2018-05-22
面试的重点HashMap,实现原理，扩展什么的，1.7和1.8的区别。还有和hashtable的异同。还有Juc下面集合的熟悉程度。 作者回复	2018-05-23
下两篇就是	2018-05-23
Hesher	2018-05-22
我觉得使用queue来实现VIP业务就行了，检查这个队列深度，大于0时就优先处理。分布式环境下用MQ实现。 13683815260	2018-05-22
面试重点HashMap，1.7和1.8然后还有hashtable,还有Juc下面的集合 且以深情共白头	2018-07-09
之前一直以为Verctor不属于集合，只是数组。学习了。针对VIP客户任务优先处理场景，认为采用SortSet进行，按照默认排序即可，数值越小优先级越高，和线程的优先级级别一致 作者回复	2018-07-11
和优先队列相比，不那么紧凑，例如treeset用的树比堆要多了节点开销	2018-07-11
Allen	2018-06-30
多级反馈队列	2018-06-30
Miaozhe	2018-05-28

再问个问题，为什么ArrayDeque类中的存储对象Object数组，数组的长度必须是2的n次方。	
Miaozhe	2018-05-28
今天看了一下PriorityQueue的源码，发现其是使用最小堆结构(二叉堆)，存放在数组中(数组索引对应树的从上到下，从左到右)。采用上面最小，每插入一个数据，就先与根节点比较，如果小于根节，依次换位置；大于根节点，就放在最后一个位置。	
Miaozhe	2018-05-28
杨老师，有个问题，TreeSet为什么不支持正序，只支持倒序(DescendingIterator)？Tree本身支持正序列。	
zjh	2018-05-26
比较片面的说，Java集合类底层基本上就是基于数组或者链表来实现的，数组的地址连续性决定了其随机存取速度较快，但是涉及到扩容则比较耗时，而链表则不存在扩容的性能消耗，但随机访问需要遍历地址因此相对数组要慢，所以判断一个集合的特点可以先判断是基于数组还是链表。	
L.B.Q.Y	2018-05-24
集合框架的那张图，SortedSet应该是接口，图里面画成类了，TreeSet看起来有两个父类了。	
作者回复	2018-05-25
感谢指出，已经记录下来去修正一下	
Lawt	2018-05-24
写得还挺不错的，担心36期能讲那么多没容吗？持续关注，相信您，期待	
悬崖丶	2018-05-23
想问一个小白问题，List接口继承了Collection接口，为什么List接口还要重写一遍Collection接口中的一些方法	
作者回复	2018-05-24
覆盖（override），面向对象多态的基本方面	
Miaozhe	2018-05-23
最近学习过程中，感觉看原代码比较吃力。一个Hash Map没有看明白，这么对Key的Set赋值，这么对value的Collection赋值。	
作者回复	2018-05-24
下一篇就是	
呵呵	2018-05-23
阅读速度太快了	
webwombat	2018-05-22
那个问题，应该是priority queue吧？操作系统的进程调度一般都是基于优先级队列来实现的。	
作者回复	2018-05-23
yes，人家可能进一步提出更多场景继续考	
王磊	2018-05-22
使用优先级队列，按照任务按照优先级排好序，这样优先级高的任务被优先处理。	

