

# JAVA

## 算法(算法图解/漫画算法)

### 数据结构

- 数组
  - `byte[][] b2 = new byte[1][1];`  
`byte b22[][] = new byte[2][2];`  
`byte[] b3 = { (byte)1,(byte)2,3 };`  
`byte[] b32 = new byte[]{ (byte)1,(byte)2,3 };`
- 链表
  - 单项链表: 单节点 节点中保存: 值, 下一个节点地址
    - HashMap(链表深度<8时)
  - 双线链表: 头节点, 尾节点  
节点中保存: 上一个节点地址,值,下一个节点地址;  
双向循环:头结点的before节点为尾节点地址,  
尾节点的after节点为头节点  
非循环: 头结点的before节点为NULL,  
尾节点的after节点为NULL
    - LinkedHashMap
  - 获取链表长度(不含头节点): 遍历
  - 获取链表某位置节点值: 1. 遍历获取总节点数; 2. 遍历到size - index 位置节点
- 栈
  - 支持Push(入栈)和Pop(出站)操作, 维护一个栈顶索引,栈大小,栈是否空,是否满。  
Push :往栈顶插入元素  
Pop: 弹出栈顶元素
  - 栈分类
    - 顺序栈(数组实现)
    - 链式栈(链表实现,单向链表,链表尾为栈顶)
  - 使用场景
    - 表达式计算
      - 表达式 $1+(4-6*3)/2=2$ 
        - 1. 表达式表示为后缀表达式 $1463*-2/+$ 。
        - 2. 使用栈来计算: 碰见数字就入栈, 碰见符号先运算
    - 递归(  
斐波那契数列( $f(1)=f(2)=1;f(n)=f(n-1)+f(n-2),n \geq 3;$ ),  
阶乘( $0!=1, f(n)=n*f(n-1)$ ))
      - 递归本身就是入栈
    - 括号匹配(匹配输入的括号是否全匹配)
      - 输入一个字符串 里面只含有 [ , ] , ( , ) 四种括号 ; 现要求判断这个字符串 是否满足括号匹配 。如 ([])() 是匹配的 ([])是不匹配的
      - 栈空时, 将字符Push栈中, 并判断, 当当前字符和栈顶字符匹配时, 出栈; 循环结束后, 若栈为空则都匹配, 反之不匹配

- #2个栈实现消息队列
  - 栈特点: 先进后出(FILO)
  - 队列特定: 先进先出(FIFO)
  - 实现: 栈1(S1),栈2(S2),  
生产数据时, 将数据Push到S1,  
消费数据时, 从S2中Pop数据,S2数据为空时,将S1数据全部Pop到S2中, 然后再从S2Pop数据  
(有部分人在S2 Pop数据后, 把剩余数据Pop回S1)
- java.util.Stack extends Vector
  - 数组实现, 默认大小10  
扩容 2倍大小+1
- 队列
- 哈希表

## 排序算法(10大经典排序算法)

- 冒泡排序
    - 概述: 重复走访过要排序的元素列, 依次比较两个相邻的元素, 如果顺序错误, 就把他们交换过来, 直到没有相邻的元素需要交换
    - 提供一个待排序数据, 依次比较相邻2个数的大小,按从小到大排序  
双层循环: 外层循环次数:  $arr.length - 1$   
内存循环次数:  $arr.length - i - 1$ 
      - 1. 在第一趟比较完成后, 最后一个数一定是数组中最大的一个数, 所以在比较第二趟的时候, 最后一个数是不参加比较的。
    - 2. 在第二趟比较完成后, 倒数第二个数也一定是数组中倒数第二大数, 所以在第三趟的比较中, 最后两个数是不参与比较的。
    - 3. 依次类推, 每一趟比较次数减少依次
      - 时间复杂度:  $O(n^2)$  稳定性: 稳定
      - 优化: 设置是否完成标记位, 若一趟中没有任何更新,则表示数组已排序完成, 结束后续的排序
- ```
int[] arr = new int[10];
boolean isFinished = false;
for (int i = 0; i < arr.length - 1; i++) {
    for (int j = 0; j < arr.length - i - 1; j++) {
        if (arr[j] > arr[j + 1]) {
            temp = arr[j];
            arr[j] = arr[j + 1];
            arr[j + 1] = temp;
            isFinished = true;
        }
    }
    if (!isFinished) {
        break;
    }
    isFinished = false;
}
```
- 选择排序

- 概述(原理):  
第一次从待排序的数据元素中选出最小(或最大)的一个元素, 存放到队列的起始位置;  
然后再从剩余的未排序的元素中选出最小(或最大)的一个元素, 然后放到已排序队列尾。  
以此类推, 直到待排序元素个数为0。  
(简单描述:待排序数组的每趟循环中, 选出最小(或最大)值放到已排序队列尾(第一趟循环时, 放到队列的其实位置), 知道所有循环完成)
- 时间复杂度:  $O(n^2)$  不稳定排序算法
- ```
public void sort(int[] arr) {
    for (int i = 0; i < arr.length-1; i++) {
        int minIdx = i;
        for (int j = i+1; j < arr.length; j++) {
            if (arr[minIdx] > arr[j]) {
                minIdx = j;
            }
        }
        if (i != minIdx) {
            int temp = arr[i];
            arr[i] = arr[minIdx];
            arr[minIdx] = temp;
        }
    }
}
```
- 插入排序
  - 概述(原理): 通过构建有序队列, 对于未排序数据, 在已排序序列中向后往前扫描, 找到相应的位置并插入。
  - 步骤:
    1. 将待排序序列的第一个元素看做有序序列, 把第二个元素到最后一个元素看着待排序序列;
    2. 从头到尾, 依次扫描待排序序列, 将待排序序列的每个元素, 从已排序序列尾依次向前扫描, 找到合适的位置插入(若在已排序序列中找到相同值, 则将该元素插入到已存在元素后)
  - 时间复杂度:  $O(n^2)$  稳定排序算法
  - ```
public void sort(int[] arr){
    for (int i = 1; i < arr.length; i++) {
        int nextVal = arr[i];
        int j = i-1;
        //目标值从已排序队列尾依次扫描,处理已排序队列值位置
        for (; j >= 0 && nextVal < arr[j]; j--) {
            arr[j + 1] = arr[j];
        }
        //将目标值插入到合适的位置
        arr[j + 1] = nextVal;
    }
}
```
  - 优化: 折半插入排序
    - 概述: 原理同直接插入排序;  
不同: 在扫描已排序序列时, 使用折半查找法查找插入位置
    - ```
public void binaryInsertSort(int[] arr) {
    for (int i = 1; i < arr.length; i++) {
        int nextVal = arr[i];
        //折半查找扫描位置
```

```

int low = 0;
int high = i - 1;
while (low <= high) {
    int mid = (high + low) / 2;
    if (arr[mid] > nextVal) {
        high = mid - 1;
    } else {
        low = mid + 1;
    }
}
for (int j = i; j > high+1; j--) {
    arr[j] = arr[j - 1];
}
arr[high + 1] = nextVal;
}
}

```

- 希尔排序

- 概述(原理):

希尔排序(Shell's Sort)又称 递减增量排序(Diminishing Increment Sort)，是一种插入排序的更高效的改进版本

基本思想:

先将整个待排序序列分割成若干个子序列分别进行插入排序，待整个序列基本有序时，再对全体记录进行一次插入排序。

分割增量逐次递减，直到1

希尔增量: 每次增量为原来的一半  $gap = gap / 2$

- 最差时间复杂度 $O(n^2)$  稳定性: 不稳定

- ```
public void sort(int[] arr) {
    //希尔增量,最坏时间复杂度 $O(n^2)$ 
    int gap = arr.length / 2;
    while (gap > 0) {
        //内部为插入排序
        for (int i = gap; i < arr.length; i++) {
            int tmp = arr[i];
            int j = i - gap;
            while (j >= 0 && arr[j] > tmp) {
                arr[j + gap] = arr[j];
                j -= gap;
            }
            arr[j + gap] = tmp;
        }
        //逐步减小增量,直到1
        gap = gap / 2;
    }
}
```

- 归并排序

- 概述: 建立在归并操作上的有效算法，是分治法的一种典型应用。

将已有序列的子序列合并，得到完整的有序序列；即先使每个子序列有序，再使序列段有序。

将2个有序子序列合成一个1个完整序列，成为二路归并。

原理:

1. 将序列每相邻的两个数字进行归并操作(merge),形成 $\text{floor}(n/2+n\%2)$ 个序列, 排序后每个序列包含2个元素,
2. 将上一步的序列再次归并, 形成 $\text{floor}(n/4)$ 个序列, 每个序列包含4个元素  
重复2步骤, 直到所有排序完毕。

- 稳定性: 稳定

- 时间复杂度:  $O(n \log n)$   
速度仅次于快速排序

- 实现方法

- ■ 1. 自上而下的递归
- ■ 2. 自下而上的迭代

- @Override

```
public void sort(int[] arr) {  
    int[] temp = new int[arr.length]; // 在排序前, 先建好一个长度等于原数组长度的临时数组, 避免递归中频繁开辟空间  
    sort(arr, 0, arr.length - 1, temp);  
}  
  
private static void sort(int[] arr, int left, int right, int[] temp) {  
    if (left < right) {  
        int mid = (left + right) / 2;  
        sort(arr, left, mid, temp); // 左边归并排序, 使得左子序列有序  
        sort(arr, mid + 1, right, temp); // 右边归并排序, 使得右子序列有序  
        merge(arr, left, mid, right, temp); // 将两个有序子数组合并操作  
    }  
}  
  
private static void merge(int[] arr, int left, int mid, int right, int[] temp) {  
    int i = left; // 左序列指针  
    int j = mid + 1; // 右序列指针  
    int t = 0; // 临时数组指针  
    while (i <= mid && j <= right) {  
        if (arr[i] <= arr[j]) {  
            temp[t++] = arr[i++];  
        } else {  
            temp[t++] = arr[j++];  
        }  
    }  
    while (i <= mid) { // 将左边剩余元素填充进temp中  
        temp[t++] = arr[i++];  
    }  
    while (j <= right) { // 将右序列剩余元素填充进temp中  
        temp[t++] = arr[j++];  
    }  
    t = 0;  
    // 将temp中的元素全部拷贝到原数组中  
    while (left <= right) {  
        arr[left++] = temp[t++];  
    }  
}
```

- Arrays.sort() 为优化版的归并排序

- 快速排序

- 概述:

快速排序又是一种分治法的典型应用。本质上是冒泡排序的递归分治法。

原理：通过一趟排序将要排序的序列分割成2个独立部分，其中一部分的所有数据比另一部分的所有数据都要小；然后再按此方法对这两部分数据分别进行快速排序，整个过程可以递归进行，以此达到整个数据变成有序序列

- 算法步骤

- ■ 1. 从数列中挑出一个元素，称为 "基准" (pivot) ；
- ■ 2. 重新排序数列，所有元素比基准值小的摆放在基准前面，所有元素比基准值大的摆在基准的后面（相同的数可以到任一边）。在这个分区退出之后，该基准就处于数列的中间位置。这个称为分区 (partition) 操作；
- ■ 3. 递归地 (recursive) 把小于基准值元素的子数列和大于基准值元素的子数列排序；

- 稳定性: 不稳定

- 时间复杂度:  $O(n \log(n))$

- @Override

```
public void sort(int[] arr) {
    quickSort(arr, 0, arr.length - 1);
}

private int[] quickSort(int[] arr, int left, int right) {
    if (left < right) {
        int partitionIndex = partition(arr, left, right);
        quickSort(arr, left, partitionIndex - 1);
        quickSort(arr, partitionIndex + 1, right);
    }
    return arr;
}

private int partition(int[] arr, int left, int right) {
    // 设定基准值 (pivot)
    int pivot = left;
    int index = pivot + 1;
    for (int i = index; i <= right; i++) {
        if (arr[i] < arr[pivot]) {
            swap(arr, i, index);
            index++;
        }
    }
    swap(arr, pivot, index - 1);
    return index - 1;
}

private void swap(int[] arr, int i, int j) {
    int temp = arr[i];
    arr[i] = arr[j];
    arr[j] = temp;
}
```

- 堆排序
- 计数排序
- 桶排序
- 基数排序

## 排序算法的稳定性

- 概述:

假定在待排序的记录序列中，存在多个具有相同的关键字的记录，若经过排序，这些记录的相对次序保持不变，即在原序列中， $r[i]=r[j]$ ，且 $r[i]$ 在 $r[j]$ 之前，而在排序后的序列中， $r[i]$ 仍在 $r[j]$ 之前，则称这种排序算法是稳定的；否则称为不稳定的。

- 常见排序算法的稳定性

- 不稳定

- 堆排序、快速排序、希尔排序、直接选择排序

- 稳定

- 基数排序、冒泡排序、直接插入排序、折半插入排序、归并排序

- 稳定性的意义:

如果只是简单的进行数字的排序，那么稳定性将毫无意义。

如果排序的内容仅仅是一个复杂对象的某一个数字属性，那么稳定性依旧将毫无意义

如果要排序的内容是一个复杂对象的多个数字属性，但是其原本的初始顺序毫无意义，那么稳定性依旧将毫无意义。

除非要排序的内容是一个复杂对象的多个数字属性，且其原本的初始顺序存在意义，那么我们需要在二次排序的基础上保持原有排序的意义，才需要使用到稳定性的算法。

例如要排序的内容是一组原本按照价格高低排序的对象，如今需要按照销量高低排序，使用稳定性算法，可以使得想同销量的对象依旧保持着价格高低的排序展现，只有销量不同的才会重新排序。

(当然，如果需求不需要保持初始的排序意义，那么使用稳定性算法依旧将毫无意义)

换句话说，以某种关键字的方式排序后，能不影响到其他关键字原来排序结果的方法就是稳定的，比如一开始按照价格高低排序结果为 a(10元，卖了5个) b(8元，卖了20个) c(6元，卖了20个) d(4元，卖了30个)，则按照销量重拍后如果保持 d(30个，价格为4元) b(20个，价格为8元) c(20个，价格为6元) a(5个，价格为10元)，则说明该方法为稳定的，而如果出现c在b前，破坏了排序前b在c前的顺序，则说明这个方法是不稳定的

## 查找算法

- 二分查找

- (树)广度优先搜索

- 它遍历一个层的所有节点，然后再进入下一层。

这种遍历也称为层序遍历，它从根开始，从左到右访问树的所有层。

(层序遍历需要用到队列)

- (树)深度优先搜索

- 在遍历下一个同级对象之前，它会在每个子节点中尽可能深入

- 有前序遍历，中序遍历，后序遍历

(针对父节点的遍历次序区分的)

## 动态规划

## 递归

- 分而治之，分治法

## 狄克斯特拉算法

## 贪婪算法

## 树

- 树

- 树 (Tree) 是 $n$  ( $n \geq 0$ )个结点的有限集。 $n=0$ 时称为空树。在任意一颗非空树中：
  - 1) 有且仅有一个特定的称为根 (Root) 的结点；
  - 2) 当 $n > 1$ 时，其余结点可分为 $m$  ( $m > 0$ )个互不相交的有限集 $T_1$ 、 $T_2$ 、.....、 $T_n$ ，其中每一个集合本身又是一棵树，并且称为根的子树。此外，树的定义还需要强调以下两点：
  - 1)  $n > 0$ 时根结点是唯一的，不可能存在多个根结点，数据结构中的树只能有一个根结点。
  - 2)  $m > 0$ 时，子树的个数没有限制，但它们一定是互不相交的。

- 节点的度

- 结点拥有的子树数目称为结点的度。

- 节点关系

- 双亲节点
- 孩子节点
- 兄弟节点

- 节点层次

- 根为第一层，根的孩子为第二层，以此类推

- 树的深度

- 树中结点的最大层次数称为树的深度或高度

- 二叉树(BinaryTree)

- 二叉树是 $n$  ( $n \geq 0$ )个结点的有限集合，该集合或者为空集（称为空二叉树），或者由一个根结点和两棵互不相交的、分别称为根结点的左子树和右子树组成。

- 特点

- 1. 每个节点最多有2棵子树，所以二叉树中不存在度大于2的节点
- 2. 左子树和右子树是有顺序的，次序不能任意颠倒
- 3. 即使树中某节点只有1棵子树，也要区分它是左子树还是右子树

- 性质

- 1. 在二叉树的第 $i$ 层上最多有 $2^{(i-1)}$ 个节点
- 2. 二叉树中如果深度为 $k$ ，那么最多有 $2^k - 1$ 个节点( $k \geq 1$ )
- 3.  $n_0 = n_2 + 1$   $n_0$ 表示度数为0的节点， $n_2$ 表示度数为2的节点数
- 4. 在完全二叉树中，具有 $n$ 个节点的完全二叉树的深度为 $\lfloor \log_2 n \rfloor + 1$  其中 $\lfloor \log_2 n \rfloor$ 是向下取整
- 5. 若对含 $n$ 个节点的完全二叉树从上到下切从左至右进行1至 $n$ 的编号，则完全二叉树中任意一个编号 $i$ 的特性为
  - (1) 若 $i=1$ ,则该节点是二叉树的跟，无双亲，否则，编号为 $\lfloor i/2 \rfloor$ 的节点为其双亲节点
  - (2) 若 $2i > n$ ，则该节点无左孩子，否则，编号为 $2i$ 的节点为其左孩子节点
  - (3) 若 $2i+1 > n$ ，则该节点无右孩子节点，否则编号为 $2i+1$ 的节点为其右孩子节点

- 斜树

- 左斜二叉树
  - 所有节点只有左子树
- 右斜二叉树
  - 所有节点只有右子树

- 满二叉树

- 在一棵二叉树中，如果所有分支节点都存在左子树和右子树，并且所有叶子都在同一层上，这样的二叉树成为满二叉树
- 特点
  - 1. 叶子只能出现在最下一层。出现在其它层就不可能达成平衡
  - 2. 非叶子节点的度一定是2



- 3. 在同样深度的二叉树中，满二叉树的节点个数最多，叶子数量最多。
- 完全二叉树
  - 对一棵具有n个节点的二叉树按层编号，如果编号为i( $1 \leq i \leq n$ )的节点与同样深度的满二叉树中编号为i的节点在二叉树中位置完全相同，则这棵二叉树成为完全二叉树
  - 特点
    - 1. 叶子节点只能出现在最下层和次下层
    - 2. 最下层的叶子节点集中在树的左部
    - 3. 倒数第二层若存在叶子节点，一定在右部连续位置
    - 4. 如果节点度为1，则该节点只有左孩子，没有右子树
    - 5. 同样节点数目的二叉树，完全二叉树深度最小
    - (满二叉树一定是完全二叉树，反之不一定)
- 二叉树的存储结构
  - 顺序存储(一维数组)
    - 二叉树的顺序存储结构就是使用一维数组存储二叉树中的结点，并且结点的存储位置，就是数组的下标索引。
    - 顺序存储一般适用于完全二叉树
  - 二叉链表(带有数据和2个指针域的链表)
    - 既然顺序存储不能满足二叉树的存储需求，那么考虑采用链式存储。由二叉树定义可知，二叉树的每个结点最多有两个孩子。因此，可以将结点数据结构定义为一个数据和两个指针域。
- 二叉树遍历
  - 从二叉树的根节点出发，按照某种次序一次访问二叉树中的所有节点，使得每个节点被访问一次，且仅被访问一次。
  - 四种遍历方式
    - 前序遍历（深度优先搜索）
      - 从二叉树的根节点出发，当第一次到达结点时就输出节点数据，按照先向左再向右的方向访问
      - ABDH I E J C F G
    - 中序遍历（深度优先搜索）
      - 从二叉树的根节点出发，当第二次到达节点时输出结点数据（第一次到达时不输出），按照先向左再向右的访问访问
      - HDI BJE A F C G
    - 后序遍历（深度优先搜索）
      - 从二叉树的根节点出发，当第三次到达结点时就输出结点数据，按照先向左再向右的方向访问
      - HID JEB F G C A
    - 层序遍历（广度优先搜索）
      - 按照树的层次自上而下遍历二叉树
      - ABCDEFGHIJ
  - 已知前序遍历序列和后序遍历序列，不可以唯一确定一棵二叉树。
- 二叉树的操作
  - <https://www.baeldung.com/java-binary-tree>
  - 创建树(使用二叉链表)
  - 插入数据
    - 1. 根为空时，先创建根

- 2. 若值大于当前节点，则将新节点放到右子树
- 3. 若值小于当前节点，则将新节点放到左子树
- 4. 当当前节点值为null时，则将新节点插入该位置
- ```
private Node<T> insertToNode(Node<T> curr, T value) {
    if (curr == null) {
        return new Node<>(value);
    }
    if (curr.compare(value) > 0) {
        curr.left = insertToNode(curr.left, value);
    } else if (curr.compare(value) < 0) {
        curr.right = insertToNode(curr.right, value);
    }
    //若值相同,则节点已存在
    return curr;
}
```
- 查找数据
  - 1. 值等于根节点值时，返回根
  - 2. 值大于当前节点值时，搜索节点右子树
  - 3. 值小于当前节点值时，搜索节点左子树
  - ```
public Node<T> get(Node<T> curr, T value) {
    if (curr == null) {
        return null;
    }
    if (curr.compare(value) == 0) {
        return curr;
    }
    return curr.compare(value) > 0 ? get(curr.left, value) : get(curr.right, value);
}
```
- 删除数据
  - ```
public Node<T> remove(T value) {
    return root = remove(root, value);
}

/**
```
- 删除节点(节点可能需要重组)
- @param curr
- @param value
- @return
- ```
*/
private Node<T> remove(Node<T> curr, T value) {
    if (curr == null) {
        return null;
    }
    //节点重组
    if (curr.compare(value) == 0) {
        // Node to delete found
        // 一个节点没有子节点
        if (curr.left == null && curr.right == null) {
            return null;
        }
        //一个节点只有一个子节点-在父节点中，我们用唯一的子节点替换该节点。
        if (curr.right == null) {
```

```

        return curr.left;
    }
    if (curr.left == null) {
        return curr.right;
    }
    //一个节点有两个孩子 -这是最复杂的情况，因为它需要对树进行重组
    //从右结点中找到最小值的节点
    T smallestValue = (T) findSmallestValue(curr.right);
    curr.value = smallestValue;
    //将最小值节点删除
    curr.right = remove(curr.right, smallestValue);
    return curr;
}
if (curr.compare(value)>0) {
    curr.left = remove(curr.left, value);
    return curr;
}
curr.right = remove(curr.right, value);
return curr;
}
/**

```

- 找出结点中最小值的结点

- @param root

- @return

```

*/
private T findSmallestValue(Node<T> root) {
    return root.left == null ? root.value : (T)findSmallestValue(root.left);
}

```

- 4种遍历

- 前序遍历

- public void prevOrderIterator(Node<T> curr) {
 if (curr == null) {
 return;
 }
 System.out.print(curr.getValue() + ",");
 prevOrderIterator(curr.left);
 prevOrderIterator(curr.right);
 }

- 中序遍历

- public void midOrderIterator(Node<T> curr) {
 if (curr == null) {
 return;
 }
 midOrderIterator(curr.left);
 System.out.print(curr.getValue() + ",");
 midOrderIterator(curr.right);
 }

- 后序遍历

- public void postOrderIterator(Node<T> curr) {
 if (curr == null) {
 return;
 }

```

    }
    postOrderIterator(curr.left);
    postOrderIterator(curr.right);
    System.out.print(curr.getValue() + ",");
}

```

#### ■ 层序遍历

```

■ public void levelOrderIterate(Node<T> root) {
    if (root == null) {
        return;
    }
    //add(队列满,抛异常)/offer(队列满,返回false)/put(队列满,阻塞)
    //remove(移除数据,队列空,抛异常)/poll(取数据,队列空,返回false)/take(取数据,队列空,阻塞)
    //element(从队列头查询元素,队列空,抛异常)/peek(队列为空,返回null)
    Queue<Node> queue = new LinkedList();
    ((LinkedList<Node>) queue).add(root);
    Node<T> front;
    while (!queue.isEmpty()) {
        front = queue.remove();
        if (front.left != null) {
            ((LinkedList<Node>) queue).add(front.left);
        }
        if (front.right != null) {
            ((LinkedList<Node>) queue).add(front.right);
        }
        System.out.print(front.getValue()+"");
    }
}

```

#### ■ 获取根深度

```

■ public int length(Node<T> root) {
    return root == null ? 0 : 1 + Math.max(length(root.left), length(root.right));
}

```

#### ■ 二叉树的打印

##### ■ 水平树(前序遍历, 深度优先)

```

■ /**

```

##### ■ 打印水平树2-中序遍历

```

*/

```

```

public void printHorizontalTree2() {
    System.out.println(inOrderPrint(tree.getRoot(), new StringBuffer(), "", ""));
}

public String inOrderPrint(BinaryTree.Node<T> root, StringBuffer sb, String pointer, String parentPadding) {
    if (root == null) {
        return "";
    }
    String paddings = getPadding(level(root)) + parentPadding;
    inOrderPrint(root.getLeft(), sb, LEFT_SUBTREE_FLAG2, paddings);
    sb.append(paddings + " ");
    sb.append(pointer).append(root.getValue()).append(NEWLINE_FLAG);
    inOrderPrint(root.getRight(), sb, RIGHT_SUBTREE_FLAG, paddings);
    return sb.toString();
}

```

```

    }
    public String getPadding(int level) {
        StringBuffer sb = new StringBuffer();
        for (int i = 0; i < level; i++) {
            sb.append(" ");
        }
        return sb.toString();
    }
}

```

- 垂直树(层序遍历, 广度优先)
- 平衡二叉树(AVLTree)
  - 自平衡二叉查找树
  - 在AVL树中, 任一节点对应的两颗子树的最大高度差为1, 因此也被称为高度平衡树。插入和查找的最坏时间复杂度 $O(\log n)$ 。增加和删除元素的操作则可能需要一次或多次树旋转, 以实现树平衡
  - 平衡因子
    - 某个节点的左子树与右子树的高度(深度)差即为该节点的平衡因子(BF,BalanceFactor);  
平衡二叉树中不存在平衡因子大于1的节点。  
在平衡二叉树中, 节点的平衡因子只能取0,1,-1分别对应左右子树等高, 左子树高, 右子树高
  - 旋转
    - AVL树的4中节点插入方式
      - LL 在A的左子树根节点的左子树上插入节点而破坏平衡, 右旋
      - RR 在A的右子树根节点的右子树上插入节点而破坏平衡, 左旋
      - LR 在A的左子树的根节点的右子树上插入节点而破坏平衡, 先左旋后右旋
      - RL 在A的右子树根节点的左子树上插入节点而破坏平衡, 先右旋后左旋
- 子主题 4
- 红黑树

## 设计模式

---

### 七大设计原则

- 单一职责原则(Single Responsibility Principle)
  - 一个类或者模块应该有且只有一个改变的原因  
一个类/接口/方法只负责一项职责或职能
  - 优点
    - 降低类的复杂度;
    - 提高类的可读性, 因为类的职能单一, 看起来比较有目的性, 显得简单;
    - 提高系统的可维护性, 降低变更程序引起的风险。
- 开闭原则(Open Close Principle)
  - 一个软件实体,如类、模块和函数应该对扩展开放,对修改关闭.即一个软件实体应该通过扩展来实现变化,而不是通过修改已有的代码来实现变化。
  - 优点
    - 开闭原则有利于进行单元测试
    - 开闭原则可以提高复用性
    - 开闭原则可以提高可维护性

- 面向对象开发的要求
- 里氏替换原则(Liskov Substitution Principle)
  - 里氏代换原则(Liskov Substitution Principle LSP)面向对象设计的基本原则之一。  
里氏代换原则中说，任何基类可以出现的地方，子类一定可以出现。  
LSP是继承复用的基石，只有当衍生类可以替换掉基类，软件单位的功能不受到影响时，基类才能真正被复用，而衍生类也能够在基类的基础上增加新的行为。里氏代换原则是对“开-闭”原则的补充。  
实现“开-闭”原则的关键步骤就是抽象化。而基类与子类的继承关系就是抽象化的具体实现，所以里氏代换原则是对实现抽象化的具体步骤的规范
  - 具体约束
    - 子类必须实现父类的抽象方法，但不得重写父类的非抽象(已实现的)方法。
    - 子类中可增加自己特有的方法。(可以随时扩展)
    - 当子类覆盖或者实现父类的方法时,方法的前置条件(方法形参)要比父类输入参数更加宽松。否则会调用到父类的方法。
    - 当子类的方法实现父类的抽象方法时，方法的后置条件（即方法的返回值）要比父类更严格。否则会调用到父类的方法。
  - 最佳实践
    - 我们最好将父类定义为抽象类，并定义抽象方法，让子类重新定义这些方法，当父类是抽象类时候，父类不能实例化
- 依赖倒置原则(Dependence Inversion Principle)
  - 这个是开闭原则的基础，具体内容：针对接口编程，依赖于抽象而不依赖于具体。
    - 1、上层模块不应该依赖底层模块，它们都应该依赖于抽象。
    - 2、抽象不应该依赖于细节，细节应该依赖于抽象。
- 接口隔离原则(Interface Segregation Principle)
  - 1、客户端不应该依赖它不需要的接口。
  - 2、类间的依赖关系应该建立在最小的接口上。  
含义是：要为各个类建立它们需要的专用接口，而不要试图去建立一个很庞大的接口供所有依赖它的类去调用。
- 迪米特法则(Demeter Principle)
  - 一个实体应当尽量少的与其他实体之间发生相互作用，使得系统功能模块相对独立  
如果两个软件实体无须直接通信，那么就不应当发生直接的相互调用，可以通过第三方转发该调用。其目的是降低类之间的耦合度，提高模块的相对独立性。
- 组合复用原则(Composite Reuse Principle)
  - 尽量使用合成/聚合的方式，而不是使用继承

## 设计模式(23)

- 创建型(5)
  - 工厂方法(FactoryMethod)
    - 概述
      - 定义一个用于创建对象的接口，让子类决定实例化哪一个类。FactoryMethod使一个类的实例化延迟到其子类。
        - 普通工厂模式
        - 多个工厂方法模式
        - 静态工厂方法模式
    - 样例
      - JDK: J.U.C.ThreadFacotry
      - Mybatis: org.apache.ibatis.datasource.DataSourceFactory

- JndiDataSourceFactory
  - PooledDataSourceFactory
- Mybatis: org.apache.ibatis.transaction.TransactionFactory
  - JdbcTransactionFactory
  - ManagedTransactionFactory
- 适用性
  - 1. 当一个类不知道它所必须创建的对象类的時候。
  - 2. 当一个类希望由它的子类来指定它所创建的对象的时候。
  - 3. 当类将创建对象的职责委托给多个帮助子类中的某一个，并且你希望将哪一个帮助子类是代理者这一信息局部化的时候。
- 抽象工厂(AbstractFactory)
  - 概述
    - 提供一个创建一系列相关或相互依赖对象的接口，而无需指定它们具体的类。  
抽象工厂模式是对工厂方法模式的再升级，但是二者面对的场景稍显差别。  
工厂方法模式面对的目标一般都是单类的，就比如《Java设计模式之《工厂方法模式》及使用场景》中所举的例子，目标就是桌子这一类商品。  
抽象工厂模式面对的是一个组合体。
  - 样例
    - 动物抽象工厂,生产猫,狗
  - 适用性
    - 1. 一个系统要独立于它的产品的创建、组合和表示时。
    - 2. 一个系统要由多个产品系列中的一个来配置时。
    - 3. 当你要强调一系列相关的产品对象的设计以便进行联合使用时。
    - 4. 当你提供一个产品类库，而只想显示它们的接口而不是实现时。
- 单例模式(Singleton)
  - 概述
    - 保证一个类仅有一个实例，并提供一个访问它的全局访问点。
  - 样例
    - 双重检查锁(DCL)单例 (延迟模式,利用synchronized,volatile)
    - 枚举单例(立即模式,利用枚举属性在编译后为new枚举对象加载)
    - 静态内部类(延迟模式,利用静态属性和静态代码块实现单例对象的加载)
  - 用途
    - 配置文件加载
    - SpringBean加载
- 建造者模式(Builder)
  - 概述
    - 将一个复杂对象的构建与它的表示分离，使得同样的构建过程可以创建不同的表示。
  - 样例
    - StringBuilder,StringBuffer
    - Guava中的ImmutableMap.ImmutableSet等
    - Mybatis中的Example, SqlSessionFactoryBuilder 等
- 原型模式(Prototype)
  - 概述
    - 用原型实例指定创建对象的种类，并且通过拷贝这些原型创建新的对象。
  - 样例

- 实现Cloneable，重写clone方法创建新对象  
深复制可以使用ObjectInput/OutputStream
- 结构型(7)
  - 适配器模式(Adapter)
    - 概述
      - 将一个类的接口转换成客户希望的另外一个接口。Adapter模式使得原本由于接口不兼容而不能一起工作的那些类可以一起工作。  
适配器模式有三种：类适配器、对象适配器、接口适配器
        - 类适配器
          - 通过继承类，实现对某个类方法的调用
        - 对象适配器
          - 通过传入类实例对象，实现对类方法的调用
        - 接口适配器
          - 对于一个存在n个方法的接口，只会用到某几个方法，通过构造抽象类实现(实现方法体为空)该接口的所有方法，  
然后继承该抽象类实现所需的方法。
    - 适用性
      - 1.你想使用一个已经存在的类，而它的接口不符合你的需求。
      - 2.你想创建一个可以复用的类，该类可以与其他不相关的类或不可预见的类（即那些接口可能不一定兼容的类）协同工作。
      - 3.（仅适用于对象Adapter）你想使用一些已经存在的子类，但是不可能对每一个都进行子类化以匹配它们的接口。对象适配器可以适配它的父类接口。
    - 样例
      - JDK(对象适配器): java.util.Arrays.asList()
      - SpringMVC(接口适配器):  
org.springframework.web.servlet.config.annotation.  
WebMvcConfigurerAdapter  
implements WebMvcConfigurer(interface)
  - 桥接模式(Bridge)
    - 概述
      - 将抽象部分与它的实现部分分离，使它们都可以独立地变化。  
这种模式涉及到一个作为桥接的接口，使得实体类的功能独立于接口实现类。这两种类型的类可被结构化改变而互不影响。
    - 适用性
      - 1.你不希望在抽象和它的实现部分之间有一个固定的绑定关系。  
例如这种情况可能是因为，在程序运行时刻实现部分应可以被选择或者切换。
      - 2.类的抽象以及它的实现都应该可以通过生成子类的方法加以扩充。  
这时Bridge模式使你可以对不同的抽象接口和实现部分进行组合，并分别对它们进行扩充。
      - 3.对一个抽象的实现部分的修改应对客户不产生影响，即客户的代码不必重新编译。
      - 4.正如在意图一节的第一个类图中所示的那样，有许多类要生成。  
这样一种类层次结构说明你必须将一个对象分解成两个部分。
      - 5.你想在多个对象间共享实现（可能使用引用计数），但同时要求客户并不知道这一点。
    - 注意点
      - 1、定义一个桥接口，使其与一方绑定，这一方的扩展全部使用实现桥接口的方式。



2、定义一个抽象类，来表示另一方，在这个抽象类内部要引入桥接口，而这一方的扩展全部使用继承该抽象类

- 样例

- JDK: JDBC数据库访问接口

- 组合器模式(Composite)

- 概述

- 将对象组合成树形结构以表示"部分-整体"的层次结构。"Composite使得用户对单个对象和组合对象的使用具有一致性。"

- 适用性

- 1.你想表示对象的部分-整体层次结构。
  - 2.你希望用户忽略组合对象与单个对象的不同，用户将统一地使用组合结构中的所有对象。

- 样例

- 文件目录
  - 教材的章节知识点

- 装饰模式(Decorator)

- 概述

- 动态地给一个对象添加一些额外的职责。就增加功能来说，Decorator模式相比生成子类更为灵活。

- 装饰者与被装饰者需要有共同的超类或接口

- 适用性

- 1.在不影响其他对象的情况下，以动态、透明的方式给单个对象添加职责。
  - 2.处理那些可以撤消的职责。
  - 3.当不能采用生成子类的方法进行扩充时。

- 样例

- 饮料: 雪碧，美年达饮料的描述和价格
  - JavaIO InputStream

- 与代理模式的区别

- 装饰模式在于对原有系统业务的扩展或增强
  - 代理模式在于增加一些与业务无关的操作

- 外观模式(Facade)

- 概述

- 为子系统的一组接口提供一个一致的界面，Facade模式定义了一个高层接口，这个接口使得这一子系统更加容易使用。

- 适用性

- 1.当你要为一个复杂子系统提供一个简单接口时。子系统往往因为不断演化而变得越来越复杂。大多数模式使用时都会产生更多更小的类。这使得子系统更具可重用性，也更容易对子系统进行定制，但这也给那些不需要定制子系统的用户带来一些使用上的困难。

- Facade可以提供简单的缺省视图，这一视图对大多数用户来说已经足够，而那些需要更多的可定制性的用户可以越过facade层。

- 2.客户程序与抽象类的实现部分之间存在着很大的依赖性。引入facade将这个子系统与客户以及其他的子系统分离，可以提高子系统的独立性和可移植性。

- 3.当你需要构建一个层次结构的子系统时，使用facade模式定义子系统中每层的入口点。如果子系统之间是相互依赖的，你可以让它们仅通过facade进行通讯，从而简化了它们之间的依赖关系。

- 样例

- Controller 层访问Service层方法，可将各个Service注册到一个门面上，在Controller可以直接通过门面入口访问各个Service
  - 享元模式(Flyweight)
    - 概述
      - 运用共享技术有效地支持大量细粒度的对象。
    - 适用性
      - 1.一个应用程序使用了大量的对象。
      - 2.完全由于使用大量的对象，造成很大的存储开销。
      - 3.对象的大多数状态都可变为外部状态。
      - 4.如果删除对象的外部状态，那么可以用相对较少的共享对象取代很多组对象。
      - 5.应用程序不依赖于对象标识。由于Flyweight对象可以被共享，对于概念上明显有别的对象，标识测试将返回真值。
    - 样例
      - JVM字符串常量池
  - 代理模式(Proxy)
    - 概述
      - 为其他对象提供一种代理以控制对这个对象的访问。
        - 与装饰模式类似，代理模式更适用于增加无关业务的处理；或对原始对象的保护
    - 适用性
      - 1.远程代理（RemoteProxy）为一个对象在不同的地址空间提供局部代表。
      - 2.虚代理（VirtualProxy）根据需要创建开销很大的对象。
      - 3.保护代理（ProtectionProxy）控制对原始对象的访问。
      - 4.智能指引（SmartReference）取代了简单的指针，它在访问对象时执行一些附加操作。
    - 3种代理
      - 静态代理
        - 即普通代理，代理类中将被代理类作为属性来调用被代理类的方法
      - 动态代理
        - 即使用JDK内置方法生成代理类：
          - 调用java.lang.reflect.Proxy类的静态方法newProxyInstance即可，该方法会返回代理类对象
          - static Object newProxyInstance(ClassLoader loader, Class<?>[] interfaces,InvocationHandler h)
          - 接收的三个参数依次为：
            - ClassLoader loader：指定当前目标对象使用类加载器，写法固定
            - Class<?>[] interfaces：目标对象实现的接口的类型，写法固定
            - InvocationHandler h：事件处理接口，需传入一个实现类，一般直接使用匿名内部类
      - Cglib代理
        - /\*\*
- Cglib子类代理工厂
  - \*/
  - public class ProxyFactory implements MethodInterceptor{
  - // 维护目标对象
  - private Object target;
  - public ProxyFactory(Object target) {

```

        this.target = target;
    }
    // 给目标对象创建一个代理对象
    public Object getProxyInstance(){
        //1.工具类
        Enhancer en = new Enhancer();
        //2.设置父类
        en.setSuperclass(target.getClass());
        //3.设置回调函数
        en.setCallback(this);
        //4.创建子类(代理对象)
        return en.create();
    }
    @Override
    public Object intercept(Object obj, Method method, Object[] args, MethodProxy proxy)
    throws Throwable {
    }
}

```

- 静态代理和JDK代理有一个共同的缺点，就是目标对象必须实现一个或多个接口，加入没有，则可以使用Cglib代理。

- 样例

- Spring中的对象使用CGLIB动态代理

- 行为型(11)

- 责任链模式(COR)

- 概述

- 使多个对象都有机会处理请求，从而避免请求的发送者和接收者之间的耦合关系。将这些对象连成一条链，并沿着这条链传递该请求，直到有一个对象处理它为止。

这一模式的想法是，给多个对象处理一个请求的机会，从而解耦发送者和接受者。

- 适用性

- 1.有多个的对象可以处理一个请求，哪个对象处理该请求运行时刻自动确定。
      - 2.你想在不明确指定接收者的情况下，向多个对象中的一个提交一个请求。
      - 3.可处理一个请求的对象集合应被动态指定。

- 样例

- Servlet.Jsp中的filter

- 编码方式

- 1. 责任链模式写法1: 接口方法中传入责任链对象,有责任链对象调度各个责任对象处理
- ```
void handle(String request, IResponsibility responsibility);
```

- 2. 写法2: 责任对象中设置下一个处理者责任对象

```
IResponsibility next;
```

- 命令模式(Command)

- 概述

- 将一个请求封装为一个对象，从而使你可用不同的请求对客户进行参数化；对请求排队或记录请求日志，以及支持可撤消的操作。

- 适用性

- 1.抽象出待执行的动作以参数化某对象。

- 2.在不同的时刻指定、排列和执行请求。
  - 3.支持取消操作。
  - 4.支持修改日志，这样当系统崩溃时，这些修改可以被重做一遍。
  - 5.用构建在原语操作上的高层操作构造一个系统。
- 样例
- 参与者
  - 接收者: 命令最终执行对象
  - 命令接口/命令接口实现: 创建不同的命令实现对象，封装接收者,定义对外接口调用接收者具体实现
  - 请求者: 封装对命令接口的调用，提供给客户端
  - 客户端: 负责创建命令，并指定其接收者，并使用请求者执行操作
- 解释器模式(Interpreter)
  - 概述
    - 给定一个语言，定义它的文法的一种表示，并定义一个解释器，这个解释器使用该表示来解释语言中的句子
  - 适用性
    - 当一个语言需要解释执行时，并且可以将语言中的句子表示为抽象语法树时，可以使用解释器模式
  - 参与者
    - 定义一个解释器的抽象表达式，抽象一个解释操作，该接口为抽象语法树的所有节点共享
    - 实现一个终结符表达式操作
    - 实现一个非终结符表达式操作
    - 定义Context,包含除解释器外的全局信息
    - 定义Client,构建(或被给定)表示该文法定义的语言中一个特定的句子的抽象语法树。  
该抽象语法树由NonterminalExpression和TerminalExpression的实例装配而成。  
调用解释操作。
  - 样例
- 迭代器模式(Iterator)
  - 概述
    - 提供一种方法顺序遍历聚合对象的各种元素，而又暴露对象的内部表示
  - 适用性
    - 访问一个聚合对象而无需暴露对象的内部表示
    - 支持集合对象的多种表示
    - 为遍历不同的聚合对象提供一个统一的接口（即多态迭代）
  - 参与者
    - Iterator 迭代器，定义访问和遍历元素的接口
    - ConcreteIterator 具体的迭代器实现；对该聚合对象记录遍历位置
    - Aggregate 聚合对象，定义创建相关迭代器对象的接口
    - ConcreteAggregate 具体的聚合对象，实现创建迭代器对象的接口，返回ConcreteIterator
  - 样例
    - Iterable  
List  
Collection
- 中介者模式(Mediator)

- 概述
  - 用一个中介者来封装一系列对象的交互。中介者使各对象不需要显示的相互引用，从而使耦合松懈，并且能够独立改变它们的交互
- 适用性
  - 1. 一组对象以定义良好但复杂通信，使得类结构体层级复杂且难以理解
  - 2. 一个对象引用很多其他对象，并且直接调用对象进行通信，使得该对象难以复用
  - 3. 想定制一个分布在多个类的行为，而不想生成太多的子类
- 参与者
  - 1. 中介者: 定义各同事间交互的操作接口
  - 2. 中介者具体实现: 了解维护各同事对象，协调各同事对象间的相互操作
  - 3. 同事类: 每个同事类都知道中介者；每个同事对象与其他对象交互时，都与中介者通信

- 样例

- 备忘录模式(Memento)

- 概述
  - 在不破坏封装的前提下获取对象的内部状态，并在对象外部保存该状态，使它在之后可以恢复到先前保存的状态
- 适用性
  - 1. 保存一个对象在某个时刻的(部分)状态，这样能够在需要的时候恢复到先前的状态
  - 2. 如果用一个接口来让其他对象直接获取到这些状态，则会暴露对象的实现逻辑和破坏封装性
- 参与者
  - 1. Memento 备忘录，保存原发器的内部状态
  - 2. Originator 原发器，创建备忘录；使用备忘录恢复先前状态
  - 3. Caretake 管理者：保存备忘录，但不能操作和检查备忘录内容
- 样例(应用场景)
  - 浏览器回退操作
  - 编辑撤销与重做
  - 棋盘悔棋

- 观察者模式(Observer)

- 概述
  - 定义对象间的一种一对多的依赖关系，在一个对象发生改变时，所有依赖它的对象都会被通知并自动更新
- 适用性
  - 1. 当一个抽象模型有2个方面，其中一个方面依赖于另一个方面。将二者封装在独立的对象中，以使它们可以各自独立修改和复用
  - 2. 当一个对象改变时需要通知其他对象，而不知道有多少对象需要改变
  - 3. 当一个对象必须通知其他对象，而不能确定其他对象是谁
- 参与者
  - 1. Subject 目标对象，知道所有的观察者；可以有任意多个观察者观察同一目标；提供注册和删除观察者的接口
  - 2. Observer 观察者，提供更新观察者的接口
  - 3. ConcreteSubject: 具体目标对象，维护观察者列表。状态在发生改变时，通知所有观察者
  - 4. ConcreteObserver: 实现更新接口，使状态与目标状态一致
- 样例(应用场景)

- 下单后的各个渠道的通知
- 状态模式(State)
  - 概述
    - 在一个对象的状态改变时允许改变它的行为，看起来这个对象改变了它的类
  - 适用性
    - 1. 一个对象的行为取决于它的状态，并且在运行时根据状态改变它的行为
    - 2. 一个对象中包含庞大的多分支条件语句，且这些分支依赖于对象的状态  
这些状态通常由一个或多个常量枚举组成  
State模式将每个分支条件封装成一个独立的类中，使得可以根据具体的状态封装成具体的对象，而该对象可以独立的变化却不依赖其他对象。
  - 参与者
    - 1. Context 上下文对象, 定义客户端所期望的接口，维护一个ConcreteState实例，以保存当前状态
    - 2. State 抽象状态，定义一个与Context状态相关的一个接口
    - 3. ConcreteStatesubclasses 具体状态子类，各子类实现一个与Context状态相关的一个接口
  - 样例(应用场景)
    - 1. 订单状态的切换处理
  - 与策略模式的区别
    - 1. 状态模式的着重点在于不同的状态切换做不同的事情；  
策略模式是根据具体情况选择策略，不进行状态切换。
    - 2. 状态模式在不同的状态下事情不同，而策略模式做的是同意件事情，如支付功能，支付宝，微信支付属于不同的策略，可以相互替代。但状态模式每个状态的功能不同，不能相互替代。  
状态模式各子状态类依赖Context上下文状态  
策略模式不依赖Context, Context中调用各策略类。
- 策略模式(Strategy)
  - 概述
    - 定义一系列算法，将它们——分装起来，使得可以相互替换。策略模式使算法可以独立于它的使用者变化。
  - 适用性
    - 1. 许多相关的类仅仅是"行为各异"，"策略模式"提供了用多个行为的一个行为来配置一个类的方法
    - 2. 需要使用算法的不同变体
    - 3. 算法使用客户不应该知道的数据。可以使用策略模式避免暴露复杂的，与算法相关的数据结构
    - 4. 一个类中定义了多个行为，并且这些行为以个条件语句的形式出现；可以将相关条件分支移入他们的Strategy类中以替换条件语句
  - 参与者
    - 1. Strategy 抽象策略接口，定义全部算法公共接口
    - 2. ConcreteStrategy 具体抽象类，实现具体算法
    - 3. StrategyContext 策略上下文，维护一个ConcreteStrategy对象，持有一个Strategy引用，  
定义一个接口让Strategy访问它的数据
  - 样例(应用场景)
    - 多渠道通知
    - 生成多线路视频地址
- 模版方法(TemplateMethod)

- 概述

- 定义一个操作中算法的框架，而将一些步骤延迟到子类中。模版方法模式使得子类可以不改变一个算法的结构即可重新定义该算法的某些特定步骤。模版方法是基于代码复用技术

- 适用性

- 1. 一次性实现算法的公共部分，将可变的行为由子类实现
- 2. 各个子类中公共的行为应提出在父类中实现以避免代码重复，将代码中不同的部分封装成新的操作，用新的操作的模版方法代替这些不同的代码
- 3. 控制子类的扩展

- 参与者

- 1. AbstractClass 抽象类，定义一系列基本操作，这些操作可以是抽象的也可以是具体的。每个操作对应算法的一个步骤，在子类中可以重新定义或实现这些操作。
- 2. ConcreteClass 具体实现类，实现父类中声明的抽象基本操作已完成子类特定算法的步骤，也可以覆盖父类已经实现的具体基本操作。

- 样例

- Servlet: Servlet(Server Applet)是Java Servlet的简称，在每一个 Servlet 都必须实现 Servlet 接口，GenericServlet 是个通用的、不特定于任何协议的Servlet，它实现了 Servlet 接口，而 HttpServlet 继承于 GenericServlet，实现了 Servlet 接口，为 Servlet 接口提供了处理HTTP协议的通用实现，所以我们定义的 Servlet 只需要继承 HttpServlet 即可。

- 访问者模式(Visitor)

- 概述

- 表示一个作用于某对象结构的各元素的操作  
它使你可以在不改变各元素的类的前提下定义作用于这些元素的新操作。
- 访问者模式是一种将数据操作和数据结构分离的设计模式。

- 适用性

- 1.一个对象结构包含很多类对象，它们有不同的接口，而你想对这些对象实施一些依赖于其具体类的操作。
- 2.需要对一个对象结构中的对象进行很多不同的并且不相关的操作，而你想避免让这些操作“污染”这些对象的类。  
Visitor使得你可以将相关的操作集中起来定义在一个类中。  
当该对象结构被很多应用共享时，用Visitor模式让每个应用仅包含需要用到操作。
- 3.定义对象结构的类很少改变，但经常需要在此结构上定义新的操作。  
改变对象结构类需要重定义对所有访问者的接口，这可能需要很大的代价。  
如果对象结构类经常改变，那么可能还是在这些类中定义这些操作较好。

- 参与者

- 1.Visitor  
为该对象结构中ConcreteElement的每一个类声明一个Visit操作。  
该操作的名字和特征标识了发送Visit请求给该访问者的那个类。  
这使得访问者可以确定正被访问元素的具体类。  
这样访问者就可以通过该元素的特定接口直接访问它。
- 2.ConcreteVisitor  
实现每个由Visitor声明的操作。  
每个操作实现本算法的一部分，而该算法片断乃是对应于结构中对象的类。  
ConcreteVisitor为该算法提供了上下文并存储它的局部状态。  
这一状态常常在遍历该结构的过程中累积结果。

- 3.Element  
定义一个Accept操作， 它以一个访问者为参数。
- 4.ConcreteElement  
实现Accept操作， 该操作以一个访问者为参数。
- 5.ObjectStructure  
能枚举它的元素。  
可以提供一个高层的接口以允许该访问者访问它的元素。  
可以是一个复合或是一个集合， 如一个列表或一个无序集合。
- 样例(应用场景)

## JAVA基础

---

### JVM

- JVM基础知识
  - 构成图谱
    - .Class文件  
使用javap -v 指令能看到易于我们阅读的信息
      - 1. 0xCAFEBAE (4bytes 魔术)
      - 2. minor\_version (2bytes 副版本号)
      - 3. major\_version (2bytes 主版本号)  
主版本号和次版本号在class文件中各占两个字节， 副版本号占用第5、6两个字节， 而主版本号则占用第7， 8两个字节。  
jdk1.1:45 , jdk1.5:49,jdk1.6:50 (0x32) ,jdk1.7:51,jdk1.8:51  
一个JVM实例只能支持特定范围内的主版本号 (Mi 至Mj) 和0 至特定范围内 (0 至m) 的副版本号。假设一个Class 文件的格式版本号为V， 仅当 $Mi.0 \leq v \leq Mj.m$ 成立时， 这个Class 文件才可以被此Java 虚拟机支持  
JVM认为加载不了这个class文件， 会抛出我们经常见到的"  
java.lang.UnsupportedClassVersionError: Bad version number in .class  
file "Error 错误  
javap -v Math 查看.class版本
      - 4. constant\_pool\_count (2bytes 常量池计数器)  
常量池计数器默认从1开始而不是从0开始， 而是以1开始， 即count=1时,常量池有0个常量； count=2时， 常量池有1个常量； 即常量池数量+1  
原因： 在指定class文件规范的时候， 将索引#0项常量空出来是有特殊考虑的， 这样当： 某些数据在特定的情况下想表达“不引用任何一个常量池项”的意思时， 就可以将其引用的常量的索引值设置为#0来表示。
      - 5. cp\_info (常量池数据区,数量:constant\_pool\_count-1)
        - 字面量(Literal)  
哪些字面量会进入常量池中？  
结论：  
final类型的8种基本类型的值会进入常量池。  
非final类型（包括static的）的8种基本类型的值， 只有double、float、long 的值会进入常量池。  
常量池中包含的字符串类型字面量（双引号引起来的字符串值） 。
          - 文本字符
          - 被申明为final的常量值
          - 基本数据类型值
          - 其他
        - 符号引用(Symbolic Reference)



- 类和结构的完全限定名

对于某个类或接口而言，其自身、父类和继承或实现的接口的信息会被直接组装成CONSTANT\_Class\_info常量池项放置到常量池中；

类中或接口中使用到了其他的类，只有在类中实际使用到了该类时，该类信息才会在常量池中有对应的CONSTANT\_Class\_info常量池项；

类中或接口中仅仅定义某种类型的变量，JDK只会将变量的类型描述信息以UTF-8字符串组成CONSTANT\_Utf8\_info常量池项放置到常量池中，上面在类中的private Date date;JDK编译器只会将表示date的数据类型的“Ljava/util/Date”字符串放置到常量池中。

Object类，在源文件中的全限定名是 java.lang.Object 。而class文件中的全限定名是将点号替换成“/” 。Object类在class文件中的全限定名是 java/lang/Object 。源文件中一个类的名字，在class文件中是用全限定名表述的。

- 字段名称和描述符

- 各类型的描述符

对于字段的数据类型，其描述符主要有以下几种

基本数据类型 (byte、char、double、float、int、long、short、boolean)：

除long 和boolean，其他基本数据类型的描述符用对应单词的大写首字母表示。long 用J 表示，boolean 用Z 表示。

void：描述符是V。

对象类型：描述符用字符 L 加上对象的全限定名表示，如 String 类型的描述符为Ljava/lang/String 。

数组类型：每增加一个维度则在对应的字段描述符前增加一个 [，如一维数组 int[] 的描述符为 [I，二维数组 String[][] 的描述符为 [[Ljava/lang/String 。

- 方法名称和描述符

方法的描述符比较复杂，包括所有参数的类型列表和方法返回值。它的格式是这样的：

(参数1类型 参数2类型 参数3类型 ...)返回值类型

注意事项：

不管是参数的类型还是返回值类型，都是使用对应字符和对应字符串来表示的，并且参数列表使用小括号括起来，并且各个参数类型之间没有空格，参数列表和返回值类型之间也没有空格。

特殊方法的方法名：

首先要明确一下，这里的特殊方法是指的类的构造方法和类型初始化方法。构造方法就不用多说了，至于类型的初始化方法，对应到源码中就是静态初始化块。也就是说，静态初始化块，在class文件中是以一个方法表述的，这个方法同样有方法描述符和方法名，具体如下：

类的构造方法的方法名使用字符串 表示

静态初始化方法的方法名使用字符串 表示。

除了这两种特殊的方法外，其他普通方法的方法名，和源文件中的方法名相同。

总结：

方法和字段的描述符中，不包括字段名和方法名，字段描述符中只包括字段类型，方法描述符中只包括参数列表和返回值类型。

无论method()是静态方法还是实例方法，它的方法描述符都是相同的。

尽管实例方法除了传递自身定义的参数，还需要额外传递参数this，但是这一点不是由方法描述符来表达的。参数this的传递，是由Java虚拟机实现在调用实例方法所使用的指令中实现的隐式传递。

- 6. access\_flags (2bytes 访问标志)

访问标志，access\_flags 是一种掩码标志，用于表示某个类或者接口的访问权限及基础属性。

- 7. this\_class (2bytes 类索引)  
类索引, this\_class的值必须是对constant\_pool表中项目的一个有效索引值。constant\_pool表  
在这个索引处的项必须为CONSTANT\_Class\_info 类型常量, 表示这个 Class 文件所定义的类或接口。
- 8. super\_class (2bytes 父类索引)  
父类索引, 对于类来说, super\_class 的值必须为 0 或者是对constant\_pool表中项目的一个有效索引值。  
如果它的值不为 0, 那 constant\_pool 表在这个索引处的项必须为 CONSTANT\_Class\_info 类型常量, 表示这个 Class 文件所定义的类的直接父类。当前类的直接父类, 以及它所有间接父类的access\_flag 中都不能带有 ACC\_FINAL 标记。  
对于接口来说, 它的Class文件的super\_class项的值必须是对constant\_pool表中项目的一个有效索引值。constant\_pool表在这个索引处的项必须为代表 java.lang.Object 的 CONSTANT\_Class\_info 类型常量。  
如果 Class 文件的 super\_class的值为 0, 那这个Class文件只可能是定义的是 java.lang.Object类, 只有它是唯一没有父类的类。
- 9. interfaces\_count (2bytes 接口计数器)  
接口计数器, interfaces\_count的值表示当前类或接口的【直接父接口数量】
- 10. 接口数据区  
接口表, interfaces[]数组中的每个成员的值必须是一个对constant\_pool表中项目的一个有效索引值, 它的长度为 interfaces\_count。每个成员 interfaces[i] 必须为CONSTANT\_Class\_info类型常量, 其中 【 $0 \leq i < \text{interfaces\_count}$ 】。  
在interfaces[]数组中, 成员所表示的接口顺序和对应的源代码中给定的接口顺序 (从左至右) 一样, 即interfaces[0]对应的是源代码中最左边的接口。
- 11. fields\_count (2bytes 字段计数器)  
字段计数器, fields\_count的值表示当前 Class 文件 fields[]数组的成员个数。  
fields[]数组中每一项都是一个field\_info结构的数据项, 它用于表示该类或接口声明的类字段或者实例字段。
- 12. field\_info (字段信息数据区)  
字段表, fields[]数组中的每个成员都必须是一个fields\_info结构的数据项, 用于表示当前类或接口中某个字段的完整描述。  
fields[]数组描述当前类或接口声明的所有字段, 但不包括从父类或父接口继承的部分。
- 13. methods\_count (2bytes 方法计数器)  
方法计数器, methods\_count的值表示当前Class 文件 methods[]数组的成员个数。Methods[]数组中每一项都是一个 method\_info 结构的数据项。
- 14. method\_info (方法信息数据区)  
方法表, methods[] 数组中的每个成员都必须是一个 method\_info 结构的数据项, 用于表示当前类或接口中某个方法的完整描述。  
如果某个method\_info 结构的access\_flags 项既没有设置 ACC\_NATIVE 标志也没有设置ACC\_ABSTRACT 标志, 那么它所对应的方法体就应当可以被 Java 虚拟机直接从当前类加载, 而不需要引用其它类。  
method\_info结构可以表示类和接口中定义的所有方法, 包括实例方法、类方法、实例初始化方法和类或接口初始化方法。  
【methods[]数组只描述当前类或接口中声明的方法, 不包括从父类或父接口继承的方法】。

- 15. attribute\_count (2bytes 属性计数器)
 

属性计数器, attributes\_count的值表示当前 Class 文件attributes表的成员个数。attributes表中每一项都是一个attribute\_info 结构的数据项。
- 16. attribute\_info (属性信息数据区)
 

属性表, attributes 表的每个项的值必须是attribute\_info结构

在Java 7 规范里, Class文件结构中的attributes表的项包括下列定义的属性: InnerClasses、EnclosingMethod、Synthetic、Signature、SourceFile、SourceDebugExtension、Deprecated、RuntimeVisibleAnnotations、RuntimeInvisibleAnnotations以及BootstrapMethods属性。

对于支持 Class 文件格式版本号为 49.0 或更高的 Java 虚拟机实现, 必须正确识别并读取attributes表中的Signature、RuntimeVisibleAnnotations和RuntimeInvisibleAnnotations属性。对于支持Class文件格式版本号为 51.0 或更高的 Java虚拟机实现, 必须正确识别并读取 attributes表中的BootstrapMethods属性。Java 7 规范 要求任一 Java 虚拟机实现可以自动忽略 Class 文件的 attributes表中的若干 (甚至全部) 它不可识别的属性项。任何本规范未定义的属性不能影响Class文件的语义, 只能提供附加的描述信息。
- 类加载器子系统
  - 加载>Loading)
    - Bootstrap Class Loader
 

负责加载J A VA\_HOME\lib 目录中的, 或通过-Xbootclasspath参数指定路径中的, 且被虚拟机认可 (按文件名识别, 如rt.jar) 的类。由C++实现, 不是ClassLoader子类
    - Extension Class Loader
 

负责加载J A VA\_HOME\lib\ext 目录中的, 或通过java.ext.dirs系统变量指定路径中的类库

      - 加载过程
 

在加载的过程中,JVM主要做3件事情

通过一个类的全限定名来获取定义此类的二进制字节流(class文件)在程序运行过程中,当要访问一个类时,若发现这个类尚未被加载,并满足类初始化的条件时, 就根据要被初始化的这个类的全限定名找到该类的二进制字节流,开始加载过程。

将这个字节流的静态存储结构转化为方法区的运行时数据结构。

在内存中创建一个该类的java.lang. Class对象,作为方法区该类的各种数据的访问入口
      - 加载源
 

加载源:

zip包: jar , war, ear 等

其他文件生成: JSP生成的class

数据库中: 将二进制字节流存储至数据库中,然后在加载时从数据库中读取.有些中间件会这么做,用来实现代码在集群间分发

网络

运行时计算生成: 动态代理技术,用ProxyGenerator.generateProxyClass为特定接口生成形式为"\*\$Proxy"的代理类的二进制字节流
      - 类和数组加载的区别
 

数组类和非数组类的类加载是不同的, 具体情况如下:

非数组类: 是由类加载器来完成。

数组类: 数组类本身不通过类加载器创建, 它是由java虚拟机直接创

建，但数组类与类加载器有很密切的关系，因为数组类的元素类型最终要靠类加载器创建。

#### ■ 加载过程注意点

- JVM规范并未给出类在方法区中存放的数据结构
- JVM规范并没有指定Class对象存放的位置  
HotSpot将Class对象存放在方法区  
类实例存放在堆区
- 加载阶段和链接阶段是交叉的  
类加载的过程中每个步骤的开始顺序都有严格限制,但每个步骤的结束顺序没有限制。也就是说,类加载过程中,必须按照如下顺序开始:  
加载 -> 链接 -> 初始化  
但结束顺序无所谓,因此由于每个步骤处理时间的长短不一就会导致有些步骤会出现交叉

#### ■ Application Class Loader

负责加载用户路径（classpath）上的类库。

#### ■ 类加载器

- JVM的类加载是通过ClassLoader及其子类来完成的，类的层次关系和加载顺序:  
(类加载器遵循委托层次算法[Delegation Hierarchy Algorithm])

1. 加载过程中会先检查类是否被已加载，检查顺序是自底向上，从Custom ClassLoader到BootStrap

ClassLoader逐层检查，只要某个classloader已加载就视为已加载此类，保证此类在所有ClassLoader

加载一次。而加载的顺序是自顶向下，也就是由上层来逐层尝试加载此类。

检查顺序(自底向上): CustomClassLoader->ApplicationClassLoader->ExtensionClassLoader->BootStrapClassLoader

加载顺序(自上而下): BootStrap->Extension->Application->Custom

##### - 自定义类加载器

- 继承ClassLoader

- 重写findClass () 方法

- 调用defineClass () 方法

##### - 自定义类加载器作用

- JVM自带的三个加载器只能加载指定路径下的类字节码。

- 如果某个情况下，我们需要加载应用程序之外的类文件呢？比如本地D盘下的，或者去加载网络上的某个类文件，这种情况就可以使用自定义加载器了

##### - 双亲委派模型

- 避免重复加载，当父亲已经加载了该类的时候，就没有必要子ClassLoader再加载一次

- JVM在判定两个class是否相同时，不仅要判断两个类名是否相同，而且要判断是否由同一个类加载器实

例加载的

- 还要定义自己的类加载器呢？

因为Java中提供的默认ClassLoader，只加载指定目录下的jar和class，如果我们想加载其它位置的类或

jar时。

##### - 知识点

- ClassLoader 和 Class.forName加载类区别

- ClassLoader加载字节码到内存，默认不初始化类

- Class.forName加载字节码到内存后会初始化类，即调用static代码块,初始化static属

性

- 链接(Linking)

- Verify (验证)

验证阶段比较耗时,它非常重要但不一定必要(因为对程序运行期没有影响),如果所运行的代码已经被反复使用和验证过,那么可以使用 -Xverify:none 参数关闭,以缩短类加载时间。

- 验证的目的

- 字节码校验器校验字节码是否正确

- 验证的必要性

- 验证的过程

- 文件格式验证(基于二进制字节流)

本验证阶段是基于二进制字节流进行的,只有通过本阶段验证,才被允许存到方法区后面的三个验证阶段都是基于方法区的存储结构进行,不会再直接操作字节

- 元数据验证(基于方法区数据结构)

- 字节码验证(基于方法区数据结构)

- 符号引用验证(基于方法区数据结构)

- Prepare (准备)

仅仅为类变量(即static修饰的字段变量)分配内存并且设置该类变量的初始值即零值,这里不包含用final修饰的static,因为final在编译的时候就会分配了(编译器的优化),同时这里也不会为实例变量

分配初始化。类变量会分配在方法区中,而实例变量是会随着对象一起分配到Java堆中。

- 分配内存并初始化0值给所有类静态变量

- Resolve (解析)

- 所有符号内存应用被方法区(Method Area)的原始引用

- 初始化(Initialization)

- 注意点

方法是编译器自动收集类中所有类变量的赋值动作和静态语句块中的语句合并产生的,编译器收集的顺序是由语句在源文件中出现的顺序所决定的。

静态代码块只能访问到出现在静态代码块之前的变量,定义在它之后的变量,在前面的静态语句块可以赋值,但是不能访问。

实例构造器需要显式调用父类构造函数,而类的不需要调用父类的类构造函数,虚拟机会确保子类的方法执行前已经执行完毕父类的方法。因此在JVM中第一个被执行的方法的类肯定是 java.lang. Object。如果一个类/接口中没有静态代码块,也没有静态成员变量的赋值操作,那么编译器就不会为此类生成方法。

接口也需要通过方法为接口中定义的静态成员变量显示初始化。接口中不能使用静态代码块,但仍然有变量初始化的赋值操作,因此接口与类一样都会生成方法。不同的是,执行接口的方法不需要先执行父接口的方法。只有当父接口中的静态成员变量被使用到时才会执行父接口的方法。

虚拟机会保证在多线程环境中一个类的方法别正确地加锁,同步。当多条线程同时去初始化一个类时,只会有一个线程去执行该类的方法,其它线程都被阻塞等待,直到活动线程执行方法完毕。其他线程虽会被阻塞,只要有一个方法执行完,其它线程唤醒后不会再进入方法。同一个类加载器下,一个类型只会初始化一次。

- 类加载的时机

遇到 new、getstatic、putstatic 和 invokestatic 这四条指令时,如果对应的类没有初始化,则要对对应的类先进行初始化。这四个指令对应到我们Java代码中的场景分别是:

new关键字实例化对象的时候;

读取或设置一个类的静态字段(读取被final修饰,已在编译器把结果放入常量池的静态字段除外);

调用类的静态方法时。

使用 java.lang.reflect 包方法时对类进行反射调用的时候。

初始化一个类的时候发现其父类还没初始化,要先初始化其父类。

当虚拟机开始启动时,用户需要指定一个主类,虚拟机会先执行这个主类的初始化。

- 所有类静态变量赋初始值,执行静态代码块

- ...JVM内存结构

- 运行时数据区
  - 方法区
  - 堆
  - 栈
  - PC寄存器(程序计数器)
  - 本地方法栈
- 虚拟对象剖析
  - 对象创建
  - 对象访问
  - 对象分配
  - JIT 编译器
- JVM内存结构

运行时数据区

- 方法区

-XX:MaxPermSize=16m

- 为永久代的实现(PermSpace)

存放类信息,

常量池、

字符串常量池(1.7以后转到堆区)、

方法片段

- 存放各种常量池:

Class类常量池

运行时常量池

字符串常量池(<=1.6)

基本数据类型常量池

及类字节码,接口/类的全局限定名

字段名称和修饰符, 方法名称和修饰符

所有静态变量

- 方法区也会GC
- 堆区(Heap Space)

- Java堆被所有线程共享, 在Java虚拟机启动时创建。是虚拟机管理最大的一块内存。

- 存放对象实例, Java虚拟机规范的描述:

所有的对象实例以及数组都要在堆上分配

- Java堆是垃圾回收的主要区域, 主要采用分代回收算法。
- 堆分类

- <1.8

- 新生代(Young) (Eden空间[伊甸园], From Survivor空间, To Survivor空间, S0, S1)

- 老年代(Old)

- 永久代, 使用堆内存, 为方法区的实现

- >1.8

- 新生代 (Eden空间[伊甸园], From Survivor空间, To Survivor空间)、

- 老年代

- 元空间(Meta Space), 使用物理内存空间,大小受物理内存限制

存储字符串常量池的字符串引用

- Java8 中的 MetaSpace

- 为什么取消永久代

## 2. 永久代经常可能出现OOM: PermSpace

### 3. 为了融合HotSpot和JRocket,JRocket中没有永久代

- 堆内存划分
  - 堆大小= 新生代+ 老年代。堆的大小可通过参数-Xms (堆的初始容量)、-Xmx (堆的最大容量) 来指定。
  - 其中, 新生代(Young)被细分为Eden 和 两个Survivor 区域, 这两个Survivor 区域分别被命名为from 和to, 以示区分。默认的, Edem:from:to = 8:1:1。(可以通过参数-XX:SurvivorRatio 来设定。即: Eden = 8/10 的新生代空间大小, from = to = 1/10 的新生代空间大小。
  - JVM 每次只会使用Eden 和其中的一块Survivor 区域来为对象服务, 所以无论什么时候, 总是有一块Survivor 区域是空闲着的
  - 新生代实际可用的内存空间为9/10 (即90%)的新生代空间
  - jdk<=1.6时, 字符串常量池保存在方法区即永久代中
  - jdk>=1.7时, 字符串常量池保存在堆区(字符串常量池是逻辑内部表概念, 新生代,老年代是物理分层概念, 实际创建对象还是在新生代, 老年代中)
- 栈区(Java虚拟机栈 Stack Space)
  - 线程私有, 而且生命周期与线程相同, 每个Java方法在运行的时候都会创建一个栈帧(Stack Frame)
  - 栈帧
    - 随着方法创建而创建, 方法结束而销毁
    - 局部变量表
      - 基础数据类型的值
      - 对象的引用
    - 操作数栈
    - 动态连接
    - 方法返回地址
    - 附加信息
  - 栈异常
    - Java虚拟机规范中, 对该区域规定了这两种异常情况:
      - 1. 如果线程请求的栈深度大于虚拟机所允许的深度, 将会抛出StackOverflowError异常;
      - 2. 虚拟机栈可以动态拓展, 当扩展时无法申请到足够的内存, 就会抛出OutOfMemoryError异常。
- PC寄存器(程序计数器)
  - 此内存区域是唯一一个在Java的虚拟机规范中没有规定任何OutOfMemoryError异常情况的区域。
- 本地方法栈
  - 一个Native Method就是一个java调用非java代码的接口。
  - 标识符native可以与所有其它的java标识符连用, 但是abstract除外。
  - native与其它java标识符连用时, 其意义同非Native Method并无差别
  - 一个native method方法可以返回任何java类型, 包括非基本类型, 而且同样可以进行异常控制。
  - 当一个native method接收到一些非基本类型时如Object或一个整型数组时, 这个方法可以访问这些非基本型的内部, 但是这将使这个native方法依赖于你所访问的java类的实现
  - native method的存在并不会对其他类调用这些本地方法产生任何影响, 实际上调用这些方法的其他类甚至不知道它所调用的是一个本地方法。JVM将控制调用本地方法的所有细节。
  - 如果一个含有本地方法的类被继承, 子类会继承这个本地方法并且可以用java语言重写这个方法(这个似乎看起来有些奇怪), 同样的如果一个本地方法被final标识, 它被继承后不能被重写
- 执行引擎(HotSpot)
- 解释器(Interpreter)

## - JIT(JustInTime)编译器(Compiler)

热点代码才会被JIT编译器编译

### - 热点检测(HotSpot Detaction)

运行过程中会被即时编译器编译的“热点代码”有两类:

被多次调用的方法。

被多次执行的循环体

两种情况, 编译器都是以整个方法作为编译对象。这种编译方法因为编译发生在方法执行过程之中, 因此形象的称之为栈上替换 (On Stack Replacement, OSR), 即方法栈帧还在栈上, 方法就被替换了。

热点检测主要的2种方式:

基于采样的热点探测

采用这种方法的虚拟机会周期性地检查各个

线程的栈顶, 如果发现某些方法经常出现在栈顶, 那这个方法就是“热点方法”。这种探测方法的好处是实现简单高效, 还可以很容易地获取方法调用关系 (将调用堆栈展开即可), 缺点是很难精确地确认一个方法的热度, 容易因为受到线程阻塞或别的外界因素的影响而扰乱热点探测。

基于计数器的热点探测

采用这种方法的虚拟机会为每个方法 (甚至是代码

块) 建立计数器, 统计方法的执行次数, 如果执行次数超过一定的阈值, 就认为它是“热点方法”。这种统计方法实现复杂一些, 需要为每个方法建立并维护计数器, 而且不能直接获取到方法的调用关系, 但是它的统计结果相对更加精确严谨。

#HotSpot虚拟机中使用的是"基于计数器的热点探测"

为每个方法设置2个计数器(在确定虚拟机运行参数的前提下, 这两个计数器都有一个确定的阈值, 当计数器超过阈值溢出了, 就会触发JIT编译)

方法调用计数器

统计方法的调用次数

回边计数器

统计方法内循环体的执行次数

### - Server Compiler(-server)

jvm启动添加 -server 设置为Server编译器

Server Compiler则是专门面向服务器端的, 并为服务端的性能配置特别调整过的编译器, 是一个充分优化过的高级编译器Server编译器编译

用Server Compiler 来获取更好的编译质量

### - Client Compiler(-client)

jvm启动添加 -client 设置为Client编译器

Client编译器 是一个简单快速的编译器, 主要关注点在于局部优化, 而放弃许多耗时较长的全局优化手段。

Client编译器编译较Server编译器快, 用Client Compiler获取更高的编译速度

### - JIT编译器优化

- 公共子表达式的消除

- 方法内联

在使用JIT进行即时编译时, 将方法调用直接使用方法体中的代码进行替换, 这就是方法内联, 减少了方法调用过程中压栈与入栈的开销。同时为之后的一些优化手段提供条件。

- 逃逸分析(Escape Analysis: +XX:+DoEscapeAnalysis)

逃逸分析的基本行为就是分析对象动态作用域: 当一个对象在方法中被定义后, 它可能被外部方法所引用, 例如作为调用参数传递到其他地方中, 称为方法逃逸。

通过逃逸分析, Java Hotspot编译器能够分析出一个新的对象的引用的使用范围从而决定是否要将这个对象分配到堆上

逃逸分析包括:

全局变量赋值逃逸

方法返回值逃逸

实例引用发生逃逸

线程逃逸:赋值给类变量或可以在其他线程中访问的实例变量

通过JVM参数可指定是否开启逃逸分析:

(从jdk 1.7开始已经默认开始逃逸分析)

-XX:+DoEscapeAnalysis : 表示开启逃逸分析

-XX:-DoEscapeAnalysis : 表示关闭逃逸分析



- 对象的栈上内存分配

在一般情况下，对象和数组元素的内存分配是在堆内存上进行的。但是随着JIT编译器的日渐成熟，很多优化使这种分配策略并不绝对。JIT编译器就可以在编译期间根据逃逸分析的结果，来决定是否可以将对象的内存分配从堆转化为栈。

- 标量替换

标量 (Scalar) 是指一个无法再分解成更小的数据的数据。

在JIT阶段，如果经过逃逸分析，发现一个对象不会被外界访问的话，那么经过JIT优化，就会把这个对象拆解成若干个其中包含的若干个成员变量来代替。

- 同步锁消除(-XX:+EliminateLocks)

同样基于逃逸分析，当加锁的变量不会发生逃逸，是线程私有的完全没有必要加锁。在jit编译时期就可以将同步锁去掉，以减少加锁与解锁造成的资源开销。

使用参数-XX:+DoEscapeAnalysis和-XX:+EliminateLocks(锁消除必须在-server模式下)开启。

- 垃圾回收器

- 内存分配方法

1 优先在Eden分配,如果Eden空间不足虚拟机则会进行一次MinorGC

2 大对象直接接入老年代,大对象一般指的是很长的字符串或数组

3

长期存活的对象进入老年代，每个对象都有一个age，当age到达设定的年龄的时候就会进入老年代，默认是15岁

- 指针碰撞 内存地址是连续的

- Serial和ParNew收集器

- 空闲列表 内存地址不连续

- CMS收集器和Mark-Sweep收集器

- Java 本地接口

- 本地方法库

- o 面试基础问题

- String 相关

- 基本问题（字符串常量池）

- 单独使用""引号创建的字符串都是常量，编译期就已经确定存储到String Pool中。

- 使用new String("")创建的对象会存储2个对象,1个为字符串常量池对象，1个为到heap中，是运行期新创建的

- 问题升级

- 使用只包含常量的字符串连接符如"aa"+"bb"创建的也是常量，编译期就能确定已经存储到String

- Pool中

- 使用包含变量的字符串连接如"aa"+s创建的对象是运行期才创建的，存储到heap中

- 运行期调用String的intern()方法可以向String Pool中动态添加对象。

- 类型问题

- final 问题

- 值传递问题

- 字符串常量池(String Pool)

- C++维护的 StringTable

- <=jdk1.6 大小为:1009

- >=1.7 大小默认为:60013

- 可配置: -XX:StringTableSize=66666

- (<=1.6时字符串常量池在方法区即永久代,>=1.7时字符串常量池在堆区)

字符串常量池是逻辑内部表概念(实际对象创建还是在年轻代,老年代)  
堆区年轻代,老年代是物理分层概念

- `String a = new String("a")`  
会在字符串常量池保存a  
然后在堆区(Heap)创建 字符串a对象,并返回a对象地址
- `String a = "a"`  
会检查字符串常量池是否存在"a", 若不存在, 则创建  
存在则直接返回地址
- `String a = new String("a").intern()`  
同 `String a = "a"` 会从字符串常量池获取/保存字符串
- `String a = "a";`  
`String b = "b";`  
`String c = a+b;`  
`c=="ab" false`  
c只在堆区创建新String对象  
当 `c.intern()` 调用后会在字符串常量池创建对象

- 垃圾回收

- GC概念

- 对象流转过程

- 1. 向Eden区申请内存,若空间够则分配内存  
若不够则进行一次YoungGC
    - 2. YGC后再次向Eden区申请内存, 若够, 则分配内存,  
若不够则判断老年代是否够, 够则分配内存,  
若老年代也不够, 则进行1次FullGC
    - 3. FullGC后, 再次向老年代申请内存, 若够, 则分配内存  
若不够, 则抛出OOM

- 回收算法

- 引用计数器

- 对像有地方引用就+1, 失效就-1, 当计数器为0时就回收
      - Python和ActionScript3使用
      - 效率低(一直在加减计算)  
循环无法确定(循环引用无法确定)

- 根路径搜索

- 以GC roots为根搜索可达对象, 如果对象之间循环引用没有根引用则为不可达对象

- 标记-清除

- 复制算法 (新生代)

- 标记-整理 (老年代)

- 其他

- 比较

- 应用场景

- 可触性

- Stop-The-World

- 全局暂停

- 所有java代码停止, native代码可以执行, 但不能和JVM交互

- 基本上由GC清理造成  
Dump线程, 死锁检查, 堆Dump
- 回收器汇总
  - 年轻代GC和年老代并行使用
    - 年轻代
      - Serial(串行):串行回收器, 所有回收都是1个线程完成的, 回收的时候Stop-The-World, client 默认回收器
      - ParNew: 并行回收器, 多线程工作, 回收时Stop-The-World server 默认回收器
      - Parallel Scavenge: 并发回收器,  
吞吐量=程序运行时间/(JVM执行回收时间+程序运行时间)
    - 年老代
      - Serial Old,Parallel Old,CMS
      - CMS不能和Parallel Scavenge:同时使用
    - 新GC
      - G1,ZGC
- GC Roots
  - 跨代引用
  - 种类
  - 对象
- JVM 优化
  - JVM调优基本概念
    - JVM将内存区分为堆区(Heap),栈区(Stack)和方法区(Perm), 堆区存放实际的对象实例, 需要GC。  
栈区为每个方法执行开辟的栈表, 生命周期与线程相同, 存放局部变量表,操作数栈,方法返回值地址等数据
    - JVM内存模型为: 年轻代(Eden,S0,S1),老年代(OldMemory),永久代(Perm,或(java8+)元数据区),  
垃圾回收GC针对于堆区年轻带和老年代, 分为Minor GC(YGC,年轻代GC),Major GC(老年代GC,FullGC)  
对象刚创建时会存放到Eden区, 在发生YGC时, 没有引用的对象会被清除内存, 有引用的对象则被标记复制到S0(第一次YGC),S1(第二次YGC),当第在S1的对象被标记15次(默认)时, 会被复制到老年区(OldMemory)。  
对象创建时, 申请内存流程: 对象创建时先在Edon区申请空间, 若空间不足, 则进行一次MinorGC,若空间还不足则向老年代申请空间, 老年代空间不足时, 进行一次FullGC, 之后空间还不足时,抛出OOM,反之存放到老年区或Eden区。  
老年区存放大对象或大数组, 及生存时间长的对象。  
YGC,FullGC都会Stop-The-World,使程序停止事务处理, 只有GC进程允许进行垃圾回收。  
从JVM调优角度来看, 应尽量避免YGC和FullGC, 或使YGC和FullGC的时间足够短。
    - JVM调优, 主要目的是减少GC频率和FullGC次数
      - 1. Full GC 对整个堆内存进行整理, 包括Young,Old,Perm. FullGC过程比较慢, 应减少FullGC次数
      - 2. 导致FullGC的原因
        - 1. OldMemory 老年代内存不足  
调优时尽量让对象在YGC时被回收,让对象在新生代存货一段时间, 及不要创建过大的对象或数组, 避免直接在老年代创建对象

- 2. Perm Generation老年代空间不足  
增大老年代空间,避免太多静态对象,控制好新生代老年代的内存比例
  - 3. System.gc()被显示调用  
垃圾回收不要手动触发, 尽量依靠jvm自生机制  
对JVM调优过程中, 大部分工作是对FullGC的调节
- JVM调优方法和步骤
  - 1. 监控GC状态  
通过jvisualvm的visual gc插件查看GC状态, 分析JVM参数设置, 根据实际各区域内存划分和GC执行时间,判断是否需要优化
  - 2. 生成堆dump文件  
通过JMX的MBean生成当前Heap信息的hprof文件  
或使用jmap生成
  - 3.分析堆dump文件  
通过jvisualvm ,jprofiler, Mat分析dump的hprof文件
  - 4. 分析结果, 判断是否需要优化  
如果各项参数设置合理, 系统没有超时日志出现, GC频率不高, GC耗时不高, 则不需要GC优化, 如果GC时间超过1-3秒或频繁GC, 则需优化
  - 注: 如果满足下面的指标, 则一般不需要进行GC:  
Minor GC执行时间不到50ms;  
Minor GC执行不频繁, 约10秒一次;  
Full GC执行时间不到1s;  
Full GC执行频率不算频繁, 不低于10分钟1次;
  - 5. 调整GC和内存分配  
如果内存分配过大或过小, 或者采用的GC收集器比较慢, 则应该有效调整这些参数, 并找1台或多台机器进行优化对比,最后做出选择
  - 6. 不断的分析和调整  
通过不断试验调整,找出最合适的配置参数,最后应用到所有服务器
- JVM优化参数参考
  - 1. 针对JVM的设置, 一般通过-Xms,-Xmx限定最小最大值, 为防止垃圾收集器在最小、最大之间收缩产生额外时间, 通常把最大最小设置为相同值
  - 2. 年轻代与年老代将根据默认比例(1:2)分配内存,可以通过调整二者比率-XX:NewRatio来调整大小, 也可以通过-XX:NewSize,-XXMaxNewSize设置绝对大小。  
同样, 为了防止年轻代堆收缩,通常设置-XX:NewSize=-XX:MaxNewSize(或直接设置-Xmn)
  - 3. 年轻代和年老代设置多大才算合理
    - 1. 更大的年轻代必然导致更小的年老代, 大的年轻代会延长普通GC周期, 但会增加每次GC时间; 小的年老代会导致更加频繁的FullGC
    - 2. 更小的年前的必然导致更大的年老代, 小的年轻代会导致频繁YGC, 但每次GC时间很短, 大的年老代会减少FullGC的频率
    - 应根据应用程序对象生命周期的分布情况设置内存大小:  
如果应用存在大量临时对象, 应该选择更大年轻代  
如果存在相对较多的持久对象, 年老代应该适当增大。  
抉择时应根据以下2点:
      - 1. 本着FullGC尽量少的原则, 让年老代尽量缓存常用对象
      - 2. 通过观察应用一段时间, 查看峰值时年老代占多少内存, 在不影响FullGC的前提下, 根据实际情况加大年老代
  - 4. 在配置比较好的机器上(多核,大内存),可以为老年代选择并行收集算法:-XX:+UseParallelOldGC
  - 5. 线程堆栈的设置: 每个线程默认开启1M的堆栈大小, 用于存放栈臻、调用参数、局部变量表、操作数栈等, 对于大部分应用来说默认值太大, 一般256K

就够用；

理论上内存不变的情况下，减少每个线程的堆栈，可以产生更多的线程，但实际还受限于操作系统

#### ■ JVM调优年轻代/年老代大小选择总结

##### ■ 年轻代大小选择

###### ■ 响应时间优先的应用

- 年轻代尽可能的大，直到接近系统最低响应时间同时减少到达老年代的对象

###### ■ 吞吐量优先的应用

- 年轻代尽可能的大。  
因为对响应时间没有要求，垃圾回收可以并行进行，一般适合8CPU以上的应用

##### ■ 年老代大小选择

###### ■ 响应时间优先的应用

- 年老代使用并发收集器。  
但需要小心设置，一般考虑并发会话和会话持续时间等一些参数。  
如果堆设置小，可能会造成内存碎片。高回收频率及应用暂停而使用传统标记清除方式；  
如果堆设置大了，则需要较长的收集时间。  
最优化的方法，一般需要参考：  
并发垃圾收集信息  
持久代并发收集次数  
传统GC信息  
花在年轻代和年老代回收的时间比例  
减少年轻代和年老代花费的时间，一般会提高应用效率

###### ■ 吞吐量优先的应用

- 一般吞吐量优先的应用都有一个很大的年轻代和一个较小的年老代。原因是，这样可以尽可能回收掉大部分短期对象，减少中期的对象，而年老代尽存放长期存活对象。

#### ○ 参数选择

##### ■ 堆大小(新生代+老年代+永久代(<1.8))

-Xmx256M 最大堆内存大小

-Xms256M 最小堆内存大小

-Xmn128M 年轻代占堆内存大小,剩余为老年代，若配置了-Xmn则表示-XX:NewSize=-XX:MaxNewSize

- -Xmx4096m -Xms4096m 大小相等

##### ■ 栈大小

-Xss256K 每个线程的栈大小,默认1M

##### ■ 年轻代大小

-XX:NewSize 新生代初始内存大小,应小于-Xms

-XX:MaxNewSize 新生代最大内存大小,应小于-Xmx

- 官方推荐3/8，年轻代大小不要超过老年代

##### ■ 老年代和新生代比例

-XX:NewRatio=2

- 老年代和新生代比例

如:NewRatio=2 表示老年代是新生代的2倍

老年代占堆区2/3,新生代占1/3

- 新生代Eden和Survivor比例
  - XX:SurvivorRatio=1
    - 新生代里Eden和2个survivor的空间大小比例。
      - XX:SurvivorRatio=1即比例为1:1:1,即S0/S1/Eden大小相等。
      - JDK默认为8即S0:S1:Eden比例为1:1:8, Eden占新生代大小的8/10份。
- 设置新生代对象垃圾回收标记的最大年龄
  - XX:MaxTenuringThreshold=0
    - 默认为 15
- 方法区大小(1.8-),可能会进行GC
  - XX:PermSize=10M 方法区初始化内存大小
  - XX:MaxPermSize=10M 方法区最大内存带下
- 元数据区(>=1.8,MetaSpace),可能会进行GC
  - XX:MetaspaceSize=10M 元数据区初始化内存大小
  - XX:MaxMetaspaceSize=10M 元数据区最大内存大小
- JVM(HotSpot)7种垃圾回收器
 

7种垃圾收集器作用于不同的分代, 如果两个收集器之间存在连续, 就说明他们可以搭配使用。

从JDK1.3到现在, 从Serial收集器-》Parallel收集器-》CMS-》G1, 用户线程停顿时间不断缩短,

  - -XX:+UseSerialGC:设置年轻代为串行收集器
 

新生代采用复制算法,暂停所有用户线程

老年代采用标记-整理算法,暂停所有用户线程

可以和CMS,SerialOld搭配使用

    - Serial收集器是最基本、发展历史最悠久的收集器, 曾是 (JDK1.3.1之前) 虚拟机新生代收集的唯一选择。

Serial收集器是一个单线程的收集器。“单线程”的意义不仅仅是它只会使用一个CPU或一条收集器线程去完成垃圾收集工作, 更重要的是它在垃圾收集的时候, 必须暂停其他所有工作的线程, 直到它收集结束。

Serial收集器是HotSpot虚拟机运行在Client模式下的默认新生代收集器。

Serial收集器具有简单而高效, 由于没有线程交互的开销, 可以获得最高的单线程收集效率 (在单个CPU环境中) 。
  - -XX:+UseParNewGC:设置年轻代为并行收集。
 

可与CMS收集同时使用。

JDK5.0以上, JVM会根据系统配置自行设置, 所以无需再设置此值。

    - ParNew收集器是Serial收集器的多线程版本, 除了使用多条线程进行垃圾收集之外, 其余行为包括Serial收集器可用的所有控制参数、收集算法、Stop The World、对象分配规则、回收策略等都与Serial收集器一样。

ParNew收集器是许多运行在Server模式下的虚拟机首选的新生代收集器, 其中一个原因是, 除了Serial收集器之外, 目前只有ParNew收集器能与CMS收集器配合工作。

"-XX:+UseConcMarkSweepGC": 指定使用CMS后, 会默认使用ParNew作为新生代收集器。

"-XX:+UseParNewGC": 强制指定使用ParNew。
  - -XX:+UseParallelGC:设置年轻代为并行收集器
 

Parallel Scavenge JDK1.8默认年轻代收集器

多线程,复制算法

只能与SerialOld和ParallelOld搭配使用

    - Parallel Scavenge收集器是一个新生代收集器, 使用复制算法, 且是并行的多线程收集器。

Parallel Scavenge收集器关注点是达到一个可控制的吞吐量（吞吐量 = 运行用户代码时间 / （运行用户代码时间 + 垃圾收集时间）），而其他收集器关注点在尽可能的缩短垃圾收集时用户线程的停顿时间。

Parallel Scavenge收集器提供了两个参数来用于精确控制吞吐量，一是控制最大垃圾收集停顿时间的

-XX:MaxGCPauseMillis参数，二是控制吞吐量大小的

-XX:GCTimeRatio参数；

- -XX:MaxGCPauseMillis=n 参数允许的值是一个大于0的毫秒数，收集器将尽可能的保证内存垃圾回收花费的时间不超过设定的值（但是，并不是越小越好，GC停顿时间缩短是以牺牲吞吐量和新生代空间来换取的，如果设置的值太小，将会导致频繁GC，这样虽然GC停顿时间下来了，但是吞吐量也下来了）。

-XX:GCTimeRatio=n 参数的值是一个大于0且小于100的整数，也就是垃圾收集时间占总时间的比率，默认值是99，就是允许最大1%（即 $1/(1+99)$ ）的垃圾收集时间，公式为 $1/(1+n)$ 。

-XX:UseAdaptiveSizePolicy 自适应大小策略，如果这个参数打开之后，虚拟机会根据当前系统运行情况收集监控信息，动态调整新生代的比例、老年大大小等细节参数，以提供最合适的停顿时间或最大的吞吐量，这种调节方式称为GC自适应的调节策略。

-XX:ParallelGCThreads=n:设置并行收集器收集时使用的CPU数。并行收集线程数。

- Serial Old收集器(jdk1.5之前与ParallelScavenge搭配)

单线程,标记-整理算法

- Serial Old收集器是Serial收集器的老年代版本，他同样是一个单线程收集器，使用" 标记-整理" 算法。

Serial Old收集器主要用于Client模式下的虚拟机使用。

Server模式下的两大用途：一、在JDK1.5及之前的版本与Parallel Scavenge收集器搭配使用；二、作为CMS收集器的后备方案，在并发收集发生Conturrent Mode Failure时使用

- -XX:+UseParalledlOldGC:设置并行年老代收集器

多线程,标记-整理算法

Paraller Old收集器(1.6+)

- Parallel Old收集器是Parallel Scavenge收集器的老年代版本，使用多线程和“标记-整理”算法。  
在JDK1.6中才出现。

- -XX:+UseConcMarkSweepGC:设置年老代为并发收集器

多线程,标记-清除算法

- CMS收集器是一种以获取最短回收停顿时间为目标的收集器。

目前很大一部分的Java应用集中在互联网或者B/S系统的服务端上。

CMS收集器是基于“标记-清除”算法实现，它的整个运行过程可以分为：初始登记（标记一下GC Roots能直接关联到的对象，这个过程速度很快）、并发标记（进行GCRoots Tracing的过程）、重新标记（修正并发标记期间因用户线程继续运作而导致标记产生变动的那一部分对象的标记记录，速度稍慢）、并发清除（清除死亡的对象）4个步骤；其中，初始标记和重新标记仍然需要“Stop The World”。

CMS收集器运行的整个过程中，最耗费时间的并发标记和并发清楚过程收集器线程和用户线程是一起工作的，所以总体来说，CMS收集器的内存回收过程是与用户线程一起并发执行的。

优点：并发收集、低停顿。

缺点：一：CMS收集器对CPU资源非常敏感。虽然在两个并发阶段不会导致用户线程停顿，但是会因为占用了一部分线程而导致应用程序变慢，总吞吐

量下降。CMS默认启动的回收线程数是 (CPU数量+3) /4。

二：CMS收集器无法处理浮动垃圾，可能出现“Conturrent Mode Failure”失败而导致另一次Full GC产生。由于CMS并发清除阶段用户线程还在运行，伴随着程序还在产生新的垃圾，这一部分垃圾出现在标记之后，CMS无法在当次收集中处理掉它们，只能留到下次再清理，这一部分垃圾称为“浮动垃圾”。也正是由于在垃圾收集阶段用户线程还在运行，那么也就需要预留有足够的内存空间给用户线程使用，因此CMS收集器不能像其他收集器那样等待老年代填满之后再进行收集，需要预留一部分空间给并发收集时用户程序使用。可以通过“-XX:CMSInitiatingOccupancyFraction”参数设置老年代内存使用达到多少时启动收集。

三：由于CMS收集器是一个基于“标记-清除”算法的收集器，那么意味着收集结束会产生大量碎片，有时候往往还有很多内存未使用，可是没有一块连续的空间来分配一个对象，导致不得不提前触发一次Full GC。CMS收集器提供了一个“-XX:UseCMSCompactAtFullCollection”参数（默认是开启的）用于在CMS收集器顶不住要FullGC时开启内存碎片整理（内存碎片整理意味着无法并发执行不得不停顿用户线程）。参数“-XX:CMSFullGCsBeforeCompaction”来设置执行多少次不压缩的Full GC后，跟着来一次带压缩的（默认值是0，意味着每次进入Full GC时都进行碎片整理）。

■ CMS并发收集器配置:

-XX:+UseCMSCompactAtFullCollection: 使用并发收集器时，开启对老年代的压缩,默认开启。

-XX:CMSFullGCsBeforeCompaction=0: 上面配置开启的情况下，这里设置多少次Full GC后，对老年代进行压,默认为0

■ -XX:+UseG1GC :设置G1垃圾回收器(1.7+,1.9的默认垃圾回收器)

G1将新生代,老年代的物理划分取消了,取而代之的是G1将堆分成了若干区域(Region),它仍然属于分代回收器。

不过，这些区域的一部分包含新生代，新生代的垃圾收集依然采用暂停所有应用线程的方式，将存活对象拷贝到老年代或者Survivor空间。老年代也分成很多区域，G1收集器通过将对象从一个区域复制到另外一个区域，完成了清理工作。这就意味着，在正常的处理过程中，G1完成了堆的压缩（至少是部分堆的压缩），这样也就不会有cms内存碎片问题的存在了。

■ G1的堆分区(Region)中包括Eden,Servivor,Old,Humongous 4个部分

■ 在G1中，还有一种特殊的区域，叫Humongous区域。如果一个对象占用的空间超过了分区容量50%以上，G1收集器就认为这是一个巨型对象。这些巨型对象，默认直接会被分配在老年代，但是如果它是一个短期存在的巨型对象，就会对垃圾收集器造成负面影响。为了解决这个问题，G1划分了一个Humongous区，它用来专门存放巨型对象。如果一个H区装不下一个巨型对象，那么G1会寻找连续的H分区来存储。为了能找到连续的H区，有时候不得不启动Full GC。

■ G1收集器是当今收集器技术发展的最前沿成果之一；

相比其它收集器，具有如下特点：

1、并行与并发：G1能够重发利用多CPU、多核环境下的优势，使用多个CPU来缩短Stop-The-World停顿时间。

2、分代收集：与其他收集器一样，分代概念在G1中依然存在。

3、空间整合：与CMS的“标记-清理”算法不同，G1从整体来看是基于“标记-整理”来实现的收集器，从局部（两个Region之间）上来看是基于“复制”算法实现的，这两种算法都意味着G1运作期间不会产生内存空间碎片，收集后能够提供整体的可用内存。

4、可预测停顿：G1除了追求低停顿之外，还能建立可预测的停顿



时间模型，能让使用者明确指定在一个长度为M毫秒的时间片段内，消耗在垃圾收集上的时间不得超过N毫秒。

- G1收集器的运作大致可分为：

- 1、初始标记：需要停顿，耗时短；
- 2、并发标记：
- 3、最终标记：需要停顿，可并发执行；
- 4、筛选标记：

- 并行（Parallel）：指多条垃圾收集线程并行工作，但此时用户线程仍然处于等待状态。

并发（Concurrent）：指用户线程与垃圾收集线程同时执行（但不一定是并行，可能是交替执行），用户线程继续工作，而垃圾收集程序运行在另一个CPU上。

- 垃圾回收统计信息

-XX:+PrintGC

-XX:+PrintGCDetails

- -XX:+PrintGC 输出GC日志

-XX:+PrintGCDetails 输出GC的详细日志

-XX:+PrintGCTimeStamps 输出GC的时间戳（以基准时间的形式）

-XX:+PrintGCDateStamps 输出GC的时间戳（以日期的形式，如 2013-05-04T21:53:59.234+0800）

-XX:+PrintHeapAtGC 在进行GC的前后打印出堆的信息

-XX:+PrintGCApplicationStoppedTime // 输出GC造成应用暂停的时间

-Xloggc:../logs/gc.log 日志文件的输出路径

- 字符串常量池大小(数量)

Number of buckets in the interned String table

-XX:StringTableSize=66666

<=jdk1.6 默认大小为:1009

>=1.7 大小默认为:60013

- 业务场景考虑

- 高频业务处理
- 定时任务
- 服务类型

- 软硬件环境

- 机器配置
- 关联服务

- 常用JVM命令

- jvisualvm(jdk1.6+)

- 查看java进程各种状态

(VisualGC插件)

S0C 是指：Survivor0区的分配空间

S0U 是指：Survivor1区的已经使用的空间

EC是指：Eden区所使用的空间

EU是指：Eden区当前使用的空间

OC是指：老年代分配的空间

OU是指：老年代当前使用的空间

PC是指：持久待分配的空间

PU是指：持久待当前使用的空间

YGC是指：年轻代发生的次数，这里是354次

YGCT是指：年轻代发送的总时长，这里是54.272秒，因此每次年轻代发生GC，即平均每次stop-the-world的时长为 $54.272/354=0.153$ 秒。

FGC是指：年老代回收的次数，或者成为FullGC的次数。

FGCT是指：年老代发生回收的总时长。

GCT是指：包括年轻代YGC及年老代FGC的总时间。

- jconsole(jdk1.5+)
  - 查看java进程各种状态
- jps(JavaVirtualMachineProcessStatusTool)
  - 查看java进程和PID
- jstack
  - 查看java进程堆栈信息
- jmap
  - Java内存分析工具，可查看heap信息和导出heap数据  
jmap -heap PID 查看当前应用Heap情况
- iostat
- vmstat
- tcpdump
- 配置开启JMX
  - -Djava.rmi.server.hostname=192.168.32.177
  - -Dcom.sun.management.jmxremote
  - -Dcom.sun.management.jmxremote.rmi.port=1199
  - -Dcom.sun.management.jmxremote.port=1199
  - -Dcom.sun.management.jmxremote.authenticate=false
  - -Dcom.sun.management.jmxremote.ssl=false
  - -Dcom.sun.management.jmxremote.port: 这个是配置远程connection的端口号的，要确定这个端口没有被占用
  - -Dcom.sun.management.jmxremote.ssl=false
  - -Dcom.sun.management.jmxremote.authenticate=false: 这两个是固定配置，是JMX的远程服务权限的
  - -Djava.rmi.server.hostname: 这个是配置server的IP的，要使用server的IP最好在机器上先用hostname -i看一下IP是不是机器本身的IP，如果是127.0.0.1的话要改一下，否则远程的时候连不上。

- JMM 与同步

- 并发编程模型

(2个关键问题:

线程之间如何通讯及线程之间如何同步)

- 通讯
  - 共享内存模型
    - 线程之间共享程序的公共状态  
线程之间通过写-读内存中的公共状态来隐式通讯
  - 消息传递模型
    - 线程之间没有公共状态  
线程之间必须通过明确的发送消息来显示进行通讯
- 同步

指程序用于控制不同线程之间操作发生的相对顺序的机制

  - 共享内存模型
    - 同步显示进行，必须制定某方法或某段代码在线程之间互斥执行
  - 消息传递模型

- 消息的发送必须在消息的接收之前，所以同步是隐式进行的
  - Java的并发采用的是共享内存模型
    - 线程间的通讯是隐式的
- JMM(Java内存模型)
  - 基于CAS
  - JSR-133(JDK5)
  - <https://blog.csdn.net/w372426096/article/details/80898407>
  - JMM(Java Memory Model) JAVA内存模型主要目标是定义程序中的变量(指实例字段、静态字段等，不包含局部变量和方法参数，应为后2种为线程私有)在虚拟机中存储到内存与从内存读取出来的规则细节，Java内存模型规定所有变量都存储在主内存中，每条线程还有自己的工作内存，工作内存保存了该线程使用到的变量到主内存副本拷贝，线程对变量的所有操作（读取、赋值）都必须在自己的工作内存中进行而不能直接读写主内存的变量，不同线程之间无法相互直接访问对方工作内存中的变量，线程间变量值的传递均需要在主内存来完成。
  - Java JMM 内存模型是围绕并发编程中原子性、可见性、有序性三个特征来建立的
    - Java中，所有实例域、静态域、数组元素存储在堆内存中，堆内存存在线程之间共享。JMM中所指的变量即为堆内存共享变量。
  - Java线程之间通信通过JMM控制，JMM决定一个线程对共享变量的写入
    - 主内存
      - 硬件的物理内存
    - 工作内存
      - 寄存器和高速缓存
    - 工作内存中的变量是主内存中的一份拷贝
    - 线程操作,每个线程都有各自的工作内存，互相不可访问都是读写工作内存，然后通过JMM控制器刷新到主内存
      - 所有共享变量都在主内存区创建，线程读取共享变量时，现在本地内存中创建主内存共享变量副本，然后再读取副本内容；修改共享变量时，先写入本地内存，然后由JMM控制器刷新到主内存中
    - 重排序
      - 编译器重排序
        - 编译器在不改变单线程程序语义的前提下，可以重新安排语句的执行顺序
      - 处理器重排序
        - 指令重排序
        - 内存重排序
      - JAVA重排序过程: JAVA源码->编译器重排序->指令级并行重排序->内存重排序->最终执行的指令序列
      - 内存屏障指令(memory barriers)
    - as-if-serial语义
      - 不管怎么重排序，同一线程内程序的执行结果不能被改变  
编译器，runtime,处理器都必须遵守as-if-serial语义
      - 重排序对多线程的影响
        - 多线程下重排序可能会影响执行结果
      - 顺序一致性内存模型
        - 一个线程中所有操作必须按照程序的顺序执行
      - (不管是否同步)所有线程都只能看到一个单一的操作顺序执行。在顺序一致性内存模型中，每个操作都必须原子执行且立刻对所有线程可见

- o happens-before (先行发生规则)

阐述操作之间的内存可见性

Happen-Before的规则有以下几条

程序次序规则 (Program Order Rule) : 在一个线程内, 程序的执行规则跟程序的书写规则是一致的, 从上往下执行。

管程锁定规则 (Monitor Lock Rule) : 一个Unlock的操作肯定先于下一次Lock的操作。这里必须是同一个锁。同理我们可以认为在synchronized同步同一个锁的时候, 锁内先行执行的代码, 对后续同步该锁的线程来说是完全可见的。

volatile变量规则 (volatile Variable Rule) : 对同一个volatile的变量, 先行发生的写操作, 肯定早于后续发生的读操作

线程启动规则 (Thread Start Rule) : Thread对象的start()方法先行发生于此线程的没一个动作

线程中止规则 (Thread Termination Rule) : Thread对象的中止检测 (如: Thread.join(), Thread.isAlive()等) 操作, 必行晚于线程中所有操作

线程中断规则 (Thread Interruption Rule) : 对线程的interruption () 调用, 先于被调用的线程检测中断事件(Thread.interrupted())的发生

对象中止规则 (Finalizer Rule) : 一个对象的初始化方法先于一个方法执行Finalizer()方法

传递性 (Transitivity) : 如果操作A先于操作B、操作B先于操作C,则操作A先于操作C

- 1. 程序顺序规则

- 一个线程中的每个操作, 都hanpens-before于该线程中任意的后续操作

- 2. 监视器锁(管程锁定)规则

- 对一个监视器的解锁, hanpens-before于对同一个监视器的加锁

- 3. volatile变量规则

- 对一个volatile域的写, hanpens-before于后续任意对该volatile域的读

- 4. 线程启动规则

- 假定线程A在执行过程中, 通过执行ThreadB.start()来启动线程B, 那么线程A对共享变量的修改在接下来线程B开始执行后确保对线程B可见。

- start 先于 run方法

- 5. 线程终止规则

- 假定线程A在执行的过程中, 通过制定ThreadB.join()等待线程B终止, 那么线程B在终止之前对共享变量的修改在线程A等待返回后可见。

- 6. 线程中断规则

- interrupt操作,必须发生在捕获该动作之前

- 7. 传递规则

- 如果A hanpens-before 于 B,B hanpens-before于C, 则A一定hanpens-before于C

- 8. 对象终结规则

- 初始化必须发生在finalize之前

- 同步

- 可保证指令不进行重排序

- synchriized

- 保证可见性和原子性, 互斥

- volatile

- 只处理内存可见性, 非原子性;

- 保证变量在线程工作内存和主内存之间的一致

- 原理: 内存锁, cpu缓存中存储内存地址, 修改时通知总线修改, 实现同步

- 线程有自己独立的快速缓存区从主内存中读取, 修改时会让他其他快速缓存失效, 然后从主内存中读取

- 或

- 写volatile对象时, JMM会把该线程对应的本地内存中的共享变量值刷新到主内存中

读取volatile对象时，JMM会把该线程对应的本地内存置为失效，然后从主内存中读取变量

- 可见性

对一个volatile的读总能看到，任意线程对该变量最后的写

■ 原子性

对单个volatile的读写具有原子性

对volatile的符合操作不具有原子性（i++之类）

- volatile内存语义实现

- JMM通过限制编译器，处理器重排序来实现volatile写/读内存语义

1. volatile 变量写前任何操作不能重排序

2. volatile读后任何操作不能重排序

3. 第一个操作是写，第二个操作是读不能重排序

○ lock

■ 锁释放和获取的内存语义

■ 当线程释放锁时，JMM会把该线程对应的本地内存中的共享变量刷新到主内存。

■ final

■ 保证可见性

• 0 Copy

○ mmap

○ sendfile

• JMX

## 类加载

• 类代码块

○ 普通类 代码块编译后 转换为 无参构造方法体

（若已存在显示的无参构造方法，则将代码块插入到无参构造方法最开始位置）

○ 静态类代码块

• 普通类加载

○ Class.forName(className)

加载类字节码到内存，并初始化类，为类静态变量赋初始值，并执行静态代码块

○ ClassLoader.loadClass(className)

ClassLoader加载类，只是将字节码加载到内存，并不会初始化类（即为类静态变量赋初始值，调用类的static代码块）

Thread.currentThread().getContextClassLoader().loadClass(className);

○ 调用类中常量时不会初始化类

○ //调用类中静态属性时，会为类静态变量赋初始值，初始化父类及子类，及执行父类static代码块及子类static代码块

○ new 对象时，会加载并初始化父类及子类对象，执行父类static代码块和父类无参构造方法，及子类static代码块和子类构造方法和为类静态变量赋初始值，

• 内部类

```
class Test {  
    void test() {  
        class X { // "Test$1X"  
        }  
    }  
}  
  
void test2() {  
    class X { // "Test$2X"}
```

```

    }
    class Y { // "Test$1Y"
    }
    System.out.println(new X()){ } // "Test$1"
    System.out.println(new X()){ } // "Test$2"
    System.out.println(new Y()){ } // "Test$3"
}
// Test$Test2$1Foo
static class Test2 {
    void test() {
        class Foo {
            void bar(){ }
        }
    }
}
}
}

```

- 局部类LocalClass

类名为 所在类\$1类名

如: ChildClass\$1LocalClass

若相同的LocalClass名出现多次, 则类名依次为

ChildClass\$1LocalClass

ChildClass\$2LocalClass

序号从1开始

- 只能定义class,不能为enum,interface

- 成员类MemberClass

类名为 所在类\$类名

如: ChildClass\$InnerClass.class

- enum,interface编译后为static

- public static enum MemberEnumReflexClassParent { }
  - public static abstract interface MemberInterfaceReflexClassParent{ }

- class 编译后依然为class

- 成员类调用基类的成员变量时, 将该成员变量作为成员类的成员变量, 变量名为this\$0, 并被final修饰
- 并生成参数为所在类对象的构造方法

- 匿名内部类AnonymousClass

不能显示的继承和实现其他类或接口

类名为所在类\$1.class

如: ChildClass\$1.class

- new Class(){ };

- extents 基类
  - 无interface类

- new Interface(){ };

- extents Object
  - 实现 基类

- 匿名内部类被final修饰, 即 final class

- 匿名类/局部类调用所在类的成员变量和局部变量时,需要将传入的变量转换为final,防止成员变量和局部变量在匿名类中被修改, 变量在所在类运行结束后, 可能依然在匿名类/局部类中被使用, 导致内存泄漏。

匿名类/局部类调用所在类的成员变量和局部变量时,会将变量生成为自己的成员变量, 并生成

该成员变量的构造方法，在调用位置使用改构造初始化对象。，并被final修饰。  
使用所在类成员变量时，将所在类加入到成员变量中并创建创建基于所在类的构造方法，即 new Clasz(this,变量)，变量名为this\$0;  
使用所在类局部变量时，会将局部变量加入到成员变量中，并创建创建基于所在类的构造方法，，即 new Clasz(final 变量)：局部变量的变量名为val\$变量名，并被final修饰。  
创建匿名类/局部类实例，会转换为如下方式(转换为调用构造方法)：  
ChildClass.2LocalClass lc = new ChildClass.2LocalClass(this, i);  
ChildClass.2LocalClass lc2 = new ChildClass.2LocalClass(this, i);

## 注解

- 注解
  - interface Annotation
    - @interface 隐式继承(extends) java.lang.annotation.Annotation
    - 1个RetentionPolicy
    - 1~n个ElementType
  - 成员变量：注解的成员变量都是方法，可以有默认值。  
返回值为8大基本数据类型，String,Class,enum,Annotation及以上类型数组
    - value
      - 如果Annotation里有一个名为“value”的成员变量，使用该Annotation时，可以直接使用XXX(val)形式为value成员变量赋值，无须使用name=val形式。
  - implement注解
    - 注解隐式继承了Annotation接口，所以可以实现一个注解
    - 可动态为某个标识添加注解的实现
  - 主要用途
    - 生成文档
    - 编译检查
    - 编译时动态处理
    - 运行时动态处理，如反射
- 内置注解
  - @Override
    - 检查该方法是否重载方法。如果发现其父类，或者是引用的接口中并没有该方法时，会报编译错误。
  - @Deprecated
    - 标记过时方法。如果使用该方法，会报编译警告
  - @SuppressWarnings
    - 指示编译器忽略注解中声明的警告
  - 元注解
    - @Documented
      - 标记注解是否会包含在文档中
    - @Retention
      - 标记其它注解类型保留的生命周期
        - RetentionPolicy.SOURCE
          - 标记Annotation只保留在源代码中，编译器编译时，直接丢弃这种Annotation
        - RetentionPolicy.CLASS

- 标记编译器把Annotation记录在class文件中。当运行Java程序时，JVM中不再保留该Annotation
  - RetentionPolicy.RUNTIME
    - 编译器把Annotation记录在class文件中。当运行Java程序时，JVM会保留该Annotation，程序可以通过反射获取该Annotation的信息。
- @Target
  - 标记注解作用位置
    - java.lang.annotation.ElementType
      - ElementType.TYPE
        - 修饰类，接口或枚举类型
      - ElementType.FIELD
        - 修饰成员变量
      - ElementType.METHOD
        - 修饰方法
      - ElementType.PARAMETER
        - 修饰方法参数
      - ElementType.CONSTRUCTOR
        - 修饰构造器
      - ElementType.LOCAL\_VARIABLE
        - 修饰局部变量
      - ElementType.ANNOTATION\_TYPE
        - 修饰注解
      - ElementType.PACKAGE
        - 修饰包
- @Inherited
  - 标注子类自动继承父类注解  
(子类extends父类会继承注解  
子类实现接口不继承注解  
子接口extends父接口不继承注解)
- JDK7+
  - @SafeVarargs
    - JDK7, 忽略任何使用参数为泛型变量的方法或构造函数调用产生的警告
  - @FunctionalInterface
    - JDK8, 标记一个匿名函数或函数式接口
  - @Repeatable
    - JDK8, 标记某注解可以在同一声明上使用多次

## 泛型(generics)

- 泛型的本质是参数化类型
- 泛型提供了编译时类型安全监测机制
- 泛型作用于编译阶段
- Java库中 E表示集合的元素类型，K 和 V分别表示表的关键字与值的类型
- T（需要时还可以用临近的字母 U 和 S）表示“任意类型”
- 类型变量放在修饰符（这里是 public static）的后面，返回类型的前面。



- 解决容器的类型安全，使用泛型后，能让编译器在编译的时候借助传入的类型参数检查对容器的插入
- 主要为Collection类型
- 泛型的使用
  - 泛型接口(generics interface)
    - `public interface Generator<T extend Object> {...}`
  - 泛型类(generics class)
    - `public class Pair<T extend Object,U extend Object> { ... }`
  - 泛型方法(generics method)
    - `public static <T extend Object> T getMiddle(T... a) {.. }`
- 协变：子类能向父类转换 `Animal a1=new Cat();`
- 逆变：父类能向子类转换 `Cat a2=(Cat)a1;`
- 不变：两者均不能转变
- 泛型与向上转型的概念  
泛型上下边界

- `extends,super`关键字与通配符?

- `List < E extends Fruit>`
  - 作用于类，接口，方法，变量泛型定义
- `List < ? >`  
通配符
  - 具体类型为任意类型
  - 作用于方法，变量泛型定义
  - 不能向通配符泛型写入  
可以读取，但必须使用Object接
- `List < ? extends Fruit>`  
上边界通配符
  - 具体类型必须是Fruit的子类类型
  - 作用于方法，变量泛型定义
  - 上界<? extends T>不能往里存，只能往外取
  - `Plate<? extends Fruit> fruitPlate = new Plate(new Apple());`
- `List < ? supers Fruit>`  
下边界通配符
  - 泛型具体类型必须是Fruit的父类类型
  - 作用于方法，变量泛型定义
  - 下界<? super T>可以存，但往外取只能放在Object对象里
  - `Plate<? super Fruit> fruitPlate = new Plate(new Fruit());`
- //普通通配符不能存，只能用object接收读取  
`Collection<?> c = new ArrayList();`  
`Object a = c.iterator().next();`  
`//c.add(new Object());`  
`//下界通配符super可以存，但取只能用Object接`  
`Collection<? super Integer> css = new ArrayList();`  
`css.add(1);`  
`css.add(2);`  
`Object aa = css.iterator().next();`  
`//上届界通配符extend 可以取，但不能存`  
`Collection<? extends Number> cssn = new ArrayList();`

```
//cssn.add(new Integer(1));  
Number aaa = cssn.iterator().next();
```

- PECS (Producer Extends Consumer Super) 原则, 已经很好理解了:  
如果参数化类型表示一个生产者, 就使用<? extends T>; 如果它表示一个消费者, 就使用<? super T>。  
频繁往外读取内容的, 适合用上界Extends。  
经常往里插入的, 适合用下界Super。
- 泛型重载了extends, super关键字来解决通用泛型的表示
- extend还被用来指定擦除到的具体类型, 比如, 表示在运行时将E替换为Fruit,注意E表示的是一个具体的类型, 但是这里的extend和通配符连续使用<? extend Fruit>这里通配符?表示一个通用类型, 它所表示的泛型在编译的时候, 被指定的具体的类型必须是Fruit的子类。比如List<? extend Fruit> list= new ArrayList, ArrayList<>中指定的类型必须是Apple,Orange等
- 静态方法与泛型
  - 静态方法无法访问类上定义的泛型; 如果静态方法操作的引用数据类型不确定的时候, 必须将泛型定义在方法上。  
即: 如果静态方法要使用泛型的话, 必须将静态方法也定义成泛型方法
- 泛型数组
  - 在java中是“不能创建一个确切的泛型类型的数组”
    - 也就是说下面的这个例子是不可以的:  
List[] ls = new ArrayList[10];  
而使用通配符创建泛型数组是可以的, 如下面这个例子:  
List[] ls = new ArrayList[10];  
这样也是可以的:  
List[] ls = new ArrayList[10];

## 反射

- JAVA反射机制:  
在运行时, 对于任意一个类, 能够知道类的所有属性和方法;  
对于任意对象, 都能够调用任意方法和属性;  
这种动态获取信息以及动态调用对象方法的功能称为java语言的反射机制。  
动态创建类实例。
  - 主要类对象
    - java.lang.Class
    - java.lang.reflect.Field
    - java.lang.reflect.Method
    - java.lang.reflect.Constructor
    - java.lang.Package
  - 2大类方法
    - getXXXX
      - 获取所有内容, 含父类继承
    - getDeclaredXXX
      - 获取直接作用的内容, 不含父类继承
    - getGenericXXX
      - 获取带泛型的对象
  - Package
    - claz.getClassPackage()

- 获取类所在的包
- Annotation[] as = pkg.getAnnotations();
  - 获取package所有注解，包含继承注解
- Annotation a = pkg.getAnnotation(Annotation.class)
  - 根据注解类型获取注解对象
- (JDK8+)Annotation[] at = pkg.getAnnotationsByType(Annotation.class);
  - 根据注解类型获取注解对象数组，若为可重复注解时返回所有类型注解对象
- pkg.getDeclaredAnnotation...;
  - 获取直接作用于包的注解，不含继承的注解
- Class
  - 获取Class对象的3种方式
    - .getClass()
    - 类.class
    - Class.forName(String)
  - 创建类实例
    - 无参构造方法创建
      - Claz clas = Class.forName("className");  
clas.newInstance();
    - 有参构造方法
      - Constructor[] cs = (Constructor[]) claz.getConstructors();  
for (Constructor c:cs) {  
    if (c.getParameterCount() > 0) {  
        rc = (ReflexClass) c.newInstance("");  
    }  
}
  - Class<?> ec = claz.getEnclosingClass()
    - 返回该类是在那个类中定义的，比如直接定义的内部类或匿名内部类
  - Constructor<?> encs = claz.getEnclosingConstructor();
    - 该类是在哪个构造函数中定义的，比如构造方法中定义的匿名内部类
  - Method em = claz.getEnclosingMethod();
    - 该类是在哪个方法中定义的，比如方法中定义的匿名内部类
  - Class asSubclass(Class<T> clazz) 把传递的类的对象转换成代表其子类的对象
  - T cast(Object obj) 把对象转换成代表类或是接口的对象
  - getClassLoader() 获得类的加载器
  - getClass() 返回一个数组，数组中包含该类中所有公共类和接口类的对象
  - getDeclaredClasses() 返回一个数组，数组中包含该类中所有类和接口类的对象
  - forName(String className) 根据类名返回类的对象
  - getName() 获得类的完整路径名字
  - newInstance() 创建类的实例
  - getPackage() 获得类的包
  - getSimpleName() 获得类的名字
  - getSuperclass() 获得当前类继承的父类的名字
  - getInterfaces() 获得当前类实现的类或是接口

- isAnnotation() 如果是注解类型则返回true
- isAnnotationPresent(Class<? extends Annotation> annotationClass) 如果是指定类型注解类型则返回true
- isAnonymousClass() 如果是匿名类则返回true
- isArray() 如果是一个数组类则返回true
- isEnum() 如果是枚举类则返回true
- isInstance(Object obj) 如果obj是该类的实例则返回true
- isInterface() 如果是接口类则返回true
- isLocalClass() 如果是局部类则返回true
  - isMemberClass() 如果是内部类则返回true
    - .getClass().getGenericInfo() 获取泛型信息
    - .getClass().getTypeParameters()
  - Constructor
    - Constructor<?>[] cs1 = clazz.getConstructors();
      - 获取类的所有public构造方法
    - Constructor<?>[] dcs = clazz.getDeclaredConstructors();
      - 获取类定义的所有构造方法
    - Constructor<?> ics = clazz.getConstructor(String.class);
      - 获取指定参数类型的构造方法
    - getConstructor(Class...<?> parameterTypes) 获得该类中与参数类型匹配的公有构造方法
    - getDeclaredConstructor(Class...<?> parameterTypes) 获得该类中与参数类型匹配的构造方法
    - Constructor方法
      - newInstance(Object... initargs) 根据传递的参数创建类的对象
  - Field
    - getField(String name) 获得某个公有的属性对象
    - getFields() 获得所有公有的属性对象
    - getDeclaredField(String name) 获得某个属性对象
    - getDeclaredFields() 获得所有属性对象
    - Field方法
      - equals(Object obj) 属性与obj相等则返回true
      - get(Object obj) 获得obj中对应的属性值
      - set(Object obj, Object value) 设置obj中对应属性值
  - Method
    - getMethod(String name, Class...<?> parameterTypes) 获得该类某个公有的方法
    - getMethods() 获得该类所有公有的方法
    - getDeclaredMethod(String name, Class...<?> parameterTypes) 获得该类某个方法
    - getDeclaredMethods() 获得该类所有方法
  - Method方法
    - invoke(Object obj, Object... args) 传递object对象及参数调用该对象对应的方法
      - Type[] empty = em.getGenericParameterTypes(); 获取参数列表的参数泛型列表  
实际为ParameterizedType类型

- Type[] empt= em.getGenericParameterTypes();
      - ((ParameterizedType)gcm1t[0]).getActualTypeArguments()[0]
    - //获取Class[]的参数列表
      - Class[] cs = gcm1.getParameterTypes();
        - //获取定义的泛型
          - TypeVariable[] tvs = gcm1.getTypeParameters();
    - Annotation
      - getAnnotation(Class<T> annotationClass) 返回该类中与参数类型匹配的公有注解对象
      - getAnnotations() 返回该类所有的公有注解对象
      - getDeclaredAnnotation(Class<T> annotationClass) 返回该类中与参数类型匹配的所有注解对象
    - getDeclaredAnnotations() 返回该类所有的注解对象
  - 枚举Enum
    - enum关键字定义的枚举
      - 隐式继承自extends java.lang.Enum
      - 泛型为当前定义对象
    - 枚举不能定义readObject,writeObject,readObjectNoData,writeReplace,readResolve 等方法处理反序列化
      - java.io.ObjectStreamClass中定义序列化时单独对enum做了处理
      - enum 类型不支持 public 和 protected 修饰符的构造方法，因此构造函数一定要是 private 或 friendly 的。也正因为如此，所以枚举对象是无法在程序中通过直接调用其构造方法来初始化的
        - 同时在Construct.newInstance中判断了枚举不能被实例化处理：
          - Cannot reflectively create enum objects
      - 定义 enum 类型时候，如果是简单类型，那么最后一个枚举值后不用跟任何一个符号；但如果有定制方法，那么最后一个枚举值与后面代码要用分号';'隔开，不能用逗号或空格。
      - 枚举中的值在编译后会成为 final static 成员变量

## Java中的引用

影响JVM对象内存回收操作

<https://droidyue.com/blog/2014/10/12/understanding-weakreference-in-java/>

- 强引用
  - 普通编码中的对象赋值就是强引用
  - 存在强引的对象用不会被GC回收
- 软引用(java.lang.ref.SoftReference)
  - 存在软引用的对象在(不存在其他强引用)内存不足时才会被回收
  - 可用get()获取被引用对象,使用引用对象需要校验null
- 弱引用(java.lang.ref.WeakReference)
  - 存在弱引用的对象在(不存在其他强引用)GC时，会被回收
  - 可用get()获取被引用对象,使用引用对象需要校验null
- 虚引用(java.lang.ref.PhantomReference)
  - 不能使用get()获取被引用对象
    - 虚引用使用场景主要由两个。
      - 一，它允许你知道具体何时其引用的对象从内存中移除。而实际上这是Java中唯一的方式。(在对象被销毁时，把该引用加入到引用队列中)

第二点，虚引用可以避免很多析构时的问题。finalize方法可以通过创建强引用指向快被销毁的对象来让这些对象重新复活。然而，一个重写了finalize方法的对象如果想要被回收掉，需要经历两个单独的垃圾收集周期。

## 动态代理

- 静态代理
- java.lang.reflect.Proxy(基于JDK的动态代理)
- 基于CGLIB的动态代理

## 字节码操作

- BCEL(Byte Code Engineering Library)
- ASM Java字节码操作和分析框架，能够操作现有类及动态生成新类  
<https://asm.ow2.io/>
- cglib(CodeGenerationLibrary)  
基于ASM，实现类动态代理框架
  - 与Java原生Proxy动态代理的区别:  
Java原生Proxy代理只支持代理接口，无法代理没有接口的普通类  
cglib能够代理普通类和接口，是增功能更强
  - Enhancer  
Enhancer可能是CGLIB中最常用的一个类，和Java1.3动态代理中引入的Proxy类差不多(如果对Proxy不懂，可以参考这里)。和Proxy不同的是，Enhancer既能够代理普通的class，也能够代理接口。Enhancer创建一个被代理对象的子类并且拦截所有的方法调用（包括从Object中继承的toString和hashCode方法）。Enhancer不能够拦截final方法，例如Object.getClass()方法，这是由于Java final方法语义决定的。基于同样的道理，Enhancer也不能对final类进行代理操作。这也是Hibernate为什么不能持久化final class的原因。
    - ```
public class SampleClass {  
    public String test(String input){  
        return "hello world";  
    }  
}  
  
@Test  
public void testFixedValue(){  
    Enhancer enhancer = new Enhancer();  
    enhancer.setSuperclass(SampleClass.class);  
    enhancer.setCallback(new FixedValue() {  
        @Override  
        public Object loadObject() throws Exception {  
            return "Hello cglib";  
        }  
    });  
    SampleClass proxy = (SampleClass) enhancer.create();  
    System.out.println(proxy.test(null)); //拦截test，输出Hello cglib  
    System.out.println(proxy.toString());  
    System.out.println(proxy.getClass());  
    System.out.println(proxy.hashCode());  
}
```

- ImmutableBean  
通过名字就可以知道，不可变的Bean。ImmutableBean允许创建一个原来对象的包装类，这个包装类是不可变的，任何改变底层对象的包装类操作都会抛出IllegalStateException。但是我们可以直接操作底层对象来改变包装类对象。这有点类似于Guava中的不可变视图
- Bean generator  
cglib提供的一个操作bean的工具，使用它能够在运行时动态的创建一个bean。
- Bean Copier  
cglib提供的能够从一个bean复制到另一个bean中，而且它还提供了一个转换器，用来在转换的时候对bean的属性进行操作
- Javassist  
基于ASM
  - CClass,CMethod,CField等

## 并发编程

- Thread
  - Thread
    - extends Thread  
@override void run(){}
      - implements Runnable  
@override void run(){}
        - 线程状态
          - NEW  
新创建了一个线程对象。
          - RUNNABLE  
线程对象创建后，其他线程(比如main线程)调用了该对象的start()方法。该状态的线程位于可运行线程池中，等待被线程调度选中，获取cpu的使用权。
          - RUNNING  
可运行状态(runnable)的线程获得了cpu时间片 (timeslice)，执行程序代码
          - BLOCKED  
阻塞状态是指线程因为某种原因放弃了cpu使用权，也即让出了cpu timeslice，暂时停止运行。直到线程进入可运行(runnable)状态，才有机会再次获得cpu timeslice 转到运行(running)状态。阻塞的情况分三种：
            - (一). 等待阻塞：运行(running)的线程执行o.wait()方法，JVM会把该线程放入等待队列(waiting queue)中。
            - (二). 同步阻塞：运行(running)的线程在获取对象的同步锁时，若该同步锁被别的线程占用，则JVM会把该线程放入锁池(lock pool)中。
            - (三). 其他阻塞：运行(running)的线程执行Thread.sleep(long ms)或t.join()方法，或者发出了I/O请求时，JVM会把该线程置为阻塞状态。当sleep()状态超时、join()等待线程终止或者超时、或者I/O处理完毕时，线程重新转入可运行(runnable)状态。
        - DEAD  
线程run()、main()方法执行结束，或者因异常退出了run()方法，则该线程结束生命周期。死亡的线程不可再次复生。
      - 暂停线程
        - suspend  
暂停

- 问题：使用suspend暂停线程时，会导致公共同步对象被暂停线程独占，导致无法被其他线程访问
  - resume  
恢复
- 终止线程
  - thread.interrupt()
    - 中断线程
    - thread.interrupted
      - 测试当前线程是否已经中断，清除中断标记
    - thread.isInterrupted()
      - 测试当前线程是否已经中断，不清除中断标记
  - 设置对出标志位
  - stop
    - 使用stop会时线程终止时，线程内的数据结果不一致，  
如：obj.a,b="a","aa";  
thread.run(){  
a = "b";  
Thread.sleep(1000L);  
b="bb";  
}  
在b="bb"未执行前stop线程，最终b="bb"未被赋值，  
结果为: "a" "aa"
  - 异常
- 线程优先级
  - 最小为1，最大为10  
默认为5
  - 优先级有继承性，A线程中启动B,B的优先级同A
  - 优先级具有随机性，相同处理逻辑，不保证优先级高的线程先执行完
- 线程组:
  - 不指定线程组时，使用Parent线程组
- Thread.join
  - 等待线程结束
- Thread.sleep
  - 释放CPU资源，不释放锁
- synchronized
  - 可以修饰方法，代码块  
保证内存可见性及线程安全  
非公平锁  
可重入锁  
同步不具备继承性  
synchronized方法弊端：占用整个方法资源，导致其他调用方阻塞等待该锁，可用synchronized代码块  
多线程访问下会阻塞
  - synchronized可重入
    - 子类调用父类的synchronized方法，  
若父类方法中以this作为锁对象，则该this为子类实例
    - synchronized锁是作用于线程的



- 同一线程在调用自己类中其他synchronized方法/块或调用父类的synchronized方法/块都不会阻碍该线程的执行，就是说同一线程对同一个对象锁是可重入的，而且同一个线程可以获取同一把锁多次，也就是可以多次重入。因为java线程是基于“每线程（per-thread）”，而不是基于“每调用（per-invocation）”的（java中线程获得对象锁的操作是以每线程为粒度的，per-invocation互斥体获得对象锁的操作是以每调用作为粒度的）
  - 重入锁是怎么实现可重入性的：其实现方法是每个锁关联一个线程持有者和计数器，当计数器为0时表示该锁没有被任何线程持有，那么任何线程都可能获得该锁而调用相应的方法；当某一线程请求成功后，JVM会记下锁的持有线程，并且将计数器置为1；此时其它线程请求该锁，则必须等待；而该持有锁的线程如果再次请求这个锁，就可以再次拿到这个锁，同时计数器会递增；当线程退出同步代码块时，计数器会递减，如果计数器为0，则释放该锁
- synchronized(Class)
  - static synchronized{
    - 所有使用当前Class的线程持有同一类锁
- synchronized(Object)
  - synchronized(this)
    - 所有使用当前对象的线程持有同一个对象锁
- String 常量池特性
  - String常量池对同一个字符串会使用相同的内存地址，所以用synchronized(String)时为同一锁
- volatile
  - 只能修饰成员变量
    - 保证内存可见性, 主要解决多线程环境下访问资源的同步性
    - 直接操作为线程安全
    - 复合操作(i++,i=i+1)为线程不安全,因为+/-操作不是线程安全
    - 多线程访问下不会阻塞
- ThreadLocal
  - 提供了线程存储数据的能力
  - threadLocal.get()
    - threadLocal.set()
  - ThreadLocal.ThreadLocalMap
    - ThreadLocalMap为ThreadLocal的静态内部类
    - ThreadLocal 为每个Thread创建一个ThreadLocalMap;
    - ThreadLocalMap中维护了Entry<ThreadLocal,value>[] table数组(数组默认大小:16,空间不够时扩充为原空间的2倍);
    - ThreadLocal.get()时，先获取当前线程中的ThreadLocalMap(),若不存在，则创建ThreadLocalMap,然后保存当前ThreadLocal对象的null值。
    - Entry extends WeakReference
      - Entry继承自WeakReference（弱引用，生命周期只能存活到下次GC前），但只有Key是弱引用类型的，Value并非弱引用。（问题马上就来了）
    - Entry[] 索引确定:
      - firstKey.threadLocalHashCode & (INITIAL\_CAPACITY - 1)
      - 取自AtomicInteger ThreadLocal.nextHashCode 自增
      - 自增增量为: 0x61c88647
      - 0x61c88647是斐波那契散列乘数,它的优点是通过它散列(hash)出来的结果分布会比较均匀，可以很大程度上避免hash冲突
  - InheritableThreadLocal extends
    - java.lang.ThreadLocal

- 重写了ThreadLocal3个方法
  - 将只影响Thread类中的inheritableThreadLocals变量，不操作threadLocals变量
  - T childVal(T parentValue)
  - ThreadLocalMap getMap(t)
  - createMap(Thread t,firstValue)
- 创建线程默认会继承父类的inheritableThreadLocals中的值
- ThreadLocal特性
  - ThreadLocal和Synchronized都是为了解决多线程中相同变量的访问冲突问题，不同的点是Synchronized是通过线程等待，牺牲时间来解决访问冲突
  - ThreadLocal是通过每个线程单独一份存储空间，牺牲空间来解决冲突，并且相比于Synchronized，ThreadLocal具有线程隔离的效果，只有在线程内才能获取到对应的值，线程外则不能访问到想要的值。
  - 正因为ThreadLocal的线程隔离特性，使他的应用场景相对来说更为特殊一些。在android中Looper、ActivityThread以及AMS中都用到了ThreadLocal。当某些数据是以线程为作用域并且不同线程具有不同的数据副本的时候，就可以考虑采用ThreadLocal。
- 问题:
  - ThreadLocalMap Entry为什么使用弱引用
    - 为了处理非常大和生命周期非常长的线程，哈希表使用弱引用作为 key。当ThreadLocal对象被置为null时，如果Entry中key为强引用时，在线程存活周期内，ThreadLocal在堆中的内存不会被回收，因为Entry中强引用了ThreadLocal对象，导致内存泄漏。
  - 内存泄漏
    - 当ThreadLocal值使用后没有remove(),由于ThreadLocalMap的key是弱引用，而Value是强引用。这就导致了一个问题，ThreadLocal在没有外部对象强引用时，发生GC时弱引用Key会被回收，而Value不会回收。当线程没有结束，但是ThreadLocal已经被回收，则可能导致线程中存在ThreadLocalMap<null, Object>的键值对，造成内存泄露。（ThreadLocal被回收，ThreadLocal关联的线程共享变量还存在）。注意: 在调用threadLocal.get(),set(),remove()后，ThreadLocal有机会把部分key为null的值清除掉所以，使用使用ThreadLocal后及时清理ThreadLocal的值。
      - 为了防止此类情况的出现，我们有两种手段。
      - 1、使用完线程共享变量后，显示调用ThreadLocalMap.remove方法清除线程共享变量；既然Key是弱引用，那么我们要做的事，就是在调用ThreadLocal的get()、set()方法时完成后再调用remove方法，将Entry节点和Map的引用关系移除，这样整个Entry对象在GC Roots分析后就变成不可达了，下次GC的时候就可以被回收。
      - 2、JDK建议ThreadLocal定义为private static，这样ThreadLocal的弱引用问题则不存在了。
    - 内存溢出针对的是ThreadLocalMap value
- java.util.concurrent.ThreadFactory
  - 接口,只有1个 Thread newThread(Runnable r) 方法
- java.util.concurrent.ThreadPoolExecutor (ExecutorService)
  - Executors

- //创建一个可缓存线程池，如果线程池长度超过处理需要，可灵活回收空闲线程，若无可回收，则新建线程（无限创建）。  
//最大线程数Integer.MAX\_VALUE，keepAliveTime: 60s  
// SynchronousQueue 没有容量，是无缓冲等待队列，是一个不存储元素的阻塞队列  
ExecutorService cachedThreadPool = Executors.newCachedThreadPool();
  - CachedThreadPool使用的是SynchronousQueue的  
入队：offer(E e)  
出队：poll(long timeout, TimeUnit unit)  
工作线程调用poll阻塞，等待timeout时间，如果超时，则返回null并回收线程；如果在等待期内，有任务入队，则成功返回任务，继续执行线程while循环。  
在runWorker()中getTask()方法中处理的。
- //创建一个定长线程池，可控制线程最大并发数，超出的线程会在队列中等待。  
//LinkedBlockingQueue 队列最大长度:Integer.MAX\_VALUE 是一个无界缓冲等待队列  
ExecutorService fixedThreadPool = Executors.newFixedThreadPool(100);
- //创建一个定长线程池，支持定时及周期性任务执行。  
//最大线程数:Integer.MAX\_VALUE  
//DelayedWorkQueue  
ExecutorService scheduledThreadPool = Executors.newScheduledThreadPool(100);
- //创建一个单线程化的线程池，它只会用唯一的工作线程来执行任务，保证所有任务按照指定顺序(FIFO, LIFO, 优先级)执行。  
//LinkedBlockingQueue 队列最大长度:Integer.MAX\_VALUE  
ExecutorService singleThreadExecutor = Executors.newSingleThreadExecutor();
- //自定义创建线程池  
ThreadPoolExecutor executor = new ThreadPoolExecutor(0,1,0, TimeUnit.MILLISECONDS,new LinkedBlockingQueue<>(10));
- 无法加入队列时抛出RejectedExecutionException

## ■ 方法

- Future<?> submit(()->{}) 提交一个可以获取返回结果的任务  
内部创建一个 java.util.concurrent.FutureTask,然后将FutureTask 加入到execute中,并返回改对象，若执行成功，则返回null,反之抛出异常
- void execute(()->{}) 提交一个任务
- void shutdown()
- void shutdownNow()

## • Concurrent包

### ◦ 原子类

java.util.concurrent.atomic

## ■ 原理

- CAS: CompareAndSwap  
CAS需要3个操作数：内存地址(V), 旧的预期值(A) 和新值(B)。  
CAS执行时，当且仅当V符合A时,处理器采用B更新V,否则不更新;不论是否更新，都返回旧值
- 使用sun.misc.Unsafe.compareAndSwapInt()  
或sun.misc.Unsafe.compareAndSwapLong()提供

sun.misc.Unsafe.getUnsafe()方法限制了只有启动类加载器(BootstrapClassLoader)才能访问该方法

这些方法总是返回旧值

- AtomicInteger 中 getAndAdd(), addAndGet(), incrementAndGet(), decrementAndGet() 中直接调用 unsafe.getAndAddInt(this, valueOffset, -1), unsafe.getAndAddLong 等方法, 这些方法总返回旧值  
在这些方法中, 使用do{oldVal = getVal()}while(compareAndSwapInt(this,oldVal,newVal)) 一直尝试更新数据, 直到成功

- AtomicInteger  
AtomicIntegerArray
  - volatile value
- AtomicBoolean
- AtomicLong  
AtomicLongArray
- AtomicReference  
AtomicReferenceArray

o 多线程同步

- java.util.concurrent.CountDownLatch
  - 原理
    - 利用 java.util.concurrent.locks.AbstractQueuedSynchronizer 实现基于AQS计数器和AQS阻塞队列  
初始值为线程数量

1. countDown()方法:

类似AtomicInteger, 基于同步器计数

countDown()时: 利用CAS 更新state值, 每次减一

最后一个执行countDown的线程会执行doReleaseShared()来释放阻塞队列中的线程

2. await():

执行await()后在AQS阻塞队列中创建一个等待节点, 并

LockSupport.park()阻塞当前线程, 直到最后一个线程调用countDown()释放

LockSupport.unpark()当前线程

- 作用
  - 使一个或多个线程等待其他线程完成后再执行
- 用途
  - 1、项目启动时的核心服务检测
  - 2、模拟高并发
  - 3、zookeeper连接

Zookeeper, Apache curator中使用了大量CountDownLatch 等待线程执行完成操作

- 问题
  - 只能使用1次
- java.util.concurrent.CyclicBarrier 循环屏障
- 原理

- CyclicBarrier内部定义了一个可重入锁ReentrantLock (非公平锁) 对象和Condition, 每当一个线程调用await方法的时候, 计数器减1, 同时判断计数器是否为0, 为0执行runnable barrierAction方法, 重置栅栏, 并唤醒所有在lock队列里面等待的线程, 否则进入lock的等待队列。

最后一个线程调用await()后, 会执行最后的Runnable方法

所有线程进入屏障后，最后一个线程执行完Runnable方法后，所有线程突破屏障继续执行后续操作。

([https://blog.csdn.net/qg\\_39241239/article/details/87030142](https://blog.csdn.net/qg_39241239/article/details/87030142))

- 示例：赛马

- 作用

- 让一组线程到达一个屏障（也可以叫同步点）时被阻塞，直到最后一个线程 到达屏障时，所有被屏障拦截的线程才会继续运行,当前线程被阻塞。

可重置，可重复执行

- 用途

- 可以用于多线程计算数据，最后合并计算结果的场景。

- 与CountDownLatch区别

- CyclicBarrier强调的是n个线程相互等待，直到完成再执行任务，计数器可以重置，复用，所以叫循环栅栏。

即n个线程等待执行完barrierAction后继续执行

- CountDownLatch允许一个或多个线程等待直到其他线程完成操作，只能使用一次。

- java.util.concurrent.Semaphore 信号量

- 原理

- 基于AQS(AbstractQueuedSynchronizer)

Semaphore内部原理是通过AQS实现的。Semaphore中定义了Sync抽象类，而Sync又继承了AbstractQueuedSynchronizer，Semaphore中对许可的获取与释放，是使用CAS通过对AQS中state的操作实现的。

Semaphore是J.U.C包下的许可控制类，维护了一个许可集，通常用于限制可以访问某些资源（物理或逻辑的）的线程数目，或对资源访问的许可控制。

- Semaphore对许可的分配有两种策略，公平策略和非公平策略，没有明确指明时，默认为非公平策略。

- 公平策略：根据方法调用顺序（即先进先出；FIFO）来选择线程、获得许可。非公平策略：不对线程获取许可的顺序做任何保证。

- 用途

- 服务限流，控制线程的并发数量

只有指定数量的线程可以执行，其他线程进入等待队列

- o Lock

- (synchronized)Object.wait/notify/notifyAll

- 1、wait()、notify/notifyAll() 方法是Object的本地final方法，无法被重写。

- wait 释放锁和CPU资源

- 2、wait()执行后拥有当前锁的线程会释放该线程锁，并处于等待状态（等待重新获取锁）

- 3、notify/notifyAll() 执行后会唤醒处于等待状态线程获取线程锁、只是notify()只会随机唤醒其中之一获取线程锁，notifyAll() 会唤醒所有处于等待状态的线程抢夺线程锁。

- java.util.concurrent.locks.ReentrantLock

- 原理

- ReentrantLock 为可重入独占锁，乐观锁(CAS)

- 利用AQS 中Node双向链表保存线程队列

- 同一时间只有一个线程在执行ReentrantLock.lock()方法后面得任务

- 公平锁

- 按加锁顺序获取锁

- 加锁

- lock.lock();

首先会判断 AQS 中的 state 是否等于 0, 0 表示目前没有其他线程获得锁, 当前线程就可以尝试获取锁。

注意:尝试之前会利用 hasQueuedPredecessors() 方法来判断 AQS 的队列中是否有其他线程, 如果有则不会尝试获取锁(这是公平锁特有的情况)。

如果队列中没有线程就利用 CAS 来将 AQS 中的 state 修改为 1, 也就是获取锁, 获取成功则将当前线程置为获得锁的独占线程 (setExclusiveOwnerThread(current))。

如果 state 大于 0 时, 说明锁已经被获取了, 则需要判断获取锁的线程是否为当前线程(ReentrantLock 支持重入), 是则需要将 state + 1, 并将值更新。

- 1. CAS判断state是否为0来表示锁是否被占用; 若未被占用, 则独占锁

3. 否则, 尝试获取锁 acquire()

4. 若尝试获取锁成功 (锁就是被当前线程占用的, 重入计数+1即可; 或者锁正好被释放)

5. 获取锁失败, 则需要创建一个Node节点 (包含了线程信息) 入队

6. 挂起线程 acquireQueued, 挂起之前, 会先尝试获取锁, 值有确认失败之后, 则挂起锁, 并设置前置Node的状态为SIGNAL (以保障在释放锁的时候, 可以保证唤醒Node的后驱节点线程)

- 写入队列

- 如果 tryAcquire(arg) 获取锁失败, 则需要用 addWaiter(Node.EXCLUSIVE) 将当前线程写入队列中。

写入之前需要将当前线程包装为一个 Node 对象(addWaiter(Node.EXCLUSIVE))。

- 非公平锁

- 随机获取锁

- ReentrantLock

- .lock() 加锁

- .unlock() 释放锁

- 小结

- 创建锁对象 Lock lock = new ReentrantLock()

- 在希望保证线程同步的代码之前显示调用 lock.lock() 尝试获取锁, 若被其他线程占用,

则阻塞

- 执行完之后, 一定得手动释放锁, 否则会造成死锁 lock.unlock(); 一般来讲, 把释放锁的逻辑, 放在需要线程同步的代码包装外的finally块中

- lock() 和 unlock() 配套使用, 不要出现一个比另一个用得多的情况

- 不要出现lock(),lock()连续调用的情况, 即两者之间没有释放锁unlock()的显示调用

- Condition与Lock配套使用, 通过 Lock#newCondition() 进行实例化

- lock() 加锁, 但不阻塞

condition.await() 释放锁,阻塞

condition.await()前需要加锁,后需要解锁

- Condition#await() 会释放lock, 线程阻塞; 直到线程中断or Condition#signal()被执行, 唤醒阻塞线程, 并重新获取lock

- Condition

- lock.newCondition() 创建条件锁

- await()

- // 使当前线程处于等待状态, 释放与Condition绑定的lock锁

// 直到 signal()方法被调用后, 被唤醒 (若中断, 就game over了)

// 唤醒后, 该线程会再次获取与条件绑定的 lock锁

- signal()

- 表示条件达成, 唤醒一个被条件阻塞的线程

- signalAll()

- 唤醒所有被条件阻塞的线程。
- 样例: ArrayBlockingQueue, LinkedBlockingQueue
- ReentrantReadWriteLock

读写锁: 读读共享, 写写互斥, 读写互斥, 写读互斥

- lock.readLock() 读锁(读读共享)
- lock() 允许多个线程同时执行lock()方法后面的代码
- lock.writeLock() 写锁

(写写互斥, 读写互斥)

- (写写互斥)writeLock().lock() 同一时间只允许一个线程执行lock()方法后面的代码
- (读写互斥) 读锁时,不能写,

写锁时不能读

- java.util.concurrent.locks

不可重入, 默认占有许可证, 可以响应中断

LockSupport.park() 获取当前线程许可, 阻塞线程

LockSupport.unpark() 取消当前线程许可, 唤醒线程

- 底层为sun.misc.Unsafe:

//park

public native void park(boolean isAbsolute, long time);

//unpark

public native void unpark(Object var1);

- 与Object类的wait/notify机制相比, park/unpark有两个优点
- 以thread为操作对象更符合阻塞线程的直观定义
- 操作更精准, 可以准确地唤醒某一个线程 (notify随机唤醒一个线程, notifyAll唤醒所有等待的线程), 增加了灵活性。
- sun.misc.Unsafe 提供CAS native操作方法

线程park/unpark

- Thread.sleep、Object.wait、LockSupport.park 区别

<https://www.cnblogs.com/tong-yuan/p/11768904.html>

- Thread.sleep(time)方法必须传入指定的时间, 线程将进入休眠状态, 通过jstack输出线程快照的话此时该线程的状态应该是TIMED\_WAITING, 表示休眠一段时间。不能主动唤醒。

通过sleep方法进入休眠的线程不会释放持有的锁

- Object.wait() 线程的状态都将是WAITING状态

必须获得对象上的锁后, 才可以执行该对象的wait方法。否则程序会在运行时抛出IllegalMonitorStateException异常。

Object.wait()方法需要在synchronized块中执行;

Object.wait()不带超时的, 需要另一个线程执行notify()来唤醒, 但不一定继续执行后续内容;

如果在wait()之前执行了notify()会怎样? 抛出IllegalMonitorStateException异常;

调用wait()方法后, 线程进入休眠的同时, 会释放持有的该对象的锁,

- 通过LockSupport.park()方法, 我们也可以让线程进入休眠。它的底层也是调用了Unsafe类的park方法, 会响应中断interrupt();

需要使用LockSupport.unpark() 唤醒;

LockSupport.park()不带超时的, 需要另一个线程执行unpark()来唤醒, 一定会继续执行后续内容;

如果在park()之前执行了unpark()会怎样? 线程不会被阻塞, 直接跳过park(), 继续执行后续内容;

调用park方法进入休眠后, 线程状态为WAITING

另外, 和wait方法不同, 执行park进入休眠后并不会释放持有的锁。

#### o 队列Queue

- java.util.concurrent.BlockingQueue  
基于ReentrantLock

- 实现(implement)

- ArrayBlockingQueue

- ArrayBlockingQueue底层是使用一个数组（Object数组）实现队列的，并且在构造ArrayBlockingQueue时需要指定容量，也就意味着底层数组一旦创建了，容量就不能改变了，因此ArrayBlockingQueue是一个容量限制的阻塞队列。

- 因此，在队列全满时执行入队将会阻塞，在队列为空时出队同样将会阻塞。

- ArrayBlockingQueue的并发阻塞是通过ReentrantLock和Condition来实现的，ArrayBlockingQueue内部只有一把锁，意味着同一时刻只有一个线程能进行入队或者出队的操作。

- LinkedBlockingQueue

- LinkedBlockingQueue是一个基于链表实现的可选容量的阻塞队列。队头的元素是插入时间最长的，队尾的元素是最新插入的。新的元素将会被插入到队列的尾部。

- 在于该队列至少有一个节点，头节点不含有元素

- 如果在初始化时没有指定容量，那么默认使用int的最大值作为队列容量

- 原理

- LinkedBlockingQueue中维持两把锁，一把锁用于入队putLock，一把锁用于出队takeLock，这也就意味着，同一时刻，只能有一个线程执行入队，其余执行入队的线程将会被阻塞；同时，可以有另一个线程执行出队，其余执行出队的线程将会被阻塞。换句话说，虽然入队和出队两个操作同时均只能有一个线程操作，但是可以一个入队线程和一个出队线程共同执行，也就意味着可能同时有两个线程在操作队列，那么为了维持线程安全，LinkedBlockingQueue使用一个AtomicInteger类型的变量表示当前队列中含有的元素个数，所以可以确保两个线程之间操作底层队列是线程安全的。

- LinkedBlockingQueue是允许两个线程同时在两端进行入队或出队的操作的，但一端同时只能有一个线程进行操作，这是通过两把锁来区分的；

- 为了维持底部数据的统一，引入了AtomicInteger的一个count变量，表示队列中元素的个数。count只能在两个地方变化，一个是入队的方法（可以+1），另一个是出队的方法（可以-1），而AtomicInteger是原子安全的，所以也就确保了底层队列的数据同步。

- LinkedBlockingDeque

- LinkedBlockingDeque是双向链表实现的双向并发阻塞队列。该阻塞队列同时支持FIFO和FILO两种操作方式，即可以从队列的头和尾同时操作(插入/删除)；并且，该阻塞队列是支持线程安全

- 1、add(E e)：在不违反容量限制的情况下，将指定的元素插入此双端队列的末尾，返回值为Boolean。
      - 2、addFirst(E e)：如果立即可行且不违反容量限制，则将指定的元素插入此双端队列的开头；如果当前没有空间可用，则抛出IllegalStateException。
      - 3、addLast(E e)：如果立即可行且不违反容量限制，则将指定的元素插入此双端队列的末尾；如果当前没有空间可用，则抛出IllegalStateException。
      - 包含队列的基本offer,put,poll,take，及offerFirst,offerLast(不阻塞)



pullFirst,putLast(阻塞)  
pollFirst,pollLast(不阻塞)  
takeFirst,takeLast(阻塞)

## ■ LinkedListBlockingQueue 与 ArrayBlockingQueue 比较

### ■ ArrayBlockingQueue

- 一个对象数组+一把锁+两个条件
- 入队与出队都用同一把锁
- 在只有入队高并发或出队高并发的情况下，因为操作数组，且不需要扩容，性能很高
- 采用了数组，必须指定大小，即容量有限

### ■ LinkedListBlockingQueue

- 一个单向链表+两把锁+两个条件
- 两把锁，一把用于入队，一把用于出队，有效的避免了入队与出队时使用一把锁带来的竞争。
- 在入队与出队都高并发的情况下，性能比ArrayBlockingQueue高很多
- 采用了链表，最大容量为整数最大值，可看做容量无限

### ■ SynchronousQueue

- SynchronousQueue没有容量，是无缓冲等待队列，是一个不存储元素的阻塞队列，会直接将任务交给消费者，必须等队列中的添加元素被消费后才能继续添加新的元素。

## ■ 操作

### ■ 入队

- add(E e): 内部调用offer，若队列满时抛出IllegalStateException("Queue full") 异常
- offer(E e): 如果队列没满，立即返回true； 如果队列满了，立即返回false-->不阻塞
- put(E e): 如果队列满了，一直阻塞，直到队列不满或者线程被中断-->阻塞

#### ■ put方法总结

- 1. LinkedListBlockingQueue不允许元素为null。
- 2. 同一时刻，只能有一个线程执行入队操作，因为putLock在将元素插入到队列尾部时加锁了
- 3. 如果队列满了，那么将会调用notFull的await()方法将该线程加入到Condition等待队列中。await()方法就会释放线程占有的锁，这将导致之前由于被锁阻塞的入队线程将会获取到锁，执行到while循环处，不过可能因为由于队列仍旧是满的，也被加入到条件队列中。
- 4. 一旦一个出队线程取走了一个元素，并通知了入队等待队列中可以释放线程了，那么第一个加入到Condition队列中的将会被释放，那么该线程将会重新获得put锁，继而执行enqueue()方法，将节点插入到队列的尾部
- 5. 然后得到插入一个节点之前的元素个数，如果队列中还有空间可以插入，那么就通知notFull条件的等待队列中的线程。
- 6. 通知出队线程队列为空了，因为插入一个元素之前的个数为0，而插入一个之后队列中的元素就从无变成了有，就可以通知因队列为空而阻塞的出队线程了。

- offer(E e, long timeout, TimeUnit unit): 在队尾插入一个元素,, 如果队列已满, 则进入等待, 直到出现以下三种情况: -->阻塞  
被唤醒  
等待时间超时  
当前线程被中断
- 出队
  - poll(): 如果没有元素, 直接返回null; 如果有元素, 出队
  - take(): 如果队列空了, 一直阻塞, 直到队列不为空或者线程被中断-->阻塞
    - take方法总结
      - 当队列为空时, 就加入到notEmpty(的条件等待队列中, 当队列不为空时就取走一个元素, 当取完发现还有元素可取时, 再通知一下自己的伙伴 (等待在条件队列中的线程); 最后, 如果队列从满到非满, 通知一下put线程。
  - poll(long timeout, TimeUnit unit): 如果队列不空, 出队; 如果队列已空且已经超时, 返回null; 如果队列已空且时间未超时, 则进入等待, 直到出现以下三种情况:  
被唤醒  
等待时间超时  
当前线程被中断
- remove
  - remove()方法用于删除队列中一个元素, 如果队列中不含有该元素, 那么返回false; 有的话则删除并返回true。入队和出队都是只获取一个锁, 而remove()方法需要同时获得两把锁fullLock(); 其他对队列 contains(),toString()等都是fullLock()

## JDK工具包

- Array 和 ArrayList  
List, LinkedList
  - Array
    - 优点: 在内存中是连续的, 速度较快, 操作简单。
    - 缺点: 定义数组时要定义其长度, 不是很灵活, 过长过短都会造成问题。不方便进行数据的添加、插入和移除
  - Array于ArrayList的区别
    - Array可以包含基本类型和对象类型, ArrayList只能包含对象类型。  
Array数组在存放的时候一定是同种类型的元素。ArrayList就存储Object[]对象数组
    - Array大小是固定的, ArrayList的大小是动态变化的。
    - ArrayList提供了更多的方法和特性, 比如: addAll(), removeAll(), iterator()等等。
    - 对于基本类型数据, 集合使用自动装箱来减少编码工作量。但是, 当处理固定大小的基本数据类型的时候, 这种方式相对比较慢。
  - ArrayList 内部为 Object[]
    - new ArrayList(Collection c)  
将c转换为数组保存到内部数组对象中, 若c中值不为Object[] 则, Array.copyOfOf为Object[]
    - 优点: 命名空间System.Collections下的一部分。大小是动态扩充与收缩的。在声明ArrayList对象时不需要指定它的长度。ArrayList继承了IList接口, 可以很方便的进行数据的添加、插入和移除。

- 缺点：当向集合插入不同类型的数据后（ArrayList将数据当作Object存储），在进行数据处理时容易出现类型不匹配的错误，使用时需要进行类型转换处理，存在装箱与拆箱操作，造成性能大量损耗的现象。
  - ArrayList 默认大小为10  
其后，默认增长数组的1/2大小  
最大为Integer.MAX\_VALUE
  - List为接口
  - LinkedList内部为链表Node  
每个节点存储Prev,Next元素索引  
双向链表
  - ArrayList和LinkedList的区别
    - ArrayList和LinkedList都实现了List接口，他们有以下不同点：
      - 1.ArrayList是基于索引的数据接口，它的底层是数组。它可以以O(1)时间复杂度对元素进行随机访问。与此对应，LinkedList是以元素列表的形式存储它的数据，每一个元素都和它的前一个和后一个元素链接在一起，在这种情况下，查找某个元素的时间复杂度是O(n)。
        - get(i)  
ArrayList 直接访问数组element[i]  
LinkedList 需要循环访问到节点i
      - 2.相对于ArrayList，LinkedList的插入，添加，删除操作速度更快，因为当元素被添加到集合任意位置的时候，不需要像数组那样重新计算大小或者是更新索引。
        - add(i,E)  
ArrayList插入元素时，将i位置的元素在数组中向后移动一位，再将E添加到i位置  
LinkedList 插入元素时，先获取到i位置的元素，  
然后添加新元素到i位置
      - 3.LinkedList比ArrayList更占内存，因为LinkedList为每一个节点存储了两个引用，一个指向前一个元素，一个指向下一个元素。
- HashMap、LinkedHashMap和HashTable
  - java.util.HashMap(1.8+)
    - 非线程安全  
HashMap继承自AbstractMap(implements Map)类，implements Map。  
可以用Collections.synchronizedMap(HashMap map)方法使HashMap具有同步的能力。
    - 并发下的问题
      - (1)多线程扩容，引起的死循环问题 (1.8已解决)
      - (2)多线程put的时候可能导致元素丢失
      - (3)put非null元素后get出来的却是null
    - 无序，根据key的hashCode值存储数据  
key允许null,且只有1个null key  
null key在HashMap第0个位置
    - 数据结构: 数组+链表(Node<K,V>[] table) 或 数组+红黑树（在JDK1.8中如果链表长度大于8的时候才转换为红黑树，平常不是），链表为单向链表  
数组默认大小16,数组每个位置叫做一个Bucket  
若hash(key)相同时,获取到当前hash位置存储的链表，若链表长度  
≥ TREEIFY\_THRESHOLD(8) 并且 数组长度 > 64 (n = tab.length) <

MIN\_TREEIFY\_CAPACITY(64) 时,将链表转为红黑树处理;

当 红黑树的长度小于6时 (UNTREEIFY\_THRESHOLD(6)) , 转换为链表存储。

HashMap底层是通过链表来解决hash冲突的。

- 为什么链表长度为8时,才转换为红黑树:

1. 随机hashCode算法下所有bin中节点的分布频率会遵循泊松分布,

2. 一个bin中链表长度达到8个元素的概率为0.00000006

3. more: less than 1 in ten million

- 为什么链表长度为6时, 由红黑树转换为链表:

4. 当链表长度为6时 查询的平均长度为  $n/2=3$

红黑树为  $\log(6)=2.6$

为8时: 链表  $8/2=4$

红黑树  $\log(8)=3$

中间有个差值7可以防止链表和树之间频繁的转换

- 扩容: 默认下 数组大小为16, 当元素超过160.75=12时, 会扩容为216=32  $Cap \ll 1$ , 即扩容1倍

capacity 的容量大小是 2 的 n 次幂

- hash值计算方式:  $(h = \text{key.hashCode()}) \wedge (h \ggg 16)$

key为null时为0

hashCode 的高 16 位不变, 低 16 位与高 16 位做一个异或。

- modCount: HashMap 在结构上被修改的次数, 结构修改是指改变HashMap中映射的次数, 或者以其他方式修改其内部结构(例如, rehash)。此字段用于使HashMap集合视图上的迭代器快速失败(fail fast)。(著名的ConcurrentModificationException便与此有关)

- 遍历过程中不能修改HashMap,否则会抛出

ConcurrentModificationException

因为  $\text{modCount} \neq \text{expectedModCount}$

- index计算:  $i = (n - 1) \& \text{hash}$

$= \text{hash} \% \text{length}$

- JDK1.8 修改

5. 数据结构:

数组+链表 为

数组+红黑树

6. 优化了hash算法

$h \wedge (h \ggg 16)$

7. 扩容后, 元素要么是在原位置, 要么是在原位置再移动2次幂的位置, 且链表顺序不变。

- 遍历方式Iterator

- HashMap转同步

- `java.util.Collections.unmodifiableMap(map)`

- LinkedHashMap extends HashMap

- 非线程安全

- 有序

记录的插入顺序, 在用Iteraor遍历LinkedHashMap时, 先得到的记录肯定是先插入的。在遍历的时候会比HashMap慢。有HashMap的全部特性。

- 插入顺序(默认)

- 访问顺序

`boolean accessOrder= false;`

`put/get`已存在的Entry时, 会把Entry移动到双向链表的表尾

- 数据结构：双向链表+HashMap  
初始化时创建了只有头结点的双向链表  
单独维护了1个双向链表,head ,tail
- 重写了Map的  
newNode,newTreeNode,replacementNode,replacementTreeNode,entrySet() 等方法  
重新实现了Entry, LinkedEntrySet,LinkedHashIterator
- HashTable
  - 线程安全  
大部分方法线程同步
  - Hashtable继承自Dictionary类.implements Map
  - 保存key/value数据,  
key/value都不能为空  
实现同HashMap  
默认容量11, 不要求底层数组容量为2 的整数次幂  
扩容为old\*2+1  
数据结构: 数组+链表
    - Key不能为null,因为完全使用Key的hashCode
  - hash值计算方式: key.hashCode()
  - Entry<K,V> tables index : (hash & 0x7FFFFFFF) % tab.length
  - 遍历方式: Enumeration
- TreeMap
  - TreeMap
    - extends AbstractMap<K,V>  
implements NavigableMap<K,V>
    - 非线程安全
    - 有序且可以进行排序  
TreeMap能够把它保存的记录根据键排序, 默认是按升序排序, 也可以指定排序的  
比较器。当用Iteraor遍历TreeMap时, 得到的记录是排过序的。  
TreeMap的键和值都不能为空。  
保证了 key 的大小排序性  
默认按key的自然序排序  
可置顶Comparator排序
      - 添加子节点时, 若无根节点, 则该节点为黑色根节点;  
若存在根节点, 则循环查找新节点的存储位置;  
找到位置后, 处理红黑树, 默认设置节点为红色
    - 数据结构: 红黑数
      - 红黑树二叉树而言我们必须增加如下规则:
        - 1、每个节点都只能是红色或者黑色
        - 2、根节点是黑色
        - 3、每个叶节点 (NIL节点, 空节点) 是黑色的。
        - 4、如果一个结点是红的, 则它两个子节点都是黑的。也就是说在一条路径上  
不能出现相邻的两个红色结点。
        - 5、从任一节点到其每个叶子的所有路径都包含相同数目的黑色节点。
    - 红黑树的操作
      - 左旋

- 右旋
    - 着色
  - 根节点为黑色
  - 若父节点为黑色时，子节点插入时为红色
- HashSet 和 TreeSet
  - HashSet
    - 继承于AbstractSet implements Set
    - 默认为HashMap实现，值为Map Key,Map Value为 Object
  - LinkedHashMap
    - extends HashSet implements Set  
由LinkedHashMap实现
  - TreeSet
    - 继承于AbstractSet implements NavigableSet(extends SortedSet)  
默认为TreeMap实现，值为Map Key,Map Value为 Object

## JDK8新特性

- Java语言新特性
  - Lambdas 与 Functional 接口
    - Lambdas
      - Lambda允许把函数作为参数传到方法中或把代码看成数据
      - Lambda 可以引用类的成员变量和局部变量，如果这些变量不是final的话，会被隐式转换为final  
(匿名类调用所在类的成员变量和局部变量时,需要将传入的变量转换为final,防止成员变量和局部变量在匿名类中被修改。)  
Lambda最终会动态生成匿名内部类(但不会生成.class文件,反射可以查看类信息), 类名如: LambdaTest\$\$Lambda\$1/4036432  
->(箭头)函数
      - Lambda可以返回一个值，若Lambda的函数体只有1行的话，return可以省略。
      - java.util.function.Consumer;  
accept(T t)
      - 如: Arrays.asList("a","b","c").forEach(e->System.out.println(e));  
或  
Arrays.asList("a","b","c").forEach(e-> {  
System.out.println(e);  
System.out.println(e);  
});
    - Functional
      - 为了友好支持Lambda，增加了函数式接口概念  
函数式接口可以隐式转换为Lambda表达式
      - @FunctionalInterface 标注函数式接口
      - 默认方法与静态方法不影响函数式接口的定义
      - java.util.function包  
常用的函数式接口
        - Consumer  
接受一个输入参数，无返回值
          - T apply(T t)

- Int,Long,Double,ObjInt... Consumer
    - Consumer<String> consumer = (x) ->
 

```
System.out.println("consumer: " + x);
consumer.accept("Hello");
```
  - Supplier
 无输入参数, 返回一个结果
    - T get()
      - Boolean,Int,Long,Double...Supplier
        - Supplier<String> supplier = () -> "Test supplier";
 supplier.get();
  - Predicate
 接受一个输入参数, 返回布尔值
 Predicate 是一个可以指定入参类型, 并返回 boolean 值的函数式接口。它内部提供了一些带有默认实现的方法, 可以被用来组合一个复杂的逻辑判断 (and, or, negate)
    - and(Predicate<? super T> other)
 negate(Predicate<? super T> other)
      - Predicate<String> predicate = (x) -> x.length() > 0;
 predicate.test("String");
  - Function
 接受一个输入参数, 返回一个结果
 Function 函数式接口的作用是, 我们可以为其提供一个原料, 他给生产一个最终的产品。通过它提供的默认方法, 组合,链行处理(compose, andThen):
    - R apply(T t)
      - Function<Integer, String> function1 = (x) -> "result: " + x;
 function1.apply(6);
  - //比较接口
 java.util.Comparator
- 接口的默认方法与静态方法
  - 接口default标识符
 default void run()
    - 如:java.lang.Iterable实现的forEach,spliterator,stream等
    - 不需要子类实现, 子类可重写
  - 接口static 标识符 声明静态方法
 static void runS()
- 方法引用
 (直接引用已有Java类或对象(实例)的方法或构造器)
  - Class::new 构造器引用
    - final Car car = Car.create(Car::new);
 final Car car2 = Car.create(() -> new Car());
  - Class::static\_method 静态方法引用
    - cars.forEach(Car::collide);
 cars2.forEach((c)->Car.collide(c));
  - Class::method 特定类的任意方法引用
    - cars.forEach(Car::repair);
 cars2.forEach((c)->c.repair());
  - instance::method 特定类的任意对象的方法引用

- cars.forEach(Car.create(Car::new)::follow);
    - cars.forEach(police::follow);
    - cars2.forEach((c)->police.follow(c));
  - 重复注解
    - @Repeatable
      - @Repeatable(Filters.class)
        - //value必须为注解数组类
        - //Filters必须与Filter的注解相同
        - @Target(ElementType.TYPE)
        - @Retention(RetentionPolicy.RUNTIME)
        - public @interface Filters{
          - Filter[] value();
        - }
        - @Target(ElementType.TYPE)
        - @Retention(RetentionPolicy.RUNTIME)
        - public @interface Filter{
          - String value();
        - }
      - class.getAnnotationsByType(Filters.class)
        - 获取相同类型的注解,返回Filter注解数组
        - Filter[] = [@Filter1,@Filter2];
      - class.getDeclaredAnnotations()
        - 获取到的注解中若存在@Repeatable注解, 则获取到的注解为注解数组, 即
        - Filters(value=[@Filter1,@Filter2])
    - 更好的类型推测机制
    - 扩展注解
      - 可以为任何东西添加注解:
      - 局部变量, 泛型类、父类、接口类的实现、方法的异常
  - @Target
    - ElementType.TYPE\_USE
    - ElementType.TYPE\_PARAMETER
  - import java.lang.annotation.ElementType;
    - import java.lang.annotation.Retention;
    - import java.lang.annotation.RetentionPolicy;
    - import java.lang.annotation.Target;
    - import java.util.ArrayList;
    - import java.util.Collection;
    - public class Annotations {
      - @Retention( RetentionPolicy.RUNTIME )
      - @Target( { ElementType.TYPE\_USE, ElementType.TYPE\_PARAMETER } )
      - public @interface NonEmpty {
        - public static class Holder< @NonEmpty T > extends @NonEmpty object {
          - public void method() throws @NonEmpty Exception [
          - }
      - }
      - @SuppressWarnings( "unused")
      - public static void main(String[] args){
        - final Holder< String > holder = new @NonEmpty Holder< string >();
        - @NonEmpty Collections @NonEmpty String > strings = new ArrayList0;
        - }
      - }



- Java 类库新特性

- java.util.Optional

null值容器类

- //map对当前Optional值进行转换, orElse: 设置默认值  
fullName.map(s->"Hey "+s+"!").orElse("Default");
- Optional.empty() 创建空的Optional对象
  - get() 会抛出NoSuchElementException
- Optional.of(obj) 创建一个包含obj的Optional对象
  - obj不能为null,否则会抛出NPE
- Optional.ofNullable(obj); 创建一个包含obj的Optional对象
  - obj可以为null, 为null时同Optional.empty()
- of.get() 获取Optional包含的值
- of.isPresent() 判断是否为空
- of.ifPresent(t-> System.out.println(t)) 若Optional包含值不为空时, 提供一个Consumer
- //为空时提供回调方法(Supplier)产生默认值  
of.orElseGet(()->"[none]")
  - of.orElseGet(Object::new)
  - orElseGet 因为使用了Supplier,所在传值时不会计算Supplier中的值  
所以当of不为空时, orElseGet中传入的值不计算
- //为空时提供默认值  
of.orElse("Default");
  - 因为orElse的参数为具体对象或值, 所以在传入参数时, 会计算参数结果, 导致不管of是否为空, 都会执行参数计算
- //为空时返回自定义异常  
of.orElseThrow(Supplier<Exception> t)
- //值转换, 返回一个Optional对象, 传入函数值为任何值  
Optional<T> map(Function<? super T, ? extends U> mapper)  
of.map(t -> t).orElseGet(()->"t")
- //值转换, 返回一个Optional对象, 传入值需为Optional对象  
<T> Optional<T> flatMap(Function<? super T, Optional> mapper)
- //值过滤, 若值不满足条件, 则返回Optional.empty();  
of.filter(t -> t != null).orElse("null");

- java.util.Stream

Stream 是对集合 (Collection) 对象功能的增强, 它专注于对集合对象进行各种非常便利、高效的聚合操作 (aggregate operation), 或者大批量数据操作 (bulk data operation) 简化了集合对象的处理

优点:

减少了迭代次数, 也避免了存储中间结果

- 生成Stream
  - Stream.generate(Obj);
  - Collection 和数组
    - Collection.stream()
    - Collection.parallelStream()
    - Arrays.stream(T array) or Stream.of()
  - BufferedReader

- java.io.BufferedReader.lines()
- 静态工厂
  - java.util.stream.IntStream.range()
  - java.nio.file.Files.walk()
- 自己构建
  - java.util.Spliterator
  - StreamSupport.stream(new Spliterator() {})
- 其它
  - Random.ints()
  - BitSet.stream()
  - Pattern.splitAsStream(java.lang.CharSequence)
  - JarFile.stream()
- 流的操作类型分为两种：
  - Intermediate: (中间操作)一个流可以后面跟随零个或多个 intermediate 操作。其目的主要是打开流，做出某种程度的数据映射/过滤，然后返回一个新的流，交给下一个操作使用。这类操作都是惰性的（lazy），就是说，仅仅调用到这类方法，并没有真正开始流的遍历。
    - map、mapToInt、mapToLong、mapToDouble、flatMap、flatMapToInt、flatMapToLong、flatMapToDouble、filter、distinct、sorted、peek、limit、skip、parallel、sequential、unordered
  - Terminal: (终止操作)一个流只能有一个 terminal 操作，当这个操作执行后，流就被使用“光”了，无法再被操作。所以这必定是流的最后一个操作。Terminal 操作的执行，才会真正开始流的遍历，并且会生成一个结果，或者一个 side effect。
    - Stream.collect()
    - Stream.count()
    - Stream.min()
    - Stream.max()
    - Stream.forEach()
    - Stream.forEachOrdered()
    - Stream.reduce()
    - Stream.toArray()
    - 短路操作(short-circuiting operations)
 

如果一个中间操作（intermediate operation）的输入是无限的，而它可以产生一个有限stream结果，那么这个中间操作是短路的（short-circuiting）；

      - Stream.allMatch()
      - Stream.anyMatch()
      - Stream.noneMatch()
      - Stream.findAny()
      - Stream.findFirst()
  - 有状态的操作（Stateful operations）
    - distinct() sorted() sorted() limit() skip()
- IntStream、LongStream、DoubleStream
  - summaryStatistics
 

统计

```
IntStream.builder().build().summaryStatistics();
```
  - 为什么单独提供IntStream,LongStream,DoubleStream
- //将Stream中的元素映射为另一种类型

```
Stream.map()
```

- //将Stream中的元素映射为另一种类型  
//flatMap是一种扁平化、一对多的操作，会返回一个新的Stream，因此flatMap实现需要返回一个新的Stream对象  
Stream.flatMap()  
flatMap(e -> Stream.of(e.split(" ")))
  - Arrays.asList("Stream operations is a new API in Java 8")  
.stream()  
.flatMap(e -> Stream.of(e.split(" ")))  
.forEach(System.out::println);
- //规约操作（reduction operation），也称作折叠（fold）。可以通过相同的合并方法将系列的输入组合成一个结果。  
Stream.reduce
  - //函数接口BinaryOperator<T> accumulator称为累加器，需满足结合律，其抽象方法为T apply(T t1, T t2); 有两个参数和返回值，且他们的类型相同。t1为之前结合的结果，t2为当前元素，该抽象方法的实现需要将t1和t2结合，并返回结果。  
Optional<T> reduce(BinaryOperator<T> accumulator)
  - T reduce(T identity, BinaryOperator<T> accumulator) 跟1) 相似，只不过需要制定计算时的初始值，
- Stream.parallelStream()  
与  
Stream.stream() 实现区别
  - parallelStream  
并行流
    - stream.collect()执行并行的条件
      - isParallel() 指定为并行操作  
同时  
collector对象支持并行操作  
同时  
(非排序 || collector对象指定非排序)  
isParallel()  
&&  
(collector.characteristics().contains(Collector.Characteristics.CONCURRENT))  
&& (!isOrdered() ||  
collector.characteristics().contains(Collector.Characteristics.UNORDERED))
      - 使用ForkJoinTask实现并行
  - stream  
串行流
- 流的基本特点
  - 集合是对一组特定类型的元素值序列提供的接口 是数据结构,提供了元素的存取
  - 流也是对一组特定类型元素值序列提供的接口,在于计算,提供了对元素序列的操作  
计算方式 比如 filter map等
  - 流只能运行一次
  - 流由源 0个或者多个中间操作以及结束操作组成
  - 流操作的方法基本上是函数式接口的实例
  - 流的中间操作是惰性的并不会立即执行 这更有利于内部迭代的优化
  - 流借助于它内部迭代特性提供了声明式的编程方式 更简洁
  - 中间操作本身会返回一个流,可以将多个操作复合叠加,形成一个更大的流水线

- 流分为顺序和并行两种方式
- Stream的每个操作被抽象为1个stage,所有操作组成了一个双向链表,这些Stream对象以双向链表的形式组织在一起,构成整个流水线,由于每个Stage都记录了前一个Stage和本次的操作以及回调函数,依靠这种结构就能建立起对数据源的所有操作  
每个stage被1个Sink包裹,为每个操作具体的行为操作,也叫做回调
- 一个sink有两种状态,初始/激活  
开始时是初始状态,begin 激活, end使之回到初始状态,可以重复利用  
accept只能在激活状态使用

Sink用于协调相邻的stage之间的数据调用

通过begin end accept方法 以及cancellationRequested短路标志位来控制处理流程,对数据进行管控

- Sink.begin(size)

Sink.end()

Sink.cancellationRequested()

Sink.accept

- void begin(long size) 开始遍历元素之前调用该方法,通知Sink做好准备。

- void end() 所有元素遍历完成之后调用,通知Sink没有更多的元素了。

- boolean cancellationRequested() 是否可以结束操作,可以让短路操作尽早结束。

- void accept(T t) 遍历元素时调用,接受一个待处理元素,并对元素进行处理。Stage把自己包含的操作和回调方法封装到该方法里,前一个Stage只需要调用当前Stage.accept(T t)方法就行了。

- Date/Time API(JSR 310)

java.time

- Clock clock = Clock.systemUTC();

- //基于Clock实现,只返回ISO-8601且无时区信息的日期部分

LocalDate localDate = LocalDate.now();

//2019-12-12

- //基于Clock实现,只返回ISO-8601且无时区信息的时间部分

LocalTime localTime = LocalTime.now();

//11:24:44.153

- //基于Clock实现,只返回ISO-8601且无时区信息的日期+时间部分

LocalDateTime localDateTime = LocalDateTime.now();

//2019-12-12T11:24:44.153

- JavaScript引擎Nashorn

- java.util.Base64

- Base64.getEncoder().encode(byte[])

- Base64.getDecoder().decode(String)

- 并行(parallel)数组

- Arrays.parallelSort()

- 并发(Concurrency)

- ConcurrentHashMap新增方法支持聚合操作

#### • Java 编译器新特性

##### ◦ 参数名字

javac -parameters启用

Parameter.getName()获取

- 不使用 -parameter 编译后获取的参数为  
arg0

使用后获取为args

#### • Java虚拟机新特性

- PermSpace (永久代)移除, 新增MetaSpace
  - XX:MetaSpaceSize
  - XX:MaxMetaSpaceSize
- 新增Java工具
  - Nashorn引擎:jjs
    - //接收参数为js文件,返回执行的结果  
jjs file.js
  - 类依赖分析器: jdeps
    - 接收.class文件,目录,jar作为输入, 结果输出到控制台

## 其他

- 异常淹没,返回值栈
  - 当Java程序的方法中有异常处理块的时候, 执行引擎可能需要处理多了返回值, 这时候, 执行引擎会将处理到的返回值, 压入到返回值栈中。
  - 不建议在finally中进行return, 因为当有异常发生且catch块中又抛出新的异常, 会淹没异常
  - try-catch中若有finally不提供catch块(catch可多个), 异常会被淹没
- //若为null抛出NPE  
Objects.requireNonNull(obj)
- Hash冲突解决方法
  - 开放地址法
    - 线型探测法:  
当冲突发生后, 直接去下一个位置找是否存在没用的位置, 例如2位置发生冲突, 然后去下一位置3查找, 如果3也被占用, 去找4, 直到问题解决
      - 问题: 这样就会导致落在区间内的关键字Key要进行多次探测才能找到合适的位置, 并且还会继续增大这个连续区间, 使探测时间变得更长, 这样的现象被称为“一次聚集 (primary clustering)”, 也就是说越后面的数, 如果发生hash冲突, 探测的时间越长, 因为前面的数都已经将很多可用区域占了。
    - 平方探测法:  
当冲突发生后, 当直接每次增长 $i$ 的2 倍, 即 $2 \text{ (hash值)} + (-) i^2$
  - 再哈希法(双哈希法):  
多个不同的Hash函数, 当发生冲突时, 使用第二个, 第三个, ..., 等哈希函数计算地址, 直到无冲突。虽然不易发生聚集, 但是增加了计算时间。
  - 链地址法:  
链地址法的基本思想是: 每个哈希表节点都有一个next指针, 多个哈希表节点可以用next指针构成一个单向链表, 被分配到同一个索引上的多个节点可以用这个单向链表连接起来
    - HashMap 使用该方法解决hash冲突
  - 建立公共溢出区:  
将哈希表分为基本表和溢出表两部分, 凡是和基本表发生冲突的元素, 一律填入溢出表
- clone实现方式
  - 浅拷贝: java.lang.Object.clone()  
对象实现Cloneable接口并重写Object类中的clone()方法  
对象中的封装类型数据在对象clone以后引用的同一个对象
  - 深拷贝:

1. 对象实现Serializable接口, 通过对象的序列化和反序列化实现克隆

2. 对象实现Cloneable接口并重写Object类中的clone()方法, 手动拷贝, 将封装数据类型copy一次赋值

- equals实现原则  
java.lang.Object
  - 自反性 reflexive
    - 对于任何非null的引用值x, x.equals(x)必须返回true。  
x.equals(x) 必须为true
  - 对称性 symmetric
    - 对于任何非null的引用值x和y,当且仅当y.equals(x)返回true时, x.equals(y)必须返回true  
x.equals(y)的返回要和x.equals(y)的返回保持一致
  - 传递性 transitive
    - 对于任何非null的引用值x,y和z, 如果x.equals(y)返回true, 并且y.equals(z)返回true, 那么x.equals(z)返回true  
x.equals(y)==true , y.equals(z)==true , 则 x.equals(z) == true
  - 一致性 consistent
    - 对于任何非null的引用值x和y, 只要equals的比较操作在对象中所用的信息没有被修改, 多次调用x.equals(y)就会一致地返回
  - 任何非null引用值x ,x.equals(null) 必须返回false

## DB

---

### NoSQL(Not Only SQL)

NoSQL, 泛指非关系型的数据库

- Redis  
(Key/Value)
  - Redis: 内存Key-Value数据库  
<https://redis.io/>  
<http://www.redis.cn/>
    - Redis数据全部保存在内存中  
只有save后会持久化到磁盘
    - Redis key是二进制安全的, 可以使用任何二进制序列作为key
    - 空字符串也是有效key
    - 规则
      - key 不要太长, 不仅因为消耗内存, 而且在数据中查找这类键值的计算成本很高。
      - key 不要太短
      - 最好使用一种模式,如 user:1000:password按:分割的模式
  - 基础数据类型
    - Strings 字符串
      - 内容长度不能超过512M
      - SET 设置值: set mykey myvalue
      - GET 获取值: get mykey
      - SETNX key value 当key不存在时设置值

- SETEX key seconds value 设置key有效期, 单位s
- PSETEX key milliseconds value 设置key有效期,单位ms
- SET key val NX [EX seconds] [PX milliseconds] 当key不存在时设置值: set mykey newval nx
- SET key val XX [EX seconds] [PX milliseconds] 当key存在时设置值: set mykey newval xx
- INCR原子递增
  - set counter 100  
incr counter
  - INCR 命令将字符串值解析成整型, 将其加一, 最后将结果保存为新的字符串值,
  - 类似的命令有INCRBY, DECR 和 DECRBY。
- INCRBY原子递增指定值
  - SET mykey "10"  
INCRBY mykey 5  
(integer 15)
- DECR, DECRBY 原子递减
- GETSET 设置新值并返回旧值
- MSET批量设置多个keyvalue
  - mset a 10 b 20
- MGET 批量获取多个key的值
  - mget a b  
(  
1) "10"  
2) "20"  
3) "30"  
)
- STRLEN key 获取key value的长度
- APPEND key value 在value 尾部追加value
- GETRANGE key start end 获取指定范围的字符串
- SETRANGE key offset value 重写指定位置开始的字符串为新字符串
- Hashes 散列
  - 由field和关联的value组成的map。field和value都是字符串的hash的域数量没有限制(除内存外)
  - HSET key field val 设置HASH 值
  - HMSET mykey field1 val1 key2 field2 val2 ... 设置多个HASH field value值
    - hmset user:1000 username antirez birthyear 1977 verified 1
  - HSETNX key field val 当field不存在时设置HASH field, 存在时不处理
  - HMGET mykey field1 field2 .. 获取多个HASH key的值
- > hmget user:1000 username birthyear no-such-field  
1) "antirez"  
2) "1977"  
3) (nil)
  - HGET mykey field1 获取HASH field的值

- > hget user:1000 username  
"antirez"
  - HGETALL mykey 获取所有HASH field/value对
- redis> HSET myhash field1 "Hello"  
(integer) 1  
redis> HSET myhash field2 "World"  
(integer) 1  
redis> HGETALL myhash  
1) "field1"  
2) "Hello"  
3) "field2"  
4) "World"
  - HEXISTS mykey field 获取HASH field 是否存在
- HEXISTS myhash field1  
(integer) 0
  - HDEL mykey field1 field2 ... 删除HASH 多个field
- redis> HDEL myhash field2  
(integer) 0
  - HINCRBY key field increment 原子递增(increment)HASH field值
- redis> HSET myhash field 5  
(integer) 1  
redis> HINCRBY myhash field 1  
(integer) 6
  - HINCRBYFLOAT key field increment 原子递增浮点数 HASH field值
  - HKEYS key 返回所有HASH field名称
  - HLEN key 获取HASH field数量
  - HSCAN key cursor [MATCH pattern] [COUNT count] 同SCAN
  - HSTRLEN key field 获取HASH field字符串长度
  - HVALS key 获取所有HASH values
- redis> HSET myhash field1 "Hello"  
(integer) 1  
redis> HSET myhash field2 "World"  
(integer) 1  
redis> HVALS myhash  
1) "Hello"  
2) "World"
- Lists 列表
  - 按插入顺序排序的字符串元素的集合，由链表实现
  - LPUSH mykey val1 val2 ... 向list头部添加多个元素，返回list 长度
  - RPUSH mykey val1 val2 ... 向list尾部添加多个元素，返回list 长度
  - LRange mykey startIdx stopIdx 从list中取出一定范围的元素
    - LRange 带有两个索引，一定范围的第一个和最后一个元素。这两个索引都可以为负来告知Redis从尾部开始计数，因此-1表示最后一个元素，-2表示list中的倒数第二个元素，以此类推
  - LINDEX key index 获取list中某个索引的元素值
  - LPOP mykey 从头部删除并返回删除的元素
  - RPOP mykey 从尾部删除并返回删除的元素



- POP 空list时返回null
- LTRIM mykey startIdx stopIdx 把list从左边截取指定长度。
  - 从0位起截取到2索引位, 3个元素: ltrim mykey 0 2
- list上的阻塞操作(可实现生产者, 消费者队列)
  - BLPOP mylist1 mylist2 ... TIMEOUT 阻塞式从头部删除元素
    - TIMEOUT 为指定阻塞时间, 若为0时, 表示一直阻塞
  - BRPOP mylist1 mylist2 ... TIMEOUT 阻塞式从尾部删除元素
  - 多批量回复(multi-bulk-reply): 具体来说:  
 当没有元素可以被弹出时返回一个 nil 的多批量值, 并且 timeout 过期。  
 当有元素弹出时会返回一个双元素的多批量值, 其中第一个元素是弹出元素的 key, 第二个元素是 value。  
 (
 

```
redis> RPUSH list1 a b c
(integer) 3
redis> BRPOP list1 list2 0
1) "list1"
2) "c"
)
```
- RPOPLPUSH source destination
- 原子性地返回并移除存储在 source 的列表的最后一个元素 (列表尾部元素), 并把该元素放入存储在 destination 的列表的第一个元素位置 (列表头部)。  
 返回值:  
 bulk-string-reply: 被移除和放入的元素  
 - 例如: 假设 source 存储着列表 a,b,c, destination存储着列表 x,y,z。执行 RPOPLPUSH 得到的结果是 source 保存着列表 a,b, 而 destination 保存着列表 c,x,y,z。
- BRPOPLPUSH
- LLEN mylist 获取list长度
- LREM key count element 删除列表中的指定元素
- count > 0: 删除等于element从头到尾移动的元素。  
 count < 0: 删除等于element从尾到头移动的元素。  
 count = 0: 删除所有等于的元素element。
  - ```
redis> RPUSH mylist "hello"
(integer) 1
redis> RPUSH mylist "hello"
(integer) 2
redis> RPUSH mylist "foo"
(integer) 3
redis> RPUSH mylist "hello"
(integer) 4
redis> LREM mylist -2 "hello"
(integer) 2
redis> LRANGE mylist 0 -1
1) "hello"
2) "foo"
redis>
```
  - 例如, LREM list -2 "hello"将删除"hello"存储在中的列表中最后两个出现的list。

- LSET key index element 设置列表中某个位置的元素
- redis> RPUSH mylist "one"  
(integer) 1  
redis> RPUSH mylist "two"  
(integer) 2  
redis> RPUSH mylist "three"  
(integer) 3  
redis> LSET mylist 0 "four"  
"OK"  
redis> LSET mylist -2 "five"  
"OK"  
redis> LRANGE mylist 0 -1  
1) "four"  
2) "five"  
3) "three"  
redis>
- Sets 集合
  - 不重复且无序的字符串元素的集合。
  - SADD key value1 ,value2 ... 添加新元素到set中
  - SMEMBERS key 获取set中的所有元素
    - redis> smembers myset1
  - 数据量大时，生产环境慎用，该操作会导致服务器阻塞  
使用SSCAN替代
  - SISMEMBER key value 检测元素是否存在，1:是 0:否
  - SINTER key [key ...] 返回从所有给定集合的交集得到的集合成员
  - key1 = {a,b,c,d}  
key2 = {c}  
key3 = {a,c,e}  
SINTER key1 key2 key3 = {c}
  - SINTERSTORE destination key [key ...] 获取给定集合的交集并保存到destination
  - SDIFF key [key ...] 返回给定集合的差集
  - key1 = {a,b,c,d}  
key2 = {c}  
key3 = {a,c,e}  
SDIFF key1 key2 key3 = {b,d}
  - SDIFFSTORE destination key [key ...] 获取给定集合的差集并保存到destination  
中
  - SUNION key [key ...] 获取给定集合的合集
  - key1 = {a,b,c,d}  
key2 = {c}  
key3 = {a,c,e}  
SUNION key1 key2 key3 = {a,b,c,d,e}
  - SUNIONSTORE destination key [key ...]  
获取给定集合的合集并保存到destination中
    - SCARD key 获取集合中的元素个数

- SPOP key [count] 随机删除集合中的1个或多个元素，并返回(count参数始于3.2版本)
  - redis> spop myset 1
- SREM key member [member ...] 删除集合中的多个元素，返回从集合中删除的元素数量
- redis> SREM myset "four"  
(integer) 0
- SRANDMEMBER key [count] 从集合中随机获取指定数量的元素（不删除集合中的元素）
- SMOVE source destination member 将集合中的元素移动到另一个集合中，返回元素是否移动成功，1：是，0：否
- redis> SMOVE myset myotherset "two"  
(integer) 0
- SSCAN
- Sorted Sets 有序集合
  - 类似Sets,但是每个字符串元素都关联到一个叫score浮动数值（floating number value）。里面的元素总是通过score进行着排序，所以不同的是，它是可以检索的一系列元素。
  - 排序规则
    - 如果A和B是两个具有不同分数的元素，那么如果A.score是> B.score，则A>B。
    - 如果A和B的分数完全相同，那么如果A字符串在字典上大于B字符串，则A>B。A和B字符串不能相等，因为排序集仅具有唯一元素。
- ZADD key [NX|XX] [CH] [INCR] score member [score member ...] 将具有指定分数的成员添加到有序集合中，可指定多个分数/成员对。返回添加到有序集合中的元素数量，不含更新分数的现有元素。  
+inf和-inf值也是有效值
  - XX：只更新已存在元素
  - NX：元素不存在时，创建新元素
- ZRANGE key start stop [WITHSCORES] 返回有序列表中指定范围的元素，如果得分相同，将按字典排序。带有WITHSCORES 时同时返回每个元素的分数
- redis> ZRANGE myzset 0 1 WITHSCORES
  - 1) "one"
  - 2) "1"
  - 3) "two"
  - 4) "2"
- ZREVRANGE key start stop [WITHSCORES] 返回有序集key中，指定区间内的成员。其中成员的位置按score值递减(从大到小)来排列。具有相同score值的成员按字典序的反序排列。
- ZCARD key 返回有序列表的基数（元素数）
- redis> ZCARD myzset  
(integer) 2
- ZCOUNT key min max 返回有序集合中的元素数量，元素得分介于min和max间，默认包含min max的分数。  
(min (max 时不包含值为min max分数的元素)
  - redis> ZADD myzset 1 "one"  
(integer) 1
  - redis> ZADD myzset 2 "two"

```
(integer) 1
redis> ZADD myzset 3 "three"
(integer) 1
redis> ZCOUNT myzset -inf +inf
(integer) 3
redis> ZCOUNT myzset (1 3
(integer) 2
redis>
```

- redis> zcount myzset (1 3
 (integer) 1
- ZINCRBY key increment member 递增元素在有序集合中的分数
- ZSCORE key member 返回有序集合中元素的分数
- redis> ZADD myzset 1 "one"
 (integer) 1
 redis> ZSCORE myzset "one"
 "1"
 redis>
- ZREM key member [member ...] 删除有序集合中多个元素
- 等: <https://redis.io/commands/zunionstore>

#### ■ Bitmaps (Bit arrays)

- simply bitmaps: 通过特殊的命令, 你可以将 String 值当作一系列 bits 处理: 可以设置和清除单独的 bits, 数出所有设为 1 的 bits 的数量, 找到最前的被设为 1 或 0 的 bit, 等等

#### ■ hyperloglogs

- 被用于估计一个 set 中元素数量的概率性的数据结构
- PFADD key element [element ...]

### ○ 命令

- KYES pattern 列出所有匹配的key
  - 生产环境慎用, 该操作会导致服务器阻塞  
使用SCAN替代
- EXISTS mykey 查询key是否存在, 返回1/0
- DEL mykey 删除key, 返回1/0
- TYPE mykey 返回key对应的存储类型
  - string
  - hash
  - list
  - set
  - zset
- EXPIRE mykey secondVal 设置key有效期, 精度可以使用毫秒或秒, 默认: s
  - 设置key5s 超时 expire key 5
- TTL mykey 获取key剩余有效时间
- key的自动创建和删除
  - 推入元素之前创建空的 list, 或者在 list 没有元素时删除它。在 list 为空时删除 key, 并在用户试图添加元素 (比如通过 LPUSH) 而键不存在时创建空 list, 是 Redis 的职责。
  - 三条规则来概括
    - 当我们向一个聚合数据类型中添加元素时, 如果目标键不存在, 就在添加元素前创建空的聚合数据类型。

- 当我们从聚合数据类型中移除元素时，如果值仍然是空的，键自动被销毁。
- 对一个空的 key 调用一个只读的命令，比如 LLEN（返回 list 的长度），或者一个删除元素的命令，将总是产生同样的结果。该结果和对一个空的聚合类型做同个操作的结果是一样的。
- SCAN
  - SCAN 命令用于迭代当前数据库中的key集合。  
SSCAN 命令用于迭代SET集合中的元素。  
HSCAN 命令用于迭代Hash类型中的键值对。  
ZSCAN 命令用于迭代SortSet集合中的元素和元素对应的分值
  - SCAN 命令用于迭代当前数据库中的key集合。
    - SCAN cursor [MATCH pattern] [COUNT count]
    - SCAN命令是一个基于游标的迭代器。这意味着命令每次被调用都需要使用上一次这个调用返回的游标作为该次调用的游标参数，以此来延续之前的迭代过程  
当SCAN命令的游标参数被设置为 0 时，服务器将开始一次新的迭代，而当服务器向用户返回值为 0 的游标时，表示迭代已结束。
    - SCAN命令的返回值 是一个包含两个元素的数组，第一个数组元素是用于进行下一次迭代的新游标，而第二个数组元素则是一个数组，这个数组中包含了所有被迭代的元素。  
在第二次调用 SCAN 命令时，命令返回了游标 0，这表示迭代已经结束，整个数据集已经被完整遍历过了。  
full iteration：以 0 作为游标开始一次新的迭代，一直调用 SCAN 命令，直到命令返回游标 0，我们称这个过程为一次完整遍历。
    - Scan命令的保证（SCAN命令以及其他增量式迭代命令）
      - 从完整遍历开始直到完整遍历结束期间，一直存在于数据集内的所有元素都会被完整遍历返回；这意味着，如果有一个元素，它从遍历开始直到遍历结束期间都存在于被遍历的数据集当中，那么 SCAN 命令总会在某次迭代中将这个元素返回给用户。
      - 同样，如果一个元素在开始遍历之前被移出集合，并且在遍历开始直到遍历结束期间都没有再加入，那么在遍历返回的元素集中就不会出现该元素。
    - SCAN命令每次执行返回的元素数量
      - SCAN增量式迭代命令并不保证每次执行都返回某个给定数量的元素,甚至可能会返回零个元素，但只要命令返回的游标不是 0，应用程序就不应该将迭代视作结束。
      - 默认COUNT=10，即最大返回10条记录
    - cursor 游标
      - 在开始一个新的迭代时，游标必须为 0。  
增量式迭代命令在执行之后返回的，用于延续迭代过程的游标。
    - SCAN, SSCAN, HSCAN 和 ZSCAN 命令都返回一个包含两个元素的 multi-bulk 回复：回复的第一个元素是字符串表示的无符号 64 位整数（游标），回复的第二个元素是另一个 multi-bulk 回复，包含了本次被迭代的元素。  
SCAN 命令返回的每个元素都是一个key。  
SSCAN 命令返回的每个元素都是一个集合成员。  
HSCAN 命令返回的每个元素都是一个键值对，一个键值对由一个键和一个值组成。  
ZSCAN命令返回的每个元素都是一个有序集合元素，一个有序集合元素由一个成员（member）和一个分值（score）组成。
  - SSCAN 命令用于迭代SET集合中的元素。

- HSCAN 命令用于迭代Hash类型中的键值对。
  - ZSCAN 命令用于迭代SortSet集合中的元素和元素对应的分值
- 其他功能
  - Geospatial 地理空间半径查询
    - GEOADD key longitude latitude member [longitude latitude member ...]
      - 存储在ZSET中
      - 将指定的地理空间位置（纬度、经度、名称）添加到指定的key中。这些数据将会存储到sorted set这样的目的是为了使用GEORADIUS或者GEORADIUSBYMEMBER命令对数据进行半径查询等操作。返回添加到ZSET中的元素数目，不包含更新score的元素
        - redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269 37.502669 "Catania"
        - (integer) 2
        - redis> GEODIST Sicily Palermo Catania
        - "166274.15156960039"
        - redis> GEORADIUS Sicily 15 37 100 km
        - 1) "Catania"
        - redis> GEORADIUS Sicily 15 37 200 km
        - 1) "Palermo"
        - 2) "Catania"
        - redis>
    - GEODIST key member1 member2 [unit]
      - 返回两个给定位置之间的距离。
      - 如果两个位置之间的其中一个不存在，那么命令返回空值。
      - 指定单位的参数 unit 必须是以下单位的其中一个：
      - m 表示单位为米。
      - km 表示单位为千米。
      - mi 表示单位为英里。
      - ft 表示单位为英尺。
      - 如果用户没有显式地指定单位参数，那么 GEODIST 默认使用米作为单位。
      - redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269 37.502669 "Catania"
      - (integer) 2
      - redis> GEODIST Sicily Palermo Catania
      - "166274.15156960039"
      - redis> GEODIST Sicily Palermo Catania km
      - "166.27415156960038"
      - redis> GEODIST Sicily Palermo Catania mi
      - "103.31822459492736"
      - redis> GEODIST Sicily Foo Bar
      - (nil)
      - redis>
      - 计算出的距离会以双精度浮点数的形式被返回。如果给定的位置元素不存在，那么命令返回空值。
    - GEOHASH key member [member ...]

- 返回一个或多个位置元素的 Geohash 表示。  
通常使用表示位置的元素使用不同的技术，使用Geohash位置52点整数编码。由于编码和解码过程中所使用的初始最小和最大坐标不同，编码的编码也不同于标准。此命令返回一个标准的Geohash，在维基百科和 geohash.org网站都有相关描述
  - redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269 37.502669 "Catania"  
(integer) 2  
redis> GEOHASH Sicily Palermo Catania  
1) "sqc8b49rny0"  
2) "sqdtr74hyu0"  
redis>
- GEOPOS key member [member ...]
  - 从key里返回所有给定位置元素的位置（经度和纬度）。
    - redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269 37.502669 "Catania"  
(integer) 2  
redis> GEOPOS Sicily Palermo Catania NonExisting  
1) 1) "13.361389338970184"  
2) "38.115556395496299"  
2) 1) "15.087267458438873"  
2) "37.50266842333162"  
3) (nil)  
redis>
  - array-reply, 具体的:  
GEOPOS 命令返回一个数组，数组中的每个项都由两个元素组成：第一个元素为给定位置元素的经度，而第二个元素则为给定位置元素的纬度。当给定的位置元素不存在时，对应的数组项为空值。
- GEORADIUS key longitude latitude radius m|km|ft|mi [WITHCOORD] [WITHDIST] [WITHHASH] [COUNT count] [ASC|DESC]
  - 以给定的经纬度为中心，返回键包含的位置元素当中，与中心的距离不超过给定最大距离的所有位置元素。  
范围可以使用以下其中一个单位：  
m 表示单位为米。  
km 表示单位为千米。  
mi 表示单位为英里。  
ft 表示单位为英尺。  
在给定以下可选项时，命令会返回额外的信息：  
WITHDIST: 在返回位置元素的同时，将位置元素与中心之间的距离也一并返回。距离的单位和用户给定的范围单位保持一致。  
WITHCOORD: 将位置元素的经度和纬度也一并返回。  
WITHHASH: 以 52 位有符号整数的形式，返回位置元素经过原始 geohash 编码的有序集合分值。这个选项主要用于底层应用或者调试，实际中的作用并不大。  
命令默认返回未排序的位置元素。通过以下两个参数，用户可以指定被返回位置元素的排序方式：  
ASC: 根据中心的位置，按照从近到远的方式返回位置元素。  
DESC: 根据中心的位置，按照从远到近的方式返回位置元素。
  - redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269 37.502669 "Catania"

```
(integer) 2
redis> GEORADIUS Sicily 15 37 200 km WITHDIST
1) 1) "Palermo"
2) "190.4424"
2) 1) "Catania"
2) "56.4413"
redis> GEORADIUS Sicily 15 37 200 km WITHCOORD
1) 1) "Palermo"
2) 1) "13.361389338970184"
2) "38.115556395496299"
2) 1) "Catania"
2) 1) "15.087267458438873"
2) "37.50266842333162"
redis> GEORADIUS Sicily 15 37 200 km WITHDIST WITHCOORD
1) 1) "Palermo"
2) "190.4424"
3) 1) "13.361389338970184"
2) "38.115556395496299"
2) 1) "Catania"
2) "56.4413"
3) 1) "15.087267458438873"
2) "37.50266842333162"
redis>
```

- 返回:

bulk-string-reply, 具体的:

在没有给定任何 WITH 选项的情况下, 命令只会返回一个像 ["New York","Milan","Paris"] 这样的线性 (linear) 列表。

在指定了 WITHCOORD、WITHDIST、WITHHASH 等选项的情况下, 命令返回一个二层嵌套数组, 内层的每个子数组就表示一个元素。

在返回嵌套数组时, 子数组的第一个元素总是位置元素的名字。至于额外的信息, 则会作为子数组的后续元素, 按照以下顺序被返回:

以浮点数格式返回的中心与位置元素之间的距离, 单位与用户指定范围时的单位一致。

geohash 整数。

由两个元素组成的坐标, 分别为经度和纬度。

举个例子, GEORADIUS Sicily 15 37 200 km WITHCOORD WITHDIST 这样的命令返回的每个子数组都是类似以下格式的:

```
["Palermo","190.4424","13.361389338970184","38.115556395496299"]]
```

- GEORADIUSBYMEMBER key member radius m|km|ft|mi [WITHCOORD] [WITHDIST] [WITHHASH] [COUNT count]

- 这个命令和 GEORADIUS 命令一样, 都可以找出位于指定范围内的元素, 但是 GEORADIUSBYMEMBER 的中心点是由给定的位置元素决定的, 而不是像 GEORADIUS 那样, 使用输入的经度和纬度来决定中心点指定成员的位置被用作查询的中心。

关于 GEORADIUSBYMEMBER 命令的更多信息, 请参考 GEORADIUS 命令的文档。

- redis> GEOADD Sicily 13.583333 37.316667 "Agrigento"

```
(integer) 1
```

```
redis> GEOADD Sicily 13.361389 38.115556 "Palermo" 15.087269
37.502669 "Catania"
```

```
(integer) 2
```



```
redis> GEORADIUSBYMEMBER Sicily Agrigento 100 km
1) "Agrigento"
2) "Palermo"
redis>
```

- 可用业务场景: 附近的人, 距离位置
- Lua Scripting LUA脚本
- Transactions 事务
  - MULTI、EXEC、DISCARD 和 WATCH 是 Redis 事务相关的命令
    - 事务是一个单独的隔离操作: 事务中的所有命令都会序列化、按顺序地执行。事务在执行的过程中, 不会被其他客户端发送来的命令请求所打断。
    - 事务是一个原子操作: 事务中的命令要么全部被执行, 要么全部都不执行。
  - EXEC 命令负责触发并执行事务中的所有命令
    - 如果客户端在使用 MULTI 开启了一个事务之后, 却因为断线而没有成功执行 EXEC, 那么事务中的所有命令都不会被执行。
    - 另一方面, 如果客户端成功在开启事务之后执行 EXEC, 那么事务中的所有命令都会被执行。
  - 事务中的错误
    - 事务在执行 EXEC 之前, 入队的命令可能会出错。比如说, 命令可能会产生语法错误(参数数量错误, 参数名错误, 等等), 或者其他更严重的错误, 比如内存不足(如果服务器使用 maxmemory 设置了最大内存限制的话)。
    - 命令可能在 EXEC 调用之后失败。举个例子, 事务中的命令可能处理了错误类型的键, 比如将列表命令用在了字符串键上面, 诸如此类。
      - 对于发生在 EXEC 执行之前的错误, 客户端以前的做法是检查命令入队所得的返回值: 如果命令入队时返回 QUEUED, 那么入队成功; 否则, 就是入队失败。如果有命令在入队时失败, 那么大部分客户端都会停止并取消这个事务。  
不过, 从 Redis 2.6.5 开始, 服务器会对命令入队失败的情况进行记录, 并在客户端调用 EXEC 命令时, 拒绝执行并自动放弃这个事务。
  - MULTI 命令开启一个事务, 总是返回OK
    - MULTI 执行之后, 客户端可以继续向服务器发送任意多条命令, 这些命令不会立即被执行, 而是被放到一个队列中, 当 EXEC命令被调用时, 所有队列中的命令才会被执行。
      - > MULTI  
OK  
> INCR foo  
QUEUED  
> INCR bar  
QUEUED  
> EXEC  
1) (integer) 1  
2) (integer) 1
  - DISCARD 客户端可以清空事务队列, 并放弃执行事务
    - > SET foo 1  
OK  
> MULTI  
OK  
> INCR foo  
QUEUED

```
> DISCARD
OK
> GET foo
"1"
```

- EXEC 执行命令队列中的所有命令，  
返回一个数组，数组中的每个元素都是执行事务中的命令所产生的回复。其中，  
回复元素的先后顺序和命令发送的先后顺序一致。
- WATCH 命令可以为 Redis 事务提供 check-and-set (CAS) 行为。
  - 被 WATCH 的键会被监视，并会发觉这些键是否被改动过了。如果有至少一个被监视的键在 EXEC 执行之前被修改了，那么整个事务都会被取消，EXEC 返回nil-reply来表示事务已经失败。
- Persistence 磁盘持久化
  - RDB持久化方式能够在指定的时间间隔对你的数据进行快照存储。
  - AOF持久化方式记录每次对服务器写的操作,当服务器重启的时候会重新执行这些命令来恢复原始的数据,AOF命令以redis协议追加保存每次写的操作到文件末尾.Redis还能对AOF文件进行后台重写,使得AOF文件的体积不至于过大。
  - 如果你只希望你的数据在服务器运行的时候存在,你也可以不使用任何持久化方式。
  - 你也可以同时开启两种持久化方式,在这种情况下,当redis重启的时候会优先载入AOF文件来恢复原始的数据,因为在通常情况下AOF文件保存的数据集要比RDB文件保存的数据集要完整。
  - 最重要的事情是了解RDB和AOF持久化方式的不同,让我们以RDB持久化方式开始:
    - RDB(Redis Database)  
对数据进行快照存储  
适用于数据备份
      - RDB的优点
        - RDB是一个非常紧凑的文件,它保存了某个时间点得数据集,非常适用于数据集的备份,比如你可以在每个小时报保存一下过去24小时内的数据,同时每天保存过去30天的数据,这样即使出了问题你可以根据需求恢复到不同版本的数据集。
        - RDB是一个紧凑的单一文件,很方便传送到另一个远端数据中心或者亚马逊的S3（可能加密），非常适用于灾难恢复。
        - RDB在保存RDB文件时父进程唯一需要做的就是fork出一个子进程,接下来的工作全部由子进程来做，父进程不需要再做其他IO操作，所以RDB持久化方式可以最大化redis的性能。
        - 与AOF相比,在恢复大的数据集的时候，RDB方式会更快一些。
      - RDB的缺点
        - 如果你希望在redis意外停止工作（例如电源中断）的情况下丢失的数据最少的话，那么RDB不适合你.虽然你可以配置不同的save时间点(例如每隔5分钟并且对数据集有100个写的操作),是Redis要完整的保存整个数据集是一个比较繁重的工作,你通常会每隔5分钟或者更久做一次完整的保存,万一在Redis意外宕机,你可能会丢失几分钟的数据。
        - RDB 需要经常fork子进程来保存数据集到硬盘上,当数据集比较大的时候,fork的过程是非常耗时的,可能会导致Redis在一些毫秒级内不能响应客户端的请求.如果数据集巨大并且CPU性能不是很好的情况下,这种情况会持续1秒,AOF也需要fork,但是你可以调节重写日志文件的频率来提高数据集的耐久度。

- AOF(Append-only file, AOF)

配置: appendonly yes

通过追加命令到文件末尾, 恢复数据时重新执行AOF文件命令来达到数据恢复

- AOF优点

- 使用AOF 会让你的Redis更加耐久: 你可以使用不同的fsync策略: 无fsync,每秒fsync,每次写的时候fsync.使用默认的每秒fsync策略,Redis的性能依然很好(fsync是由后台线程进行处理的,主线程会尽力处理客户端请求),一旦出现故障, 你最多丢失1秒的数据.
    - AOF文件是一个只进行追加的日志文件,所以不需要写入seek,即使由于某些原因(磁盘空间已满, 写的过程中宕机等等)未执行完整的写入命令,你也可使用redis-check-aof工具修复这些问题.
    - Redis 可以在 AOF 文件体积变得过大时, 自动地在后台对 AOF 进行重写: 重写后的新 AOF 文件包含了恢复当前数据集所需的最小命令集合。整个重写操作是绝对安全的, 因为 Redis 在创建新 AOF 文件的过程中, 会继续将命令追加到现有的 AOF 文件里面, 即使重写过程中发生停机, 现有的 AOF 文件也不会丢失。而一旦新 AOF 文件创建完毕, Redis 就会从旧 AOF 文件切换到新 AOF 文件, 并开始对新 AOF 文件进行追加操作。
    - AOF 文件有序地保存了对数据库执行的所有写入操作, 这些写入操作以 Redis 协议的格式保存, 因此 AOF 文件的内容非常容易被读懂, 对文件进行分析 (parse) 也很轻松。导出 (export) AOF 文件也非常简单: 举个例子, 如果你不小心执行了 FLUSHALL 命令, 但只要 AOF 文件未被重写, 那么只要停止服务器, 移除 AOF 文件末尾的 FLUSHALL 命令, 并重启 Redis , 就可以将数据集恢复到 FLUSHALL 执行之前的状态。

- AOF缺点

- 对于相同的数据集来说, AOF 文件的体积通常要大于 RDB 文件的体积。
    - 根据所使用的 fsync 策略, AOF 的速度可能会慢于 RDB 。在一般情况下, 每秒 fsync 的性能依然非常高, 而关闭 fsync 可以让 AOF 的速度和 RDB 一样快, 即使在高负荷之下也是如此。不过在处理巨大的写入载入时, RDB 可以提供更有保证的最大延迟时间 (latency) 。

- fsync3种方式

- 每次有新命令追加到 AOF 文件时就执行一次 fsync : 非常慢, 也非常安全
    - 每秒 fsync 一次: 足够快 (和使用 RDB 持久化差不多), 并且在故障时只会丢失 1 秒钟的数据。
    - 从不 fsync : 将数据交给操作系统来处理。更快, 也更不安全的选择。
    - 推荐 (并且也是默认) 的措施为每秒 fsync 一次, 这种 fsync 策略可以兼顾速度和安全性。

- LRU缓存回收算法

- Redis支持LRU, LFU(4.0+)

- Maxmemory配置指令

- maxmemory配置指令用于配置Redis存储数据时指定限制的内存大小。

- 例如为了配置内存限制为100mb, 以下的指令可以放在redis.conf文件中:

maxmemory 100mb

- 设置maxmemory为0代表没有内存限制。对于64位的系统这是个默认值，对于32位的系统默认内存限制为3GB。
  - maxmemory-policy配置 (8种, LRU: 5种, LFU: 2种, 返回错误: 1种)
    - LRU是Least Recently Used的缩写，即最近最久未使用算法
    - LFU: least frequently used 最近最少使用 (最少使用频率) 算法
    - 如果没有键满足回收的前提条件的话，策略volatile-lru, volatile-random以及volatile-ttl就和noeviction 差不多了。
    - noeviction:返回错误当内存限制达到并且客户端尝试执行会让更多内存被使用的命令 (大部分的写入指令，但DEL和几个例外)
    - allkeys-lru: 尝试回收最久没使用的键 (LRU)，使得新添加的数据有空间存放。
    - volatile-lru: 尝试回收最久没使用的键 (LRU)，首先从设置了过期时间的键集合中驱逐最久没有使用的键，使得新添加的数据有空间存放。
    - allkeys-random: 回收随机的键使得新添加的数据有空间存放。
    - volatile-random: 回收随机的键使得新添加的数据有空间存放，但仅限于在过期集合的键。
    - volatile-ttl: 回收在过期集合的键，并且优先回收存活时间 (TTL) 较短的键，使得新添加的数据有空间存放。
    - allkeys-lfu: 从所有键中驱逐使用频率最少的键
    - volatile-lfu: 从所有配置了过期时间的键中驱逐使用频率最少的键
  - Replication 主从复制
    - Redis为异步复制
    - 一个Master可以有多个slave
    - Redis 复制在 master 侧是非阻塞的。
    - 复制在 slave 侧大部分也是非阻塞的。
    - 配置基本的 Redis 复制功能是很简单的：只需要将以下内容加进 slave 的配置文件：  
slaveof 192.168.1.1 6379
  - 高可用性: Sentinel(哨兵)，集群(Cluster)
    - redis集群模式为：多主从+数据分片
- o redis-cli
- 连接redis服务: redis-cli.exe -h 127.0.0.1 -p 6379
  - select [index] : 切换到指定的数据库，数据库索引号 index 用数字值指定，以 0 作为起始索引值，最新版本支持0-15个数据库。集群环境只有1个库
  - auth [password]: 简单密码认证
  - time: 返回当前服务器时间，unix时间戳
  - client list: 返回所有连接到服务器的客户端和统计数据
  - client kill [ip:port]: 关闭地址为ip:port的客户端
  - save: 将数据同步保存到磁盘
  - bgsave: 将数据异步保存到磁盘
  - lastsave: 返回上次成功将数据保存到磁盘的Unix时间戳
  - shutdown [nosave|save]: 保存数据并关闭服务
  - slaveof [master\_ip:port]: 设置主从
- o 集群
- 高可用性: Sentinel(哨兵，主从)，集群(Cluster)
- o 配置
- Redis 的配置文件位于 Redis 安装目录下，文件名为 redis.conf(Windows 名为 redis.windows.conf)。
  - 通过 CONFIG 命令查看或设置配置项
- o 调优

- 数据持久化配置
    - 高可用配置(集群)
  - 生产问题(缓存穿透)
    - 缓存雪崩
      - 原因:
        1. 多个热点Key在高并发下同时失效，导致并发请求落到后端数据库，导致后端数据库压力甚至宕机
        2. Redis 数据库崩溃，导致缓存服务失效
          - 解决方案:
        3. 不同热点Key 设置不同的超时时间，避免同时失效
        4. 设置热点Key数据永不过期
        5. Redis服务高可用: 设置主从及分布式部署
        6. 服务限流及降级，控制服务请求流量，避免压垮后端数据库
          - 缓存穿透
            - 原因:
              7. 请求先从Redis中未查询到缓存，到数据库中也未查询到缓存，高并发下，导致大量无效请求
                - 解决方案:
              8. 在代码层面，控制无效Redis Key
              9. 查询数据库后，将空结果缓存到Redis中，防止同一Key 频繁请求数据库
                - 缓存击穿
                  - 原因:
                    10. 高并发请求下，某个Redis Key过期瞬间，大量请求落到后端数据库
                      - 解决方案:
                    11. 设置热点Redis Key永不过期
                    12. 缓存失效后，使用分布式锁来控制查询，只有1个请求来查询数据和更新缓存，其他请求等待或直接返回无数据
- MongoDB  
(Document)
  - 分布式文件存储的数据库:
 

MongoDB 是一个基于分布式文件存储的数据库。由 C++ 语言编写。旨在为 WEB 应用提供可扩展的高性能数据存储解决方案。

MongoDB 是一个介于关系数据库和非关系数据库之间的产品，是非关系数据库当中功能最丰富，最像关系数据库的。

存储特性与内部原理:

<https://blog.csdn.net/gaozhigang/article/details/79241044>

Mongo操作数据时，会先在内存中处理，然后通过预写事务日志刷新到磁盘，最终数据是保存在磁盘文件上的，Index 会存在内存中(部分索引)，当内存不足时，会从文件里查询。

wiredTiger 引擎: 3.0+ ,使用预写事务日志journal，来确保数据的持久性。

MMAPv1引擎(<3.2默认): mongodb原生的存储引擎，比较简单，直接使用系统级的内存映射文件机制 (memory mapped files)

    - 所有的write请求都基于“文档级别”的lock，因此多个客户端可以同时更新一个collection中的不同文档，这种更细颗粒度的lock，可以支撑更高的读写负载和并发量。

wiredTiger每隔60秒（默认）或者待写入的数据达到2G时，mongodb将对journal文件提交一个checkpoint（检测点，将内存中的数据变更flush到磁盘中的数据文件中，并做一个标记点，表示此前的数据表示已经持久存储在了数据文件中，此后的数据变更存在于内存和journal日志）。

对于write操作，首先被持久写入journal，然后在内存中保存变更数据，条件满足后提交一个新的检测点，即检测点之前的数据只是在journal中持久存储，但并没有在mongodb的数据文件中持久化，延迟持久化可以提升磁盘效率，如果在提交checkpoint之前，mongodb异常退出，此后再次启动可以根据journal日志恢复数据。journal日志默认每个100毫秒同步磁盘一次，每100M数据生成一个新的journal文件，journal默认使用了snappy压缩，检测点创建后，此前的journal日志即可清除。

- MMAPV1在lock的并发级别上，支持到collection级别，所以对于同一个collection同时只能有一个write操作执行。

MMAPv1 为了确保数据的安全性，mongodb将所有的变更操作写入journal并间歇性的持久到磁盘上，对于实际数据文件将延迟写入，和wiredTiger一样journal也是用于数据恢复。

所有的记录在磁盘上连续存储，当一个document尺寸变大时，mongodb需要重新分配一个新的记录（旧的record标记删除，新的record在文件尾部重新分配空间），这意味着mongodb同时还需要更新此文档的索引（指向新的record的offset），与in-place update相比，将消耗更多的时间和存储开支。

由此可见，如果你的mongodb的使用场景中有大量的这种update，那么或许MMAPv1引擎并不太适合，同时也反映出如果document没有索引，是无法保证document在read中的顺序（即自然顺序）。

- 一个mongodb中可以建立多个数据库。

MongoDB的默认数据库为"db"，该数据库存储在data目录中。

MongoDB的单个实例可以容纳多个独立的数据库，每一个都有自己的集合和权限，不同的数据库也放置在不同的文件中。

- 数据库名(DB)

- 数据库名可以是满足以下条件的任意UTF-8字符串
- 不能是空字符串 ("")。
- 不得含有' ' (空格)、.、\$、/、\和\0 (空字符)。
- 应全部小写。
- 最多64字节。
- 保留名
  - admin: 从权限的角度来看，这是"root"数据库。要是将一个用户添加到这个数据库，这个用户自动继承所有数据库的权限。一些特定的服务器端命令也只能从这个数据库运行，比如列出所有的数据库或者关闭服务器。
  - local: 这个数据永远不会被复制，可以用来存储限于本地单台服务器的任意集合
  - config: 当Mongo用于分片设置时，config数据库在内部使用，用于保存分片的相关信息。

- 文档(Document)

- 文档是一组键值(key-value)对(即 BSON)。MongoDB 的文档不需要设置相同的字段，并且相同的字段不需要相同的数据类型
- 文档中的键/值对是有序的。
- 文档中的值不仅可以是在双引号里面的字符串，还可以是其他几种数据类型（甚至可以是整个嵌入的文档）。
- MongoDB区分类型和大小写。
- MongoDB的文档不能有重复的键。
- 文档的键是字符串。除了少数例外情况，键可以使用任意UTF-8字符。
- 命名规范
  - 键不能含有\0 (空字符)。这个字符用来表示键的结尾。
  - .和\$有特别的意义，只有在特定环境下才能使用。

- 以下划线"\_"开头的键是保留的(不是严格要求的)。
- 集合(Collection)
  - 命名规范
    - 集合名不能是空字符串""。
    - 集合名不能含有\0字符 (空字符)，这个字符表示集合名的结尾。
    - 集合名不能以"system."开头，这是为系统集合保留的前缀。
    - 用户创建的集合名字不能含有保留字符。有些驱动程序的确支持在集合名里面包含，这是因为某些系统生成的集合中包含该字符。除非你要访问这种系统创建的集合，否则千万不要在名字里出现\$。
  - capped collections(固定大小集合)
    - db.createCollection("mycoll", {capped:true, size:100000})
- 元数据
  - 数据库的信息是存储在集合中。它们使用了系统的命名空间： dbname.system.\*
    - dbname.system.namespaces 列出所有名字空间。
    - dbname.system.indexes 列出所有索引。
    - dbname.system.profile 包含数据库概要(profile)信息。
    - dbname.system.users 列出所有可访问数据库的用户。
    - dbname.local.sources 包含复制对端 (slave) 的服务器信息和状态。
- 基础数据类型
  - 数据存储为一个文档BSON(Binary JSON)，类似于JSON对象
  - ObjectId : 对象 ID。用于创建文档的 ID。可以很快的去生成和排序，包含 12 bytes。  
 ObjectId 是一个12字节 BSON 类型数据，有以下格式：  
 前4个字节表示时间戳  
 接下来的3个字节是机器标识码  
 紧接的两个字节由进程id组成 (PID)  
 最后三个字节是随机数。
    - 生成ObjectId(): ObjectId() | new ObjectId()
      - > var newObject = ObjectId()
      - > newObject.getTimestamp()
      - ISODate("2017-11-25T07:21:10Z")
    - 获取ObjectId字符串: ObjectId().str
      - > newObject.str
      - 5a1919e63df83ce79df8b38f
  - String : 字符串，UTF-8编码
  - Integer: 整型数值
  - Boolean: 布尔型
  - Double: 双精度浮点型
  - Min/Max keys: 将一个值与BSON元素的最低值和最高值相对比
  - Array: 用于将数据或列表或多个值存储为1个键
  - Timestamp: 时间戳，记录文件档修改或添加的具体时间
  - Object: 用于内嵌文档
  - Null: 用于创建空值
  - Symbol: 符号。基本上等同于String，不同的是，一般用于采用特殊符号类型的语言
  - Date: 日期。用于UNIX时间格式存储当前时间或日期。
    - > var mydate1 = new Date() //格林尼治时间
    - > mydate1

```
ISODate("2018-03-04T14:58:51.233Z")
```

```
> typeof mydate1
```

```
object
```

```
■ > var mydate2 = ISODate() //格林尼治时间
```

```
> mydate2
```

```
ISODate("2018-03-04T15:00:45.479Z")
```

```
> typeof mydate2
```

```
object
```

```
■ > var mydate1str = mydate1.toString()
```

```
> mydate1str
```

```
Sun Mar 04 2018 14:58:51 GMT+0000 (UTC)
```

```
> typeof mydate1str
```

```
string
```

```
■ > Date()
```

```
Sun Mar 04 2018 15:02:59 GMT+0000 (UTC)
```

```
■ Binary Data: 二进制数据
```

```
■ Code: 代码类型, 用于在文档中存储JavaScript代码
```

```
■ Regular expression: 正则表达式
```

#### ○ 连接

```
■ mongodb://[username:password@]host1[:port1][,host2[:port2],...[,hostN[:portN]]]  
[/[database][?options]]
```

```
■ https://docs.mongodb.com/manual/reference/connection-string/
```

```
■ mongodb:// 这是固定的格式, 必须要指定。
```

```
■ username:password@ 可选项, 如果设置, 在连接数据库服务器之后, 驱动都会  
尝试登陆这个数据库
```

```
■ host1 必须的指定至少一个host, host1 是这个URI唯一要填写的。它指定了要连接  
服务器的地址。如果要连接复制集, 请指定多个主机地址。
```

```
■ portX 可选的指定端口, 如果不填, 默认为27017
```

```
■ /database 如果指定username:password@, 连接并验证登陆指定数据库。若不  
指定, 默认打开 test 数据库。
```

```
■ ?options 是连接选项。如果不使用/database, 则前面需要加上/。所有连接选项  
都是键值对name=value, 键值对之间通过&或; (分号) 隔开
```

```
■ replicaSet=name 验证replica set的名称。 Impliesconnect=replicaSet.
```

```
■ slaveOk=true|false
```

```
■ true:在connect=direct模式下, 驱动会连接第一台机器, 即使这台服务  
器不是主。在connect=replicaSet模式下, 驱动会发送所有的写请求到  
主并且把读取操作分布在其他从服务器。
```

```
■ false: 在 connect=direct模式下, 驱动会自动找寻主服务器. 在  
connect=replicaSet 模式下, 驱动仅仅连接主服务器, 并且所有的读写  
命令都连接到主服务器。
```

```
■ tls=true|false
```

```
ssl=true|false
```

```
■ 是否使用SSL连接
```

```
■ safe=true|false
```

```
■ true: 在执行更新操作之后, 驱动都会发送getLastError命令来确保更新  
成功。(还要参考 wtimeoutMS).
```

```
■ false: 在每次更新之后, 驱动不会发送getLastError来确保更新成功。
```



- w=n 驱动添加 { w : n } 到getLastError命令. 应用于safe=true。
- wtimeoutMS=ms 驱动添加 { wtimeout : ms } 到 getlasterror 命令. 应用于 safe=true.
- fsync=true|false
  - true: 驱动添加 { fsync : true } 到 getlasterror 命令.应用于 safe=true.
  - false: 驱动不会添加到getLastError命令中。
- journal=true|false 如果设置为 true, 同步到 journal (在提交到数据库前写入到实体中). 应用于 safe=true
- connectTimeoutMS=ms 可以打开连接的时间。
- socketTimeoutMS=ms 发送和接受sockets的时间。
- 连接实例
  - 使用用户名fred, 密码foobar登录localhost的baz数据库。  
mongodb://fred:foobar@localhost/baz
  - 连接 replica pair, 服务器1为example1.com服务器2为example2。  
mongodb://example1.com:27017,example2.com:27017
  - 连接 replica set 三台服务器, 写入操作应用在主服务器 并且分布查询到从服务器。  
mongodb://host1,host2,host3/?slaveOk=true
    - 连接 replica set 三台服务器, 写入操作应用在主服务器 并且分布查询到从服务器。
  - 直接连接第一个服务器, 无论是replica set一部分或者主服务器或者从服务器。  
mongodb://host1,host2,host3/?connect=direct;slaveOk=true
  - 以安全模式连接到replica set, 并且等待至少两个复制服务器成功写入, 超时时间设置为2秒。  
mongodb://host1,host2,host3/?safe=true;w=2;wtimeoutMS=2000

#### ○ 基本概念


- database: 数据库
- collection: 数据库表/集合
- document: 数据记录行/文档
- field: 数据字段/域
- index: 索引
- primary key: 主键, Mongo自动将\_id(ObjectId)字段设置为主键

#### ○ 用法

- DB相关命令
  - show dbs 显示所有数据的列表。
    - > show dbs  
local 0.078GB  
test 0.078GB  
>
  - db 显示当前数据库对象或集合。
  - use <database\_name> 连接到一个指定的数据库(若不存在则在插入文档后创建)。
    - > use local  
switched to db local  
> db

local

>

- db.dropDatabase() 删除数据库
- Collection相关命令
  - show collections | show tables | db.getCollectionNames() 查看所有集合
  - db.createCollection(name, options) 创建集合  
集合可以不显示创建，插入文档时自动创建
    - capped 布尔 (可选)  
如果为 true，则创建固定集合。固定集合是指有着固定大小的集合，当达到最大值时，它会覆盖最早的文档。  
当该值为 true 时，必须指定 size 参数。
      - 判断是否为固定集合  
db.cappedLogCollection.isCapped()
      - 返序返回集合内容  
db.cappedLogCollection.find().sort({\$natural:-1})
    - autoIndexId 布尔 (可选)  
从3.2版开始不推荐使用。  
如为 false，不在 \_id 字段创建索引。默认为 true。
    - size 数值 (可选)  
为固定集合指定一个最大值，以千字节计 (KB) 。  
如果 capped 为 true，也需要指定该字段。
    - max 数值 (可选) 指定固定集合中包含文档的最大数量。
  - db.<collection\_name>.drop() 删除集合
- Document相关命令
  - 使用.操作符操作 嵌入式文档中的字段(Object 或Array)
  - db.col.find().pretty() 格式化显示结果
  - db.collection.find(query, projection) 查询文档
    - query：可选，使用查询操作符指定查询条件
    - projection：可选，使用投影操作符指定返回的键。查询时返回文档中所有键值，只需省略该参数即可（默认省略）。1返回,0:不返回
      -  db.collection.findOne({}, {field1:1 | 0})
  - db.collection.findOne(query, projection) 查询文档，只返回一个文档
    - where 条件操作符
      - 等于 {}
        - db.col.find({"by":"菜鸟教程"}).pretty()
      - 小于 {:\$lt:}
        - db.col.find({"likes":{\$lt:50}}).pretty()
      - 小于或等于 {:\$lte:}
        - db.col.find({"likes":{\$lte:50}}).pretty()
      - 大于 {:\$gt:}
        - db.col.find({"likes":{\$gt:50}}).pretty()
      - 大于或等于 {:\$gte:}
        - db.col.find({"likes":{\$gte:50}}).pretty()
      - 不等于 {:\$ne:}

- db.col.find({"likes":{\$ne:50}}).pretty()
  - or
    - \$or: [ {key1: value1}, {key2:value2}]
    - \$type 操作符
- <https://docs.mongodb.com/manual/reference/operator/query/type/index.html>  
类型, 值核对表
- db.col.find({"title" : {\$type : 2}})  
或  
db.col.find({"title" : {\$type : 'string'}})
  - limit 读取指定数量的数据记录
    - db.collection.find().limit(NUMBER)
  - skip 跳过指定数量的数据
    - db.collection.find().limit(NUMBER).skip(NUMBER)
  - sort 排序( 1:升序, -1:降序)
    - db.collection.find().sort({field1:1,field2:-1})
  - \$in \$nin
  - 正则表达式 \$regex
    - 注意事项
  - 正则表达式中使用变量。一定要使用eval将组合的字符串进行转换, 不能直接将字符串拼接后传入给表达式。否则没有报错信息, 只是结果为空!  
var name=eval("/" + 变量值key + "/i");
  - 模糊查询包含title关键词, 且不区分大小写  
title:eval("/"+title+"/i") // 等同于 title:{\$regex:title,\$Option:"\$i"}
    - regex操作符
      - {:\$regex:/pattern/, \$options:""}
      - {:\$regex:'pattern', \$options:""}
      - {:\$regex:/pattern/}
    - 正则表达式对象
      - {:/pattern/}
- options  
包括i, m, x以及s四个选项
  - i 忽略大小写, {{\$regex/pattern/i}}, 设置i选项后, 模式中的字母会进行大小写不敏感匹配。
  - m 多行匹配模式, {{\$regex/pattern/,\$options:'m'}}, m选项会更改^和\$元字符的默认行为, 分别使用与行的开头和结尾匹配, 而不是与输入字符串的开头和结尾匹配。
  - x 忽略非转义的空白字符, {{\$regex:/pattern/,\$options:'m'}}, 设置x选项后, 正则表达式中的非转义的空白字符将被忽略, 同时井号(#)被解释为注释的开头注, 只能显式位于option选项中。
  - s 单行匹配模式{{\$regex:/pattern/,\$options:'s'}}, 设置s选项后, 会改变模式中的点号(.)元字符的默认行为, 它会匹配所有字符, 包括换行符(\n), 只能显式位于option选项中。
    - 使用\$regex操作符时, 需要注意下面几个问题:
    - i, m, x, s可以组合使用, 例如:{name:{\$regex:/jk/,\$options:"si"}}
    - 在设置索引的字段上进行正则匹配可以提高查询速度, 而且当正则表达式使用的是前缀表达式时, 查询速度会进一步提高, 例如:{name:{\$regex:/^joe/}}
    - \$regex与正则表达式对象的区别
      - 在\$in操作符中只能使用正则表达式对象, 例如:{name:{\$in:[/^joe/i,/^jack/]}}
      - 在使用隐式的\$and操作符中, 只能使用\$regex, 例如:{name:{\$regex:/^jo/i,\$nin:['john']}}

- 当option选项中包含X或S选项时, 只能使用\$regex, 例如{name: {\$regex:/m.line/, \$options:"si"}}

- db.<collection\_name>.insert(document) 插入文档
- db.<collection\_name>.insertOne() 插入1个文档
- db.collection.insertOne(  
,  
{  
writeConcern:  
}  
)
  - 可以将数据定义为一个变量, 再插入
- > document=({title: 'var obj', likes: 100});  
> { "title" : "var obj", "likes" : 100 }  
db.col.insert(document)  
WriteResult({ "nInserted" : 1 })
  - db.collection.insertMany() 插入多个文档
- db.collection.insertMany(  
[ <document 1> , <document 2> , ... ],  
{  
writeConcern: ,  
ordered:  
}  
)
  - document: 要写入的文档。
  - writeConcern: 写入策略, 默认为 1, 即要求确认写操作, 0 是不要求。
  - ordered: 指定是否按顺序写入, 默认 true, 按顺序写入。

- 更新文档

```
db.collection.update(  
,  
,  
{  
  upsert: ,  
  multi: ,  
  writeConcern:  
}  
)
```

- db.collection.updateOne(, , )  
db.collection.updateMany(, , )  
db.collection.replaceOne(, , )
  - query : update的查询条件, 类似sql update查询内where后面的。
  - update : update的对象和一些更新的操作符 (如\$, \$inc...) 等, 也可以理解为sql update查询内set后面的
    - filde更新操作符
    - \$currentDate 设置字段为当前日期
- db.customers.updateOne(  
{ id: 1 },  
{  
 \$currentDate: {  
 lastModified: true,  
 "cancellation.date": { \$type: "timestamp" }  
}

```

    },
    $set: {
      "cancellation.reason": "user request",
      status: "D"
    }
  }
)

- $inc 为字段增加指定值
  - { $inc: { :, :, ... } }
- $min 当指定值小于当前值时, 将值更新为指定值
  - { $min: { :, ... } }
- $max 当指定值大于当前值时, 将值更新为指定值
  - { $max: { :, ... } }
- $mul 当前值与指定值相乘, 若当前字段不存在, 则用0相乘
> - { "id" : 1, "item" : "ABC", "price" : NumberDecimal("10.99"), "qty" : 25 }
db.products.update(
  { id: 1 },
  { $mul: { price: NumberDecimal("1.25"), qty: 2 } }
)
{ "id" : 1, "item" : "ABC", "price" : NumberDecimal("13.7375"), "qty" : 50 }
- $rename 重命名字段名
  - { $rename: { :, :, ... } }
- $setOnInsert 当upsert=1时为插入数据时, 插入指定数据
- db.collection.update(
,
{ $setOnInsert: { :, ... } },
{ upsert: true }
)
- $unset 删除字段
  - { $unset: { :, "", ... } }
- array 更新操作符
- $ 更新指定数组中的指定元素为新元素, 而无需指定具体索引
$.val
$.field:val
- db.collection.update(
{ : value ... },
{ : { ".$" : value } }
)
- >db.students.insert([
{ "id" : 1, "grades" : [ 85, 80, 80 ] },
{ "id" : 2, "grades" : [ 88, 90, 92 ] },
{ "_id" : 3, "grades" : [ 85, 100, 90 ] }
])
db.students.updateOne(
{ _id: 1, grades: 80 },
{ $set: { "grades.$" : 82 } }
> )
{ "id" : 1, "grades" : [ 85, 82, 80 ] }
{ "id" : 2, "grades" : [ 88, 90, 92 ] }
{ "_id" : 3, "grades" : [ 85, 100, 90 ] }

```

- db.collection.update(
  - { <query selector> },
  - { : { "array.\$.field" : value } }
 )
  - {
    - \_id: 4,
    - grades: [

```
{ grade: 80, mean: 75, std: 8 },
{ grade: 85, mean: 90, std: 5 },
{ grade: 85, mean: 85, std: 8 }
```

```
]
>}
db.students.updateOne(
  { _id: 4, "grades.grade": 85 },
  { $set: { "grades.$std" : 6 } }
>)
{
  "_id" : 4,
  "grades" : [
    { "grade" : 80, "mean" : 75, "std" : 8 },
    { "grade" : 85, "mean" : 90, "std" : 6 },
    { "grade" : 85, "mean" : 85, "std" : 8 }
  ]
}
```

- db.students.updateOne(
  - {

```
_id: 5,
grades: { $elemMatch: { grade: { $lte: 90 }, mean: { $gt: 80 } } }
```

```
},
{ $set: { "grades.$std" : 6 } }
)
- $[] 更新字段中的所有元素
.$[]:val
- { : { ".$[]" : value } }
```

```
db.collection.updateMany(
  { <query conditions> },
  { : { ".$[]" : value } }
)
- $[] 按条件更新数组中的部分数据
```

- db.collection.updateMany(
  - { <query conditions> },
  - { : { ".\$[]" : value } },
  - { arrayFilters: [ { : <condition> } ] }
 )
  - { "id" : 1, "grades" : [ 95, 92, 90 ] }
  - { "id" : 2, "grades" : [ 98, 100, 102 ] }

```
>{ "_id" : 3, "grades" : [ 95, 110, 100 ] }
db.students.update(
  {},
  { $set: { "grades.$[element]" : 100 } },
  { multi: true,
```

```
arrayFilters: [ { "element": { $gte: 100 } } ]
```

```
}
>)
```

```
{ "id" : 1, "grades" : [ 95, 92, 90 ] }
```

```
{ "id" : 2, "grades" : [ 98, 100, 100 ] }
```

```
{ "_id" : 3, "grades" : [ 95, 100, 100 ] }
```

- \$addToSet 将元素添加到数组中(若元素中已存在, 则不插入)

- db.collection.update(
 { <query> },
 { \$addToSet: { field: val } }
 )

- db.foo.update(
 { \_id: 1 },
 { \$addToSet: { colors: "c" } }
 )

- \$pop 删除数组中的第一个或最后一个元素

- db.collection.update( { <query> },

{ \$pop: { field: -1 | 1 } } )

- \$pull 删除数组中指定条件的元素, 或完全匹配的1个元素

- { \$pull: { : <value | condition>, : <value | condition>, ... } }

- {
 \_id: 1,
 fruits: [ "apples", "pears", "oranges", "grapes", "bananas" ],
 vegetables: [ "carrots", "celery", "squash", "carrots" ]
 }
 {
 \_id: 2,
 fruits: [ "plums", "kiwis", "oranges", "bananas", "apples" ],
 vegetables: [ "broccoli", "zucchini", "carrots", "onions" ]
 }
 >}
 db.stores.update(

```
{ },
{ $pull: { fruits: { $in: [ "apples", "oranges" ] }, vegetables: "carrots" } },
```

```
{ multi: true }
```

```
>)
```

```
{
  "id": 1,
  "fruits": [ "pears", "grapes", "bananas" ],
  "vegetables": [ "celery", "squash" ]
}
```

```
{
  "id": 2,
  "fruits": [ "plums", "kiwis", "bananas" ],
  "vegetables": [ "broccoli", "zucchini", "onions" ]
}
```

- \$pullAll 从数组中删除多个指定元素
  - { \$pullAll: { : [ , <value2> ... ], ... } }
- \$push 将元素添加到数组中
  - { \$push: { : , ... } }
  - { \$push: { : { : , ... }, ... } }
- 添加多个元素到数组中

```
db.students.update(
  { _id: 5 },
  {
    $push: {
      quizzes: {
        $each: [ { wk: 5, score: 8 }, { wk: 6, score: 7 }, { wk: 7, score: 6 } ],
        $sort: { score: -1 }
      }
    }
  }
)
```

- 修饰符
  - \$each 将多个值添加到数组中，用于\$addToSet,\$push
    - { \$addToSet: { : { \$each: [ , <value2> ... ] } } }
    - { \$push: { : { \$each: [ , <value2> ... ] } } }
  - \$position(NUMBER),与\$each配合使用 将元素插入到指定位置，用于\$push，下标从0开始

- {
  - \$push: {

```
<field>: {
  $each: [ <value1>, <value2>, ... ],
  $position: <num>
}
```

```
}
}
- $slice(NUMBER) 将数组截取为指定数量的大小
, 用于$push
```

- {
  - \$push: {

```
<field>: {
  $each: [ <value1>, <value2>, ... ],
  $slice: <num>
}
```

```
}
}
- $sort 对数组进行排序，与$each配合使用，用于$push
```



- {  
  \$push: {

```
<field>: {  
  $each: [ <value1>, <value2>, ... ],  
  $sort: <sort specification>  
}
```

```
}  
}
```

- upsert: 可选, 这个参数的意思是, 如果不存在update的记录, 是否插入objNew,true为插入, 默认是false, 不插入。

- multi: 可选, mongodb 默认是false,只更新找到的第一条记录, 如果这个参数为true,就把按条件查出来多条记录全部更新。

- writeConcern :可选, 抛出异常的级别

- 实例

- 更新匹配到的所有记录

```
db.col.update({'title':'MongoDB 教程'},{$set:{'title':'MongoDB'}},{multi:true})
```

- 全部更新

```
db.col.update( { "count" : { $gt : 3 } }, { $set : { "test2" : "OK" } },false,true );
```

- 删除文档

```
db.collection.remove(  
,
```

```
{  
  justOne: ,  
  writeConcern:  
}  
)
```

- query: (可选) 删除的文档的条件。

- justOne: (可选) 如果设为 true 或 1, 则只删除一个文档, 如果不设置该参数, 或使用默认值 false, 则删除所有匹配条件的文档。

- writeConcern : (可选) 抛出异常的级别。

- index索引相关命令

- db.collection.createIndex(keys, options) 创建索引

- options

- name string

索引的名称。如果未指定, MongoDB的通过连接索引的字段名和排序顺序生成一个索引名称。

- background Boolean

建索引过程会阻塞其它数据库操作, background可指定以后台方式创建索引, 即增加 "background" 可选参数。 "background" 默认值为false。

- unique Boolean

建立的索引是否唯一。指定为true创建唯一索引。默认值为false。

- sparse Boolean

对文档中不存在的字段数据不启用索引; 这个参数需要特别注意, 如果设置为true的话, 在索引字段中不会查询出不包含对应字段的文档。默认值为 false。

- expireAfterSeconds integer

指定一个以秒为单位的数值, 完成 TTL设定, 设定集合的生存时间。

- v index version

索引的版本号。默认的索引版本取决于mongod创建索引时运行的版本。

- weights document

索引权重值, 数值在 1 到 99,999 之间, 表示该索引相对于其他索引字段的得分权重。

- default\_language string

对于文本索引，该参数决定了停用词及词干和词器的规则的列表。默认为英语

- language\_override string

对于文本索引，该参数指定了包含在文档中的字段名，语言覆盖默认的language，默认值为language。

- db.collection.getIndexes() 查询集合索引
- db.col.dropIndex("indexName") 删除指定索引
- db.collection.dropIndexes() 删除所有索引
- db.collection.totalIndexSize() 查看集合索引大小
- 全文索引

- MongoDB 在 2.6 版本以后是默认开启全文检索的，如果你使用之前的版本，你需要使用以下代码来启用全文检索：

- >db.adminCommand({setParameter:true,textSearchEnabled:true})

或者使用命令：

```
mongo --setParameter textSearchEnabled=true
```

- 创建全文索引
  - db.collection.createIndex({post\_text:"text"})
- 查询全文索引
  - db.collection.find({\$text:{\$search:"runoob"}})

- aggregate聚合操作

- db.collection.aggregate(AGGREGATE\_OPERATION)
- AGGREGATE\_OPERATION

- \$project: 修改输入文档的结构。可以用来重命名、增加或删除域，也可以用于创建计算结果以及嵌套文档。

- 结果中就只剩下title和author2个字段了 (id默认被包含)

```
db.article.aggregate(
```

```
{ $project : {  
  _id : 0 ,  
  title : 1 ,  
  author : 1  
}});
```

- \$match: 用于过滤数据，只输出符合条件的文档。\$match使用MongoDB的标准查询操作。

- \$match用于获取分数大于70小于或等于90记录，然后将符合条件的记录送到下一阶段\$group管道操作符进行处理。

```
db.articles.aggregate( [
```

```
{ $match : { score : { $gt : 70, $lte : 90 } } },  
{ $group: { _id: null,  
  count: { $sum: 1 } } }  
]);
```

- \$limit: 用来限制MongoDB聚合管道返回的文档数。
- \$skip: 在聚合管道中跳过指定数量的文档，并返回余下的文档。
- \$unwind: 将文档中的某一个数组类型字段拆分成多条，每条包含数组中的一个值。
- \$group: 将集合中的文档分组，可用于统计结果。
- \$sort: 将输入文档排序后输出。
- \$geoNear: 输出接近某一地理位置的有序文档。

- MongoDB中的关系

- 1. 嵌入文档
- 2. 类数据维护在1条数据中

```
{  
  "id":ObjectId("52ffc33cd85242f436000001"),  
  "contact": "987654321",  
  "dob": "01-01-1991",
```

```

"name": "Tom Benzamin",
"address": [
  {
    "building": "22 A, Indiana Apt",
    "pincode": 123456,
    "city": "Los Angeles",
    "state": "California"
  },
  {
    "building": "170 A, Acropolis Apt",
    "pincode": 456789,
    "city": "Chicago",
    "state": "Illinois"
  }
]
}

```

### - 2. 引用式关系

- 两次查询，第一次查询用户地址的对象id (ObjectId)，第二次通过查询的id获取用户的详细地址信息。

```

{
  "_id": ObjectId("52ffc33cd85242f436000001"),
  "contact": "987654321",
  "dob": "01-01-1991",
  "name": "Tom Benzamin",
  "address_ids": [
    ObjectId("52ffc4a5d85242602e000000"),
    ObjectId("52ffc4a5d85242602e000001")
  ]
}

```

### - 3. 数据库应用DBRef

- 指定集合，一个文档从多个集合引用文档，我们应该使用 DBRefs。

- 形式：{ \$ref: , \$id: , \$db: }

- \$ref: 集合名称

- \$id: 引用的id

- \$db: 数据库名称，可选参数

- {
 

```

        "_id": ObjectId("53402597d852426020000002"),
        "address": {
          "$ref": "address_home",
          "$id": ObjectId("534009e4d852427820000002"),
          "$db": "runoob"},
        "contact": "987654321",
        "dob": "01-01-1991",
        "name": "Tom Benzamin"
      }
    
```
- > - >var user = db.users.findOne({"name":"Tom Benzamin"})

```

var dbRef = user.address
db[dbRef.$ref].findOne({"_id":(dbRef.$id)})

```

### • 集群 (ReplicaSet副本集)

- 原理

- mongodb的复制至少需要两个节点。其中一个是主节点，负责处理客户端请求，其余的都是从节点，负责复制主节点上的数据。  
mongodb各个节点常见的搭配方式为：一主一从、一主多从。  
主节点记录在其上的所有操作oplog，从节点定期轮询主节点获取这些操作，然后对自己的数据副本执行这些操作，从而保证从节点的数据与主节点一致。  
副本集在主机宕机后，副本会接管主节点成为主节点，不会出现宕机的情况。
- 设置
  - 1. mongod --port "PORT" --dbpath "YOUR\_DB\_DATA\_PATH" --replSet "REPLICA\_SET\_INSTANCE\_NAME"
  - mongod --port 27017 --dbpath "D:\set up\mongodb\data" --replSet rs0
  - 2. 登录client,执行 rs.initiate()来启动一个新的副本集
    - rs.conf()来查看副本集的配置
    - rs.status() 查看副本集状态
    - db.isMaster() 判断服务是否是主节点
  - 3. rs.add()方法来添加副本集的成员  
rs.add(HOST\_NAME:PORT)
    - rs.add("mongodb1.net:27017")
- 集群(Shard分片)

- <https://www.runoob.com/mongodb/mongodb-sharding.html>

- 监控

- mongostat
  - mongostat是mongodb自带的状态检测工具，在命令行下使用。它会间隔固定时间获取mongodb的当前运行状态，并输出。
- mongotop
  - mongotop也是mongodb下的一个内置工具，mongotop提供了一个方法，用来跟踪一个MongoDB的实例，查看哪些大量的时间花费在读取和写入数据。mongotop提供每个集合的水平的统计数据。默认情况下，mongotop返回值的每一秒。

## ● 查询分析

- MongoDB 查询分析可以确保我们所建立的索引是否有效，是查询语句性能分析的重要工具。  
MongoDB 查询分析常用函数有：explain() 和 hint()。

- explain()
  - db.collection.find().explain()
    - {
      - "cursor" : "BtreeCursor gender\_1\_user\_name\_1",
      - "isMultiKey" : false,
      - "n" : 1,
      - "nscannedObjects" : 0,
      - "nscanned" : 1,
      - "nscannedObjectsAllPlans" : 0,
      - "nscannedAllPlans" : 1,
      - "scanAndOrder" : false,
      - "indexOnly" : true,
      - "nYields" : 0,
      - "nChunkSkips" : 0,
      - "millis" : 0,
      - "indexBounds" : {
      - "gender" : [

```
[
  "M",
  "M"
],
"user_name": [
  [
    {
      "$minElement": 1
    },
    {
      "$maxElement": 1
    }
  ]
]
}
```

#### ■ 返回字段说明

- indexOnly: 字段为 true，表示我们使用了索引。
- cursor: 因为这个查询使用了索引，MongoDB 中索引存储在B树结构中，所以这也是使用了 BtreeCursor 类型的游标。如果没有使用索引，游标的类型是 BasicCursor。这个键还会给出你所使用的索引的名称，你通过这个名称可以查看当前数据库下的system.indexes集合（系统自动创建，由于存储索引信息，这个稍微会提到）来得到索引的详细信息。
- n: 当前查询返回的文档数量。
- nscanned/nscannedObjects: 表明当前这次查询一共扫描了集合中多少个文档，我们的目的是，让这个数值和返回文档的数量越接近越好。
- millis: 当前查询所需时间，毫秒数。

#### ■ indexBounds: 当前查询具体使用的索引。

#### ■ hint()

使用 hint 来强制 MongoDB 使用一个指定的索引

- db.collection.find().hint({field1:1,field2:1})

#### ○ Mongo中的原子操作

- mongodb不支持事务，但是mongodb提供了许多原子操作，比如文档的保存，修改，删除等，都是原子操作

#### ■ 原子操作命令

- \$set
- \$unset
- \$inc
- \$push
- \$pushAll
- \$pull
- \$addToSet
- \$pop
- \$rename
- \$bit

#### ○ 高级索引

- 查询时，无需指定索引顺序，mongo会自动优化

- 索引数组字段
  - 在数组中创建索引，默认对数组中的每个字段依次建立索引
- 索引子文档字段
  - db.collection.createIndex({"obj1.field1":1,"obj2.field2":1})
- 索引限制
  - 额外开销：每个索引占用一定存储空间在，进行插入，更新和删除操作时也需要对索引进行操作
  - 内存(RAM)使用
    - 由于索引是存储在内存(RAM)中，应该确保该索引的大小不超过内存的限制
  - 查询限制(索引不生效)
    - 正则表达式及非操作符，如 \$nin, \$not, 等。
    - 算术运算符，如 \$mod, 等。
    - \$where 子句
  - 最大范围
    - 集合中索引不能超过64个
    - 索引名的长度不能超过128个字符
    - 一个复合索引最多可以有31个字段
- MapReduce
  - > - Map-Reduce是一种计算模型，简单的说就是将大批量的工作（数据）分解（MAP）执行，然后再将结果合并成最终结果（REDUCE）。

● 基本语法：

```
db.collection.mapReduce(
  function() {emit(key,value);}, //map 函数
  function(key,values) {return reduceFunction}, //reduce 函数
  {
    out: collection,
    query: document,
    sort: document,
```

```
limit: number
}
```

><.find()查询返回结果>

- 使用 MapReduce 要实现两个函数 Map 函数和 Reduce 函数,Map 函数调用 emit(key, value), 遍历 collection 中所有的记录, 将 key 与 value 传递给 Reduce 函数进行处理。

Map 函数必须调用 emit(key, value) 返回键值对。

- 实例：

```
db.posts.mapReduce(
  function() { emit(this.user_name,1); },
  function(key, values) {return Array.sum(values)},
  {
    query:{status:"active"},
    out:"post_total"
  }
)
```

- 参数说明：

- map：映射函数 (生成键值对序列,作为 reduce 函数参数)。
- reduce 统计函数，reduce函数的任务就是将key-values变成key-value把values数组变成一

个单一的值value。

- out 统计结果存放集合 (不指定则使用临时集合,在客户端断开后自动删除)。

- 创建临时集合

out: { inline: 1 }

- 设置了 {inline:1} 将不会创建集合, 整个 Map/Reduce 的操作将会在内存中进行。

注意, 这个选项只有在结果集单个文档大小在16MB限制范围内时才有效。

- query 一个筛选条件, 只有满足条件的文档才会调用map函数。(query. limit, sort可以随意组合)
- sort 和limit结合的sort排序参数 (也是在发往map函数前给文档排序), 可以优化分组机制
- limit 发往map函数的文档数量的上限 (要是没有limit, 单独使用sort的用处不大)
- 以下实例在集合 orders 中查找 status:"A" 的数据, 并根据 cust\_id 来分组, 并计算 amount 的总和。

- 返回值说明

```
- {  
  "result": "post_total",  
  "timeMillis": 23,  
  "counts": {  
    "input": 5,  
    "emit": 5,  
    "reduce": 1,  
    "output": 2  
  },  
  "ok": 1  
}
```

- result: 储存结果的collection的名字,这是个临时集合, MapReduce的连接关闭后自动就被删除了。

- timeMillis: 执行花费的时间, 毫秒为单位
- input: 满足条件被发送到map函数的文档个数
- emit: 在map函数中emit被调用的次数, 也就是所有集合中的数据总量
- ouput: 结果集合中的文档个数 (count对调试非常有帮助)
- ok: 是否成功, 成功为1
- err: 如果失败, 这里可以有失败原因, 不过从经验上来看, 原因比较模糊, 作用不大

- 自增

- Mongo没有自增功能, 但可使用javascript函数实现自增序列

- >function getNextSequenceValue(sequenceName){

var sequenceDocument = db.counters.findAndModify(  
 {

```
    query:{_id: sequenceName },  
    update: {$inc:{sequence_value:1}},  
    "new":true  
  });
```

return sequenceDocument.sequence\_value;  
}

```
db.products.insert({  
  "_id":getNextSequenceValue("productid"),  
  "product_name":"Samsung S3",  
  "category":"mobiles"})
```

- 调优

- Memcache
- Neo4j  
(Graph)
- DB区别与选型

## RDMS

- Oracle
  - 基础数据类型
  - 用法
  - 集群
  - 配置
  - 索引
  - 调优
- Mysql
  - 基础数据类型
  - 用法
  - 集群
  - 索引
  - 配置
  - 调优
- PostgreSQL
  - 基础数据类型
  - 用法
  - 集群
  - 索引
  - 配置
  - 调优
- DB区别与选型

## LDAP

## GraphDB

- Neo4j

## Spring

---

### Spring基础

- Spring核心概念
  - 架构分析
  - IOC概念
  - AOP概念
- SpringIOC(源码)
  - IOC容器初始化流程
    - 加载解析XML流程
    - 解析BeanDefinition流程
      - 创建BeanDefinition流程
      - PropertyValue属性解析流程



- 注册BeanDefinition流程
- Bean创建流程
  - Bean实例化流程
    - 构造器反射创建对象
  - 属性填充流程(DI)
    - StringTypeValue属性处理流程
      - 类型转换
    - RuntimeBeanReference属性处理流程
  - Bean初始化流程
    - BeanPostProcessor执行流程
    - AOP产生动态代理对象流程
- IOC模块BeanFactory
  - 解析BeanFacotry继承体系
  - 解析BeanDefinition继承体系
  - 解析ApplicationContext继承体系
- ApplicationContext
  - 循环依赖问题
- SpringAOP
  - AOP核心概念
  - AOP源码
- SpringTx(事务)
- Spring常见面试问题解析
- SpringMVC
  - SpringMVC执行流程解析
  - 六大组件介绍
    - MVCE大组件分析(DispatcherServlet. Handler、 View)
    - 其他三大组件分析(HandlerMapping. HandlerAdapter, ViewResolver)
  - 注解解析
    - @Controller
    - ...
  - SpringMVC源码解析
    - DispatcherServlet流程
      - 初始化流程
      - 访问处理流程
      - 拦截器处理流程
    - RequestMappingHandlerMapping工作流程
      - 初始化流程
      - 处理流程
    - RequestMappingHandlerAdapter工作流程
      - 初始化流程
      - 处理流程
      - 参数绑定流程
        - 类型转换
      - 返回值处理流程

- SpEL

## SpringCloud

- SpringBoot
- SpringCloud架构
  - SpringCloud架构结构
  - Gateway
  - 授权中心
  - 认证中心
  - 配置中心
  - 消息中心
  - 通用服务
- SpringCloud相关技术
  - Zookeeper
  - Eureka
  - Feign(Ribbon+Hystrix)
- 注解解析
- 常用包
- 定时任务
  - ElasticJob
  - Scheduling

如何在SpringCloud中只有1个实例执行定时任务:  
使用  
`org.springframework.cloud.client.discovery.DiscoveryClientgetInstances(serviceName);`  
获取服务注册列表中各服务IP, 若当前服务IP为最小/最大时执行任务

  - Quartz

## Spring工具包

### 分页

### ORM

---

## Mybatis

- 架构流程解析
- Configuration、SqlSessionFactory、SqlSession、SqlSource、SqlNode、TypeHandler解析
- 配置文件解析
- SqlSession执行流程
- 四大组件
  - Executor
  - StatementHandler
  - ParameterHandler
  - ResultSetHandler
- 自定义插件
- 注解解析
- findField方式查询

- 实现工具

## Hibernate

- 架构解析
- 用法解析
- 注解解析

## 容器/Server中间件/分布式应用程序协调服务

---

### 分布式应用程序协调服务

- Zookeeper
  - Zookeeper是一个开源的分布式协调服务,为Hadoop的子项目  
是为应用提供一致性的软件, 功能包括: 配置维护、域名服务、分布式同步、组服务等  
主要包含文件系统+监听通知服务
  - Zookeeper包含一个简单的原语集, 提供Java和C的接口, 提供分布式独享锁, 选举, 队列接口。
  - Zookeeper基本运行流程
    - 1、选举Leader
    - 2. 同步数据
    - 3. 选举Leader过程中算法有很多, 但要达到选举标准是一致的
    - 4.Leader要具有最高的执行ID,类似root权限
    - 5. 集群中大多数的机器得到响应并接受选出的Leader
  - Zookeeper维持一个类似文件系统的结构, 每个子目录被称为一个znode,在znode下可以创建, 删除子znode, znode可以存储数据
  - zk对于传输数据有一个 1MB 的大小限制
  - 集群配置
    - Zookeeper集群通过 conf/zoo.cfg 来配置
- 1. 生成配置文件  
配置不同的zoo.cfg, 同时配置好clientPort,server列表
- 2. 标识Server ID  
在zoo.cfg中配置的dataDir文件夹下, 创建myid文件, 并填入对应的Server ID值, 即, 1-255
- 3. 启动时, 分别zkServer start conf/zoo-1.cfg 等启动
  - zoo.cfg
    - clientPort 配置启动端口, 默认2181
      - dataDir 数据保存目录, 默认含日志文件
      - dataLogDir 数据日志保存目录
      - server.A=B:C:D[:E]  
如: server.1=127.0.0.1:2881:3881  
A为1-255数字, 表示server对应的ID  
B为server ip,默认2181  
C为server 运行端口, 即对外提供服务的端口  
D为Leader选举的端口, 用来执行选举时的通信端口,默认2182  
E为peerType配置, 用于配置peerType=observer观察者模式, 默认为participant参与选举模式
    - 集群下数据同步原理

- 集群服务数必须为奇数个( $n \geq 3$ ), 因为需要选举, 过半机制
- 集群中所有服务数据相同
- 集群角色
  - org.apache.zookeeper.server.quorum
  - .QuorumPeer.LearnerType
    - Leader(领导者)
      - LearnerType为PARTICIPANT
    - Follower(跟随者)
      - 参与投票, 提供读服务
      - LearnerType为PARTICIPANT
    - Observer(观察者)
      - LearnerType为OBSERVER
      - 不参与投票, 提供读服务
      - 与Follower的区别
        - Follower 参与投票, Observer不参与投票
        - Follower 会得到2个消息, 1个是通过广播得到proposal的内容, 一个是commit消息
        - Observer 只会得到1个commit的INFORM消息
    - 适用场景
      - 网络状态不稳定的情况下, 使用Observer模式, 提高读服务节点量
      - Zookeeper的集群读写负载较高的情况下, 使用Observer模式, 提高读服务性能, 而不影响写服务
      - 客户端多, 跨机房跨区域
    - 使用Observer模式的好处:
 可以扩展读请求服务, 接收更多的请求流量, 而不会牺牲写操作的吞吐量。因为写操作的吞吐量取决于quorum的size, 增加更多的server投票, quorum会增大, 降低写操作吞吐量
    - 使用observer模式的zoo.cfg配置:
 peerType=observer  
server.1:localhost:2181:3181:observer
  - 集群中至少需要有1个PARTICIPANT
- Zookeeper选举和同步
  - 一个 ZooKeeper 集群同一时刻只会有一个 Leader, 其他都是 Follower 或 Observer。ZooKeeper 配置很简单, 每个节点的配置文件(zoo.cfg)都是一样的, 只有 myid 文件不一样。myid 的值必须是 zoo.cfg中server.{数值} 的{数值}部分。
  - Zookeeper集群Leader选举
    - org.apache.zookeeper.server.quorum
    - .QuorumPeer#startLeaderElection
    - org.apache.zookeeper.server.quorum.QuorumPeerConfig
      - LeaderElection
      - AuthFastLeaderElection
      - FastLeaderElection (最新默认)
        - org.apache.zookeeper.server.quorum.FastLeaderElection
    - 流程简述

- 目前有5台服务器，每台服务器均没有数据，它们的编号分别是1,2,3,4,5,按编号依次启动，它们的选择过程如下：
- 服务器1启动，给自己投票，然后发投票信息，由于其它机器还没有启动所以它收不到反馈信息，服务器1的状态一直属于Looking(选举状态)。
- 服务器2启动，给自己投票，同时与之前启动的服务器1交换结果，由于服务器2的编号大所以服务器2胜出，但此时投票数没有大于半数，所以两个服务器的状态依然是LOOKING。
- 服务器3启动，给自己投票，同时与之前启动的服务器1,2交换信息，由于服务器3的编号最大所以服务器3胜出，此时投票数正好大于半数，所以服务器3成为领导者，服务器1,2成为小弟。
- 服务器4启动，给自己投票，同时与之前启动的服务器1,2,3交换信息，尽管服务器4的编号大，但之前服务器3已经胜出，所以服务器4只能成为小弟。
- 服务器5启动，后面的逻辑同服务器4成为小弟。
- 选举操作使用UDP广播消息  
DatagramSocket udpSocket
- 概念
  - ServerId: 服务器Id  
如有3台服务器，编号为1,2,3
  - Zxid: 数据ID，64位，前32位为Epoch,后32位为全局序列  
服务器中存放的最大数据ID,值越大数据越新，在选举中数据越新权重越大  
Zookeeper是要用zxid保证顺序一致性
  - Epoch:逻辑时钟，(纪元,时代,新世纪)  
或者叫投票的次数，同一轮投票过程中的逻辑时钟值是相同的。  
每投完一次票这个数值就会增加，然后与收到的其他服务器返回的投票信息中的值比较，做出不同的判断
  - Server状态: 选举状态  
LOOKING: 竞选状态  
FOLLOWING: 随从状态，同步leader状态，参与投票  
OBSERVING: 观察状态，同步leader状态，不参与投票  
LEADING: 领导者状态
- 选举消息内容
  - 投票完成后，需要将所有投票信息发送给集群中的所有机器，包含: ServerId, Zxid, Epoch 逻辑时钟, 选举状态
- 选举流程
  1. 服务启动时，读取当前Server数据及配置信息(dataDir下):  
读取Zxid, currentEpoch, acceptedEpoch  
然后创建选举线程，开始选举
  2. 发送投票信息  
首先，每个Server第一轮都会投票给自己，申请自己为Leader  
投票信息包括: 所选举Leader的Serverid, Zxid, Epoch. Epoch 会随着选举轮数增加而增加
    - 3. 接收投票信息
      - 若服务器B接收服务器A的信息(A为选举状态LOOKING)
        - 1. 判断Epoch逻辑时钟
          - a) 若收到的逻辑时钟Epoch大于当前Server的逻辑时钟。  
首先更新本Server逻辑时钟Epoch，同时清空本轮逻辑时钟收集到的其他server的

选举数据。

然后判断是否需要更新当前自己选举的Leader Serverid.

判断规则rule judging: 保存的Zxid最大值和Leader Serverid 来进行判断。先判断Zxid, Zxid大者胜出, 然后判断Leader Serverid, Leader Serverid大者胜出。

然后将自身的选举结果(Leader Serverid,Zxid,Epoch)广播给其他Server

- b) 若收到的逻辑时钟Epoch小于当前Server的逻辑时钟

说明对方Server在一个相对较早的Epoch中, 这时, 只需要将自己的状态数据(Leader Serverid, Zxid,Epoch)广播给其他Server

- c) 若收到的逻辑时钟Epoch等于当前Server的逻辑时钟

根据rule judging来选举Leader,再讲自己选举结果广播给其他Server

- 2. 其次, 判断服务器是不是已经收集到了所有的选举状态:

- 若是, 根据选举结果设置自己的角色 (FOLLOWING,LEADING) ,然后退出选举状态。

- 若没有收到所有服务器的选举状态, 则判断选举过程中最新选举的Leader是不是得到超过半数以上的服务器支持, 若是, 则尝试200ms之内接受一次数据, 若没有新数据到达, 则说明所有服务器已经默认当前结果, 然后设置自己的角色, 退出选举状态; 反之, 继续选举。

- 服务器A处在其他状态(FOLLOWING,LEADING)

- a) 逻辑时钟Epoch等于当前Server的逻辑时钟, 将该数据保存到recvset。此时的集群已经处于LEADING状态, 说明此时的集群已经选出结果。

- 若此时当前Server宣称自己为Leader, 则判断是否有半数以上的服务器选举它, 如果是, 则当前Server为LEADING状态, 反之为FOLLOWING状态, 然后退出选举。

- b) 否则, 这是一条于当前逻辑时钟不符合的消息, 说明在另一个选举中已经有了选举结构,

- 于是将改选举结果加入到outofelection集合中, 再根据outofelection来判断是否可以结束选举, 如果可以, 保存逻辑时钟, 设置选举状态, 退出选举

- Zookeeper的同步过程

- Leader

- leader需要告知其他服务器当前的最新数据, 即最大zxid是什么, 此时leader会构建一个NEWLEADER的数据包, 包括当前最大的zxid, 发送给follower或者observer, 此时leader会启动一个leanerHandler的线程来处理所有follower的同步请求, 同时阻塞主线程,

- 只有半数以上的folower同步完毕之后, leader才成为真正的leader, 退出选举同步过程。

- Follower

- 首先与leader建立连接, 如果连接超时失败, 则重新进入选举状态选举leader, 如果连接成功, 则会将自己的最新zxid封装为FOLLOWERINFO发送给leader

- 首先会尝试与leader建立连接,这里有一个机制,如果一定时间内没有连接上,就报错退出,重新回到选举状态.

- 其次在函数learner::registerWithLeader中发送FOLLOWERINFO封包,该封包中带上自己的最大数据id,也就是会告知leader本机保存的最大数据id.

- 最后,根据前面对LeaderHandler的分析,leader会根据不同的情况发送DIFF,UPTODATE,TRUNC,SNAP,依次进行处理就是了,此时follower跟leader的数据也就同步上了.

- 由于leader端发送的最后一个封包是UPTODATE,因此在接收到这个封包之后follower结束同步数据过程,发送ACK封包回复leader.

- 同步算法

- 直接差异化同步 (DIFF同步)

- 仅回滚同步, 即删除多余的事务日志 (TRUNC)

- 先回滚再差异化同步 (TRUNC+DIFF)
- 全量同步 (SNAP同步)
- 差异化同步 (DIFF) 条件:  $\text{MinCommittedLog} < \text{peerLastZxid} < \text{MaxCommittedLog}$
- 使用Socket进行数据同步
- 过半机制
  - 在领导者选举的过程中, 如果某台zkServer获得了超过半数的选票, 则此zkServer就可以成为Leader了
- 节点读写服务分工
  - ZooKeeper 集群的所有机器通过一个 Leader 选举过程来选定一台被称为『Leader』的机器, Leader服务器为客户端提供读和写服务。
  - Follower 和 Observer 都能提供读服务, 不能提供写服务。两者唯一的区别在于, Observer机器不参与 Leader 选举过程, 也不参与写操作的『过半写成功』策略, 因此 Observer 可以在不影响写性能的情况下提升集群的读性能。
  - 一个Zookeeper集群 ( $N \geq 3$ ,  $N$ 为奇数), 那么只有一个Leader (通过FastLeaderElection选主策略选取), 所有的写操作 (客户端请求Leader或Follower的写操作) 都由Leader统一处理, Follower虽然对外提供读写, 但写操作会提交到Leader, 由Leader和Follower共同保证同一个Follower请求的顺序性, Leader会为每个请求生成一个zxid (高32位是epoch, 用来标识leader选举周期, 每次一个leader被选出来, 都会有一个新的epoch, 标识当前属于哪个leader的统治时期, 低32位用于递增计数)

- zookeeper 如何保证半数提交后剩下的节点上最新的数据

- zookeeper 的leader和follower的prepare和commit时, 只要半数的节点通过就算同意, leader就会commit, 那么剩下的半数节点的数据如何同步到最新的呢?

- 剩下的节点, 会进行版本比对, 发现版本不一致的话, 会更新节点的数据。

- Session

- Session 是指客户端会话, 在讲解客户端会话之前, 我们先来了解下客户端连接。在ZooKeeper 中, 一个客户端连接是指客户端和 ZooKeeper 服务器之间的TCP长连接。

- ZooKeeper 对外的服务端口默认是2181, 客户端启动时, 首先会与服务器建立一个TCP连接, 从第一次连接建立开始, 客户端会话的生命周期也开始了, 通过这个连接, 客户端能够通过心跳检测和服务器保持有效的会话, 也能够向 ZooKeeper 服务器发送请求并接受响应, 同时还能够通过该连接接收来自服务器的 Watch 事件通知。

- Session 的 SessionTimeout 值用来设置一个客户端会话的超时时间。当由于服务器 压力太大、网络故障或是客户端主动断开连接等各种原因导致客户端连接断开时, 只要在 SessionTimeout 规定的时间内能够重新连接上集群中任意一台服务器, 那么之前创建的会话 仍然有效。

- Zookeeper会话生命周期: CONNECTING, CONNECTED, CLOSE

- 在会话到期时, 群集将删除该会话拥有的任何/所有短暂节点, 并立即通知任何/所有连接的客户端该更改 (任何监听这些znode的客户端)

- Zookeeper节点命名

- 任何unicode字符都可以在受以下约束限制的路径中使用:
    - 空字符 (\u0000) 不能是路径名的一部分。 (这会导致C绑定出现问题。)
    - 无法使用以下字符, 因为它们无法正常显示或以令人困惑的方式呈现: \u0001 - \u0019 和 \u007F - \u009F。
    - 不允许使用以下字符: \ud800 - \uF8FFF, \uFFFF0 - \uFFFFF。
    - “.” character可以用作另一个名称的一部分, 但是“.” 并且“...”不能单独用于表示沿路径的节点, 因为ZooKeeper不使用相对路径。以下内容无效: “/ a / b ./ c”或“/a/b/./c”。

- . 或 ..不能单独作为节点名使用

- 令牌“zookeeper”被保留。

- znode类型
- PERSISTENT-持久化目录节点
  - 客户端与Zookeeper断开连接后，节点依然存在
    - TTL - TTL节点
 

创建持久化节点时，可以设置接待你的TTL(毫秒)。如果节点在TTL内未修改，且没有子节点，则会被服务删除。

默认为禁用。
    - PERSISTENT\_SEQUENTIAL-持久化顺序标号目录节点
 

客户端与Zookeeper断开连接后，节点依然存在，只是Zookeeper给该节点进行%10d的顺序编号；

重新连接后，创建的顺序节点会继续按序号增加
- EPHEMERAL-临时目录节点
  - 客户端与Zookeeper断开连接后，节点被删除
- EPHEMERAL\_SEQUENTIAL-临时顺序编号目录节点
  - 客户端与Zookeeper断开连接后，节点被删除，只是Zookeeper给该节点名称进行顺序编号
- znode操作
  - <https://zookeeper.apache.org/doc/current/zookeeperProgrammers.html#Container+Nodes>
    - zkClient -server ip:port[ip2:port2...][/rootpath] 【连接ZkServer】
  - 可设置多个server连接地址，若第一个地址连接失败，则会尝试之后的地址
  - /rootpath 设置连接后使用的根节点，连接后的所有操作都基于该节点
  - create [-s] [-e] path data acl 【创建节点】
    - s 创建节点为顺序节点，创建顺序节点时，会在节点名后自动添加顺序号，编号格式为【%010d】，不足的数字位已0填充，如0000000001，最大值为 2147483647
    - e 创建节点为临时节点，节点默认为持久化节点
  - path 创建的节点路径,按/分割层级，有多层级时，父接待那必须存在才能创建；临时节点不能有子节点；
  - acl 节点的访问控制列表，控制权限，权限有5种：CREATE,READ,WRITE,DELETE,ADMIN，简写为crwda
  - 身份认证有4中方式:
  - world :默认方式，不限制访问
  - auth: 已通过认证的用户访问
  - digest: 用户名密码认证
  - ip:使用ip认证访问
    - 创建节点: create /kangspace.org "kangspace domain:kangspace.org"
  - world:anyone:crwda
  - 且节点权限为world,不限制访问
  - > Created /kangspace.org
    - 创建持久有序节点列表:
  - create /kangspace.org/psnode "SAVED PERSISTENT\_SEQUENTIAL NODES"
  - > Created /kangspace.org/psnode
  - create -s /kangspace.org/psnode/ps "001"
  - > Created /kangspace.org/psnode/ps0000000001
  - create -s /kangspace.org/psnode/ps "002"
  - > Created /kangspace.org/psnode/ps0000000002
  - create -s /kangspace.org/psnode/ps "003"
  - > Created /kangspace.org/psnode/ps0000000003
  - create -s /kangspace.org/psnode/ps "004"
  - > Created /kangspace.org/psnode/ps0000000004
  - 创建临时有序节点:
  - create /kangspace.org/esnode "SAVED EPHEMERAL\_SEQUENTIAL NODES"



```

> Created /kangspace.org/esnode
create -s -e /kangspace.org/esnode/es "001"
> Created /kangspace.org/esnode/es0000000000
create -s -e /kangspace.org/esnode/es "002"
> Created /kangspace.org/esnode/es0000000001
create -s -e /kangspace.org/esnode/es "003"
> Created /kangspace.org/esnode/es0000000002
create -s -e /kangspace.org/esnode/es "004"
> Created /kangspace.org/esnode/es0000000003

```

- ls path [WATCH] 【获取节点下子节点名称列表】  
(只含一级子节点) ; 可设置监视
  - ls /kangspace.org
    - > [esnode, psnode]
- ls2 path [WATCH] 【获取节点下子节点名称列表】 , 并包含该节点属性信息; 可设置监视
  - ls2 /kangspace.org
    - > [esnode, psnode]
    - cZxid = 0x400457aa6
    - ctime = Wed Jun 24 13:39:42 CST 2020
    - mZxid = 0x400457aa6
    - mtime = Wed Jun 24 13:39:42 CST 2020
    - pZxid = 0x400457ab8
    - cversion = 6
    - dataVersion = 0
    - aclVersion = 0
    - ephemeralOwner = 0x0
    - dataLength = 30
    - numChildren = 2
- get path [WATCH] 【获取节点数据及属性信息】 ;  
可设置监视
  - get /kangspace.org
    - > kangspace domain:kangspace.org
    - cZxid = 0x400457aa6
    - ctime = Wed Jun 24 13:39:42 CST 2020
    - mZxid = 0x400457aa6
    - mtime = Wed Jun 24 13:39:42 CST 2020
    - pZxid = 0x400457ab8
    - cversion = 6
    - dataVersion = 0
    - aclVersion = 0
    - ephemeralOwner = 0x0
    - dataLength = 30
    - numChildren = 2
- set path data [version] 【更新节点数据】  
version为要更新的数据Version , 每次更新后dataVersion+1 ,  
若输入的version不是节点当前version时 , 更新失败。
  - set /kangspace.org/esnode/en "update001"
    - > cZxid = 0x400457ac0
    - ctime = Fri Jun 26 10:37:48 CST 2020
    - mZxid = 0x400457ac5

```
mtime = Fri Jun 26 10:50:20 CST 2020
pZxid = 0x400457ac0
cversion = 0
dataVersion = 1
aclVersion = 0
ephemeralOwner = 0x172b69b32ba0016
dataLength = 9
numChildren = 0
set /kangspace.org/esnode/en "update001" 3
> version No is not valid : /kangspace.org/esnode/en
(此时version值应为1)
```

- delete path [version] 【删除节点】  
version 为要更新的数据Version  
若节点下存在子节点，则需先删除子节点才能删除父节点

- delete /kangspace.org/esnode/es1 0

#### ○ 节点监听通知机制(WATCH)

- 客户端监听它关注的目录节点，当目录节点发生变化(数据改变、被删除、子目录节点增加删除)时，Zookeeper会通知客户端

#### ■ 监听机制特点

- 1. 一次性触发，客户端监听event，再收到一次事件后，不再监听
- 2. 发送给客户端，监听事件异步发送，watcher的通知事件从server到client端是异步的；所以只有客户端监听后才能感知节点的变化；并且所有客户端收到的事件顺序是一致的。
- 3. 被设置监视的数据，Zookeeper有数据监视和子数据监视；  
getData()和exists()设置数据监视，getChildren()设置子节点监视
- 4. 注册watcher，getData，exists，getChildren
- 5. 触发watcher，create delete setData

#### ■ 可对所有读取操作设置监听

命令行操作:

ls path [WATCH]

ls p2 path [WATCH]

get path [WATCH]

Java API:

getData()

getChildren()

exists()

#### ■ 监听的语义

- Created event 创建事件: 通过exists监听
- Deleted event 删除事件: 通过exists,getData,getChildren监听
- Changed event 变更事件: 通过exists,getData 监听
- Child event 子节点事件: 通过getChildren监听

#### ■ watcher中的相关事件(及对应监听)

- 节点操作
  - 事件名称
    - 触发监听
- create("/path")
  - EventType.NodeCreated
    - exists(), getData()

- delete("/path")
  - EventType.NodeDeleted
    - exists(),getData()
- setData("/path")
  - EventType.NodeDataChanged
    - exists(),getData()
- create("/path/child")
  - EventType.NodeChildrenChanged
    - exists(), getData(), getChildren()[针对于父节点监听]
- delete("/path/child")
  - EventType.NodeChildrenChanged
    - exists(), getData()
- Java Zookeeper Api
- 应用
  - <https://segmentfault.com/a/1190000012185866>

- Master选举
  - 适用场景
    - 需要实现主从master-slave模式的应用  
(需考虑网络抖动情况下, 主节点最好不随意波动)
  - 算法
    - 最小编号
      - 假设存在选举节点/kangspace.org/leader/election,  
leader节点/kangspace.org/leader/instance,  
follower节点/kangspace.org/servers

1. 所有Server启动时, 都向选举节点下创建临时顺序节点(-e -s)
  2. 判断当前节点是否是最小编号, 若是, 则设置当前节点为Master, 其他server为slave;  
若不是最小编号, 则监听当前节点的前一个节点的删除事件, 等待下次选举(这里监听前一个节点而不监听父节点, 是为了避免羊群效应, 当监听客户端很多时, 服务器阻塞主线程来通知其他客户端);  
或监听leader节点删除事件, 当leader发生变化时触发重新选举
  3. 将Master节点信息保存到leader节点中  
将follower节点信息保存到follower节点下
    - 实现框架: Apache curator开源框架
- org.apache.curator.framework.recipes.leader.LeaderLatch
- Apache Curator框架提供的第一种Leader选举策略是Leader Latch。
- 这种选举策略, 其核心思想是初始化多个LeaderLatch, 然后在等待几秒钟后, Curator会自动从中选举出Leader。
- Leader Latch选举的本质是连接ZooKeeper, 然后在"/curator/latchPath"路径为每个LeaderLatch创建临时有序节点:
- 在创建临时节点时, org.apache.curator.framework.recipes.leader.LeaderLatch 的 checkLeadership(List<String> children) 方法会将选举路径 (/curator/latchPath) 下面的所有节点按照序列号排序, 如果当前节点的序列号最小, 则将该节点设置为leader。反之则监听比当前节点序列号小一位的节点的状态 (PS: 因为每次都会选择序列号最小的节点为leader, 所以在比当前节点序列号小一位的节点未被删除前, 当前节点是不可能变成leader的)。如果监听的节点被删除, 则会触发重新选举方法——reset()
- Apache Curator框架提供的另一种Leader选举策略是Leader Election。

这种选举策略跟Leader Latch选举策略不同之处在于每个实例都能公平获取领导权，而且当获取领导权的实例在释放领导权之后，该实例还有机会再次获取领导权。另外，选举出来的leader不会一直占有领导权，当 takeLeadership(CuratorFramework client) 方法执行结束之后会自动释放领导权

//创建 CuratorFrameworkImpl实例

```
client = CuratorFrameworkFactory.newClient(SERVER, SESSION_TIMEOUT,
CONNECTION_TIMEOUT, retryPolicy);
```

//创建LeaderSelectorListenerAdapter实例

```
CustomLeaderSelectorListenerAdapter leaderSelectorListener =
    new CustomLeaderSelectorListenerAdapter(client, PATH, "Client #" + i);
```

```
leaderSelectorListener.start();
leaderSelectorListenerList.add(leaderSelectorListener);
```

- 随机竞争分布式锁
  - 假设存在Leader节点/kangspace.org/leader/instance.

4. 服务启动时,向Leader节点尝试创建Master临时节点, 若创建成功, 则表示当前没有Master,该服务为Master;若创建失败, 则表示当前存在Master, 该服务设置为Follower。
5. 监听Master临时节点的删除事件; 当事件触发时, 所有server同时竞争分布式锁, 得到锁的服务, 去创建Master临时节点, 升级为Master, 其他服务为follower

- 多数投票

- 同Zookeeper FastLeaderElection

- 数据发布与订阅(配置中心)

- 描述: 多个订阅者对象同时监听同一主题对象, 主题对象状态变化时通知所有订阅者对象更新自身状态。

发布方和订阅方独立封装、独立改变, 当一个对象的改变需要同时改变其他对象, 并且它不知道有多少个对象需要改变时, 可以使用发布订阅模式。(类似观察者模式)

- 配置管理: 指集群中的机器拥有某些配置, 并且这些配置信息可以动态地改变, 那么可以使用发布订阅模式把配置做统一管理。

- 算法:

假设存在配置节点/kangspace.org/config

用于配置管理

6. 配置中心管理服务Manage Server 通过config节点下发配置信息
7. 各服务节点监听config节点的数据变化事件, 来动态更新自身配置

- o 分布式协调服务/通知(注册中心)

- 服务发现: 指集群中的服务上下线做统一管理, 每个服务都可以作为数据的发布方, 向集群注册自己的基本信息, 而让某些监控服务器作为订阅方。

当工作服务器的基本信息改变时, 如服务上下线, 服务角色变更等, 监控服务器可以得到通知并响应这些变化

- 算法:

假设存在server列表节点/kangspace.org/servers

8. 当服务启动时, 在servers节点下创建临时节点, 用于服务注册
9. 各服务节点可以监听servers节点的变化, 来同步各自内存中维护的服务列表

- o 分布式锁

- 分布式锁是在分布式环境下, 保护跨进程, 跨主机, 跨网络的共享资源, 实现互斥访问, 保护一致性

- 算法:

假设存在锁节点/kangspace.org/lock

10. 所有服务都向锁节点下创建临时有序节点, 并获取当前锁节点下的所有节点, 并排序, 排序后各个服务检查自身是否是最小编号节点, 最小编号的服务获取到锁。

11. 没有获取到锁的服务删除自身节点

12. 获取到锁的服务, 在执行完成后删除自身节点

- Apache curator实现了分布式锁

```
final CuratorFramework client =
```

```
CuratorFrameworkFactory.newClient(ZK_ADDRESS, new RetryNTimes(10, 5000));
```

```
client.start();
```

```
final InterProcessMutex mutex = new InterProcessMutex(client, ZK_LOCK_PATH);
```

```
//获取锁
```

```
mutex.acquire(1, TimeUnit.SECONDS)
```

```
//释放锁
```

```
mutex.release();
```

- 分布式消息队列

- Zookeeper可以通过顺序节点实现分布式队列

- 算法:

假设存在队列节点 /kangspace.org/queue

13. 生产者在队列节点上创建持久化顺序节点来存放数据

14. 消费者通过读取顺序节点来取数据

相当于对于列的put,offer操作

15. 消费者监听队列节点子节点变化, 触发消费

相当于队列的take()操作

Apache curator实现了SimpleDistributedQueue

(org.apache.curator.framework.recipes.queue.SimpleDistributedQueue)

- 不建议使用zk作为队列来使用

- zk对于传输数据有一个 1MB 的大小限制。

这就意味着实际中zk节点ZNodes必须设计的很小

但实际中队列通常都存放着数以千计的消息

- 如果有很多大的ZNodes, 那会严重拖慢的zk启动过程。

包括zk节点之间的同步过程

如果正要用zk当队列, 最好去调整initLimit与syncLimit

- 如果一个ZNode过大, 也会导致清理变得困难

Netflix不得不设计一个特殊的机制来处理这个大体积的nodes

- 如果zk中某个node下有数千子节点, 也会严重拖累zk性能

- zk中的数据都会放置在内存中。

- 分布式命名服务

- 类似IDNI的功能,

将系统中各种服务名称, 地址及目录信息保存到Zookeeper,需要的时候读取

- 分布式ID生成,

利用Zookeeper顺序节点的特性生成分布式ID

- 封装ZookeeperApi的第三方框架

- Apache curator: Zookeeper Java client

<http://curator.apache.org/>

- Eureka

- 原理

- Spring Cloud 体系中核心、默认的注册中心组件  
SpringBoot服务式注册中心
- Eureka 分为2部分
  - 服务提供者(Server)
    - 服务注册
      - Eureka Client启动时，会向Eureka Server注册信息，Server存储这些服务信息，  
并有2层缓存机制维护整个注册表
    - 提供注册表
      - 服务消费者在调用服务时，如果 Eureka Client 没有缓存注册表的话，会从 Eureka Server 获取最新的注册表  
/eureka/apps
    - 同步状态
      - Eureka Client 通过注册、心跳机制和 Eureka Server 同步当前客户端的状态。
  - 服务消费者(Client)
    - Eureka Client 是一个 Java 客户端，用于简化与 Eureka Server 的交互。  
Eureka Client 会拉取、更新和缓存 Eureka Server 中的信息  
因此当所有的 Eureka Server 节点都宕掉，服务消费者依然可以使用缓存中的信息找到服务提供者，但是当服务有更改的时候会出现信息不一致
    - Register: 服务注册
      - 当Eureka Client向Eureka Server注册时，会提供自生的元数据，如 application,hostname,ipAddr,等
    - Renew: 服务续约
      - Eureka Client 会每隔 30 秒发送一次心跳来续约。通过续约来告知 Eureka Server 该 Eureka Client 运行正常，没有出现问题。默认情况下，如果 Eureka Server 在 90 秒内没有收到 Eureka Client 的续约，Server 端会将实例从其注册表中删除，此时间可配置，一般情况不建议更改。  
(由客户端配置,服务端会保存不同服务应用的不同配置)  
服务续约任务的调用间隔时间，默认为30秒  
eureka.instance.lease-renewal-interval-in-seconds=30  
服务失效的时间，默认为90秒。  
eureka.instance.lease-expiration-duration-in-seconds=90
    - Eviction: 服务剔除
      - 当 Eureka Client 和 Eureka Server 不再有心跳时，Eureka Server 会将该服务实例从服务注册列表中删除，即服务剔除
    - Cancel: 服务下线
      - Cancel: 服务下线
    - GetRegistry: 获取注册列表信息
      - Eureka Client 从服务器获取注册表信息，并将其缓存在本地。客户端会使用该信息查找其他服务，从而进行远程调用。该注册列表信息定期（每30秒钟）更新一次。每次返回注册列表信息可能与 Eureka Client 的缓存信息不同，Eureka Client 自动处理。  
如果由于某种原因导致注册列表信息不能及时匹配，Eureka Client 则会重新获取整个注册表信息。Eureka Server 缓存注册列表信息，整个注册表以及每个应用程序的信息进行了压缩，压缩内容和没有压缩的内容完全相同。Eureka Client 和 Eureka Server 可以使用 JSON/XML 格式

进行通讯。在默认情况下 Eureka Client 使用压缩 JSON 格式来获取注册列表的信息。

- 获取服务是服务消费者的基础，所以必有两个重要参数需要注意：

#启用服务消费者从注册中心拉取服务列表的功能

eureka.client.fetch-registry=true

#设置服务消费者从注册中心拉取服务列表的间隔

eureka.client.registry-fetch-interval-seconds=30

- Remote Call: 远程调用

- 当 Eureka Client 从注册中心获取到服务提供者信息后，就可以通过 Http 请求调用对应的服务；服务提供者有多个时，Eureka Client 客户端会通过 Ribbon 自动进行负载均衡。

## ■ 自我保护机制

- 默认情况下，如果 Eureka Server 在一定的 90s 内没有接收到某个微服务实例的心跳，会注销该实例。但是在微服务架构下服务之间通常都是跨进程调用，网络通信往往会面临着各种问题，比如微服务状态正常，网络分区故障，导致此实例被注销。

- Eureka Server 在运行期间会去统计心跳失败比例在 15 分钟之内是否低于 85%，如果低于 85%，Eureka Server 即会进入自我保护机制。

- Eureka Server 进入自我保护机制，会出现以下几种情况：

(1 Eureka 不再从注册列表中移除因为长时间没收到心跳而应该过期的服务

(2 Eureka 仍然能够接受新服务的注册和查询请求，但是不会被同步到其它节点上 (即保证当前节点依然可用)

(3 当网络稳定时，当前实例新的注册信息会被同步到其它节点中

Eureka 自我保护机制是为了防止误杀服务而提供的一个机制。当个别客户端出现心跳失联时，则认为是客户端的问题，剔除掉客户端；当 Eureka 捕获到大量的心跳失败时，则认为可能是网络问题，进入自我保护机制；当客户端心跳恢复时，Eureka 会自动退出自我保护机制。

- 通过在 Eureka Server 配置如下参数，开启或者关闭保护机制，生产环境建议打开：

eureka.server.enable-self-preservation=true

#设置保护机制的阈值，默认是0.85

eureka.server.renewal-percent-threshold=0.5

## ■ 集群原理

- 再来看看 Eureka 集群的工作原理。我们假设有三台 Eureka Server 组成的集群，第一台 Eureka Server 在北京机房，另外两台 Eureka Server 在深圳和西安机房。这样三台 Eureka Server 就组建成了一个跨区域的高可用集群，只要三个地方的任意一个机房不出现问题，都不会影响整个架构的稳定性。

- Eureka Server 集群相互之间通过 Replicate 来同步数据，相互之间不区分主节点和从节点，所有的节点都是平等的。

在这种架构中，节点通过彼此互相注册来提高可用性，每个节点需要添加一个或多个有效的 serviceUrl 指向其他节点。

如果某台 Eureka Server 宕机，Eureka Client 的请求会自动切换到新的 Eureka Server 节点。当宕机的服务器重新恢复后，Eureka 会再次将其纳入到服务器集群管理之中。当节点开始接受客户端请求时，所有的操作都会进行节点间复制，将请求复制到其它 Eureka Server 当前所知的所有节点中。

另外 Eureka Server 的同步遵循着一个非常简单的原则：只要有一条边将节点连接，就可以进行信息传播与同步。所以，如果存在多个节点，只需要将节点之间两两连接起来形成通路，那么其它注册中心都可以共享信息。每个 Eureka Server 同时也是 Eureka Client，多个 Eureka Server 之间通过 P2P 的方式完成服务注册表的同步。

Eureka Server 集群之间的状态是采用异步方式同步的，所以不保证节点间的状态一定是一致的，不过基本能保证最终状态是一致的。

#### ■ Eureka 分区

Eureka 提供了 Region 和 Zone 两个概念来进行分区，这两个概念均来自于亚马逊的 AWS

##### ■ Region:

可以理解为地理上的不同区域，比如亚洲地区，中国区或者深圳等等。没有具体大小的限制。根据项目具体的情况，可以自行合理划分 region。

##### ■ Zone:

可以简单理解为 region 内的具体机房，比如说 region 划分为深圳，然后深圳有两个机房，就可以在此 region 之下划分出 zone1、zone2 两个 zone。

#### ■ Eureka 保证 AP:(Availability Partition )

优先保证可用性，提供最终一致性

- Eureka Server 各个节点都是平等的，几个节点挂掉不会影响正常节点的工作，剩余的节点依然可以提供注册和查询服务。而 Eureka Client 在向某个 Eureka 注册时，如果发现连接失败，则会自动切换至其它节点。只要有一台 Eureka Server 还在，就能保证注册服务可用(保证可用性)，只不过查到的信息可能不是最新的(不保证强一致性)。

#### ■ 配置

- 集群配置时需要开启向服务注册和拉取注册表

#向注册中心注册自己

eureka.client.register-with-eureka=true

#拉取注册表

eureka.client.fetch-registry=true

##### ■ Server:

(集群中instance.metadata-map.zone不同,其他配置相同)

//设置实例的zone为zone1

eureka.instance.metadata-map.zone=zone1

//设置client的region为region-test

eureka.client.region=region-east

//设置zone1的服务列表

eureka.client.service-url.zone1= <http://localhost:8761/eureka/>,<http://localhost:8762/eureka/>

//设置zone2的服务列表

eureka.client.service-url.zone2= <http://localhost:8763/eureka/>,<http://localhost:8764/eureka/>

//设置可用zone

eureka.client.availability-zones.region-east= zone1,zone2

##### ■ Client:

instance.metadata-map.zone不同

其他同Server配置

#### ■ Eureka工作流程

- 1. Eureka Server启动成功后，等待服务注册。在启动时，如果配置了集群，集群之间会定时通过Replicate同步注册表，每个Eureka Server都存在完整独立的注册表信息
- Server/Client:  
#注册中心路径，如果有多个eureka server，在这里需要配置其他eureka server的地址，用","进行区分，如"<http://address:8888/eureka/>,<http://address:8887/eureka/>"



- eureka.client.service-url.default-zone=http://\${eureka.instance.hostname}:\${server.port}/eureka
- 2. Eureka Client启动, 根据配置的Eureka Server去注册服务
- 3. Eureka Client每个30s向Eureka Server进行renew 心跳, 证明服务正常
  - Client:
    - #心跳间隔5s, 默认30s。每一个服务配置后, 心跳间隔和心跳超时时间会被保存在server端, 不同服务的心跳频率可能不同, server端会根据保存的配置来分别探活
    - eureka.instance.lease-renewal-interval-in-seconds=5
- 4. 当Eureka Server 90s内没收到Eureka Client的心跳时, 会剔除(eviction)该服务
  - Client:
    - #心跳超时时间10s, 默认90s。从client端最后一次发出心跳后, 达到这个时间没有再次发出心跳, 表示服务不可用, 将它的实例从注册中心移除
    - eureka.instance.lease-expiration-duration-in-seconds=10
- 5. 单位时间内Eureka Server 统计到大量的Eureka Client 没有发送心跳, 则认为可能是网络异常, 进入自我保护机制, 不再剔除心跳超时的客户端
  - Server:
    - #开启注册中心的保护机制, 默认是开启
    - eureka.server.enable-self-preservation=true
    - #设置保护机制的阈值, 默认是0.85
    - eureka.server.renewal-percent-threshold=0.5
- 6. 当Eureka Client 心跳恢复正常后, Eureka Server自动退出
- 7. Eureka Client定时全量或增量从注册中心同步注册表, 并且将注册表信息缓存到本地(第一次为全量, 定时同步时为增量)
- 8. 服务调用时, Eureka Client先从本地缓存中获取服务信息。若获取不到, 则从注册中心拉取刷新注册表, 再同步到本地缓存
  - Client:
    - #向注册中心注册自己
    - eureka.client.register-with-eureka=true
    - #同步注册表
    - eureka.client.fetch-registry=true
- 9. Eureka Client获取到目标服务器信息, 发起请求
- 10. Eureka Client正常退出时向Eureka Server发送取消(cancel) 请求, Eureka Server将实例从实例列表中删除

#### ○ 实例

- 公共依赖
  - 依赖:

```
org.springframework.boot
spring-boot-dependencies
${spring.boot.version}
pom
import
```

```
org.springframework.cloud
spring-cloud-dependencies
${spring.cloud.version}
```

```
pom
import
```

```
org.projectlombok
lombok
${lombok.version}
import
```

- eureka-server

- 依赖:

```
org.springframework.cloud
spring-cloud-context
```

```
org.springframework.cloud
spring-cloud-netflix-eureka-server
```

- 配置:

```
spring.application.name=eureka-server
server.port=8001
#服务地址
eureka.instance.hostname=localhost
eureka.instance.instance-id=${spring.application.name}
eureka.instance.virtual-host-name=${spring.application.name}
eureka.instance.prefer-ip-address=false
eureka.environment=demo
#不向注册中心注册自己
eureka.client.register-with-eureka=false
#取消检索服务
eureka.client.fetch-registry=true
#开启注册中心的保护机制，默认是开启
eureka.server.enable-self-preservation=true
#设置保护机制的阈值，默认是0.85
eureka.server.renewal-percent-threshold=0.5
#注册中心路径，如果有多个eureka server，在这里需要配置其他eureka server
的地址，用","进行区分，如"http://address:8888/eureka,http://address:8887/eureka"
eureka.client.service-url.default-
zone=http://${eureka.instance.hostname}:${server.port}/eureka
```

- 服务启动代码:

```
@SpringBootApplication
@EnableEurekaServer
public class EurekaServerApplication{
```

- eureka-client

- 配置:

```
spring.application.name=eureka-client
server.port=8002
#服务地址
eureka.instance.hostname=eureka-client
```

```
eureka.instance.instance-id=${spring.application.name}
eureka.instance.virtual-host-name=${spring.application.name}
#使用ip作为hostname
eureka.instance.prefer-ip-address=true
eureka.environment=demo
#注册中心路径，表示我们向这个注册中心注册服务，如果向多个注册中心注册，
用“，”进行分隔
eureka.client.serviceUrl.defaultZone=http://localhost:8001/eureka
#心跳间隔5s，默认30s。每一个服务配置后，心跳间隔和心跳超时时间会被保存在server端，不同服务的心跳频率可能不同，server端会根据保存的配置来分别探活
eureka.instance.lease-renewal-interval-in-seconds=5
#心跳超时时间10s，默认90s。从client端最后一次发出心跳后，达到这个时间没有再次发出心跳，表示服务不可用，将它的实例从注册中心移除
eureka.instance.lease-expiration-duration-in-seconds=10
debug=true
```

- 服务启动代码:

```
//@EnableEurekaClient中已包含@EnableDiscoveryClients
@SpringBootApplication
@EnableEurekaClient
@Slf4j
public class EurekaClientApplication {}
```

- 备注

- Eureka 客户端注册列表

<http://192.168.10.150:8001/eureka/apps>

- 若使用ServerName做RestTemplate请求，需在RestTemplate 声明的Bean上添加@LoadBalanced注解

- 与Zookeeper的区别

- 相同点

- 都可以用作服务注册中心
    - 都是分布式服务，都能保证高可用

- 不同点

- Zookeeper是文件目录结构的分布式协调服务  
Eureka 是SpringCloud默认自带SpringBoot注册中心服务
    - Zookeeper 实现的是CP模式，保证服务的强一致性，集群采用主从模式，主服务宕机从服务选举时,不能提供对外服务，主服务提供读写操作，从服务提供读服务；服务间使用长连接，keepalive同步状态；  
Eureka 实现的是AP模式，保证服务的高可用性，各Server节点都是平等的，某个Server节点宕机时不影响其他Server对外提供服务，只要有1个Server节点存活，服务就可用；服务间采用轮询机制同步状态；
    - Zookeeper开源，且支持监听(WATCH)，可用于Master选举，分布式锁，消息队列等；  
Eureka 1.x开源，2.0以后闭源；只用做服务注册，是SpringCloud默认的注册中心，不支持监听(WATCH)

## WebServer

- Nginx
- Tomcat
- Resin

# Docker

- 基础应用
- K8S

## 负载均衡

- LVS
- Nginx
- F5
- Ribbon(框架)
  - Feign = Ribbon + Hystrix
- 选型对比

## 消息中间件

---

### Kafka

- Kafka是最初由Linkedin公司开发，是一个分布式、支持分区的（partition）、多副本的（replica），基于zookeeper协调的分布式流平台(分布式消息系统)  
用Scala语言编写  
<http://kafka.apache.org/intro>
  - <https://www.jianshu.com/p/734cf729d77b>
- 流平台的3个关键功能
  - 发布和订阅记录流，类似于消息队列或企业消息传递系统
  - 以容错的持久化方式存储记录流
  - 处理产生的记录流
- 特性
  - 高吞吐量、低延迟：  
kafka每秒可以处理几十万条消息，它的延迟最低只有几毫秒，每个topic可以分多个partition，  
consumer group 对分区进行consume操作
  - 可扩展性： kafka集群支持热扩展
  - 持久性、可靠性：消息被持久化到本地磁盘，并且支持数据备份，防止数据丢失
  - 容错性：允许集群中的节点失败(若副本数量为n,则允许n-1个节点失败)
  - 高并发：支持数千个客户端同时读写
- 使用场景
  - 消息系统：  
解耦生产者和消费者，缓存消息等
  - 存储系统：  
kafka提供稳定的性能和容错的持久化
  - 日志收集：  
可用Kafka收集各种服务log，通过Kafka以统一服务的方式和开放给各个消费者
  - 流处理：  
Spark streaming 和 storm
  - 用户活动跟踪：  
Kafka经常被用来记录web用户或者app用户的各种活动，如网页浏览、搜索、点击等活动，这些活动被服务器发布到topic中，通过订阅这些topic来做实时监控分析。
  - 运营指标：  
Kafka经常被用来记录运营监控数据。

- 设计思想
  - Kafka Broker Leader的选举:  
Kafka Broker集群使用Zookeeper管理,  
注册成功的Broker为 Kafka Broker Controller  
其他Broker为 Kafka Broker follower
  - Consumer Group: 消费者组  
对于topic ,Consumer可以组成一个组,
- 一个分区partition 中消息只能被同一个组中的一个消费者消费, 其他消费者阻塞;
- 一个分区中消息可以被多个不同组同时消费;  
不同组消费的数据是相同的。
- 多个分区partition 中的消息可以被同一组中相同数量的消费者消费;  
这些消息会分布在不同的分区中, 同一组中的消费者消费不同(分区)的数据;  
一个topic有多(n)个分区时,  
当消费者组中消费者数量小于n时, 则消费者会平均监听这n个分区;  
当消费者组中消费者数量等于n时, 则每个消费者监听一个分区;  
当消费者组中消费者数量大于n时, 则有n个消费者分别监听n个分区, 剩下的消费者阻塞;  
最佳实践: topic partition 数量等于消费者组中消费者的数量  
producer流量增大时, 可扩展分区和消费者
  - Consumer Rebalance的触发条件
    - Consumer增加或删除会触发 Consumer Group的Rebalance
    - Broker的增加或者减少都会触发 Consumer Rebalance
  - Consumer 消费者: Consumer处理partition的message的时候是O(1)顺序读取, 所以需要维护消费的offset ,  
高级API offset存在于Zookeeper中, 低级API offset由自己维护。一般使用高级API
    - Consumer 的delivery guarantee , 默认是读完message先commit再处理message, autocommit默认为true;  
这种情况下如果提交完再处理消息, 消息处理失败的情况下, 若不记录就会出现消息丢失;
  - Delivery Mode 消息分发模式
  - Topic & Partition
    - topic 相当于一个队列, 生产者发送的消费者必须指定topic, kafka会均匀的把数据分不到不同的partition,  
每个partition相当于一个子queue。  
在物理存储上, 每个partition对应一个物理目录, 文件夹命名为[tpicname][partition][序号], 一个topic可以有无数个partition。  
添加新的partition后, 旧partition中的数据不会改变, 新的分区内容为空, 在随后的进入topic的消息会加入到新的分区中
  - Partition Replica (分区副本)  
每个partition可以在其他kafka broker节点上存副本, 以便某个kafka broker节点宕机不会影响kafka集群。  
存Replica是按照kafka broker的顺序存。
    - 消息传送:  
producer先把消息发送到partition leader, 再由leader发送给其他follower
    - 在向Producer发送ACK前需要保证有多少个Replica已经收到该消息:  
根据ack配的个数而定
    - 怎样处理某个Replica不工作的情况:

1. 如果这个不工作的partition replica不在ack列表中，就是producer在发送消息到partition leader上，partition leader向partition follower发送message没有响应而已，这个不会影响整个系统，也不会有什么問題。
2. 如果这个不工作的partition replica在ack列表中的话，producer发送的message的时候会等待这个不工作的partition replca写message成功，但是会等到time out，然后返回失败因为某个ack列表中的partition replica没有响应，此时kafka会自动的把这个部工作的partition replica从ack列表中移除，以后的producer发送message的时候就不会有这个ack列表下的这个部工作的partition replica了。

- 怎样处理Failed Replica恢复回来的情况：

3. 如果这个partition replica之前不在ack列表中，那么启动后重新受Zookeeper管理即可，之后producer发送message的时候，partition leader会继续发送message到这个partition follower上
4. 如果这个partition replica之前在ack列表中，此时重启后，需要把这个partition replica再手动加到ack列表中。（ack列表是手动添加的，出现某个不工作的partition replica的时候自动从ack列表中移除的）

kafka 2.5.0中 broker server上线后，会重新加入rsr列表

- Partition leader与follower

- partition也有leader和follower之分。  
leader是主partition，producer写kafka的时候先写partition leader，再由partition leader push给其他的partition follower。  
partition leader与follower的信息受Zookeeper控制，一旦partition leader所在的broker节点宕机，zookeeper会冲其他的broker的partition follower上选择follower变为partition leader。

- Topic分配partition和partition replica的算法

- (1) 将Broker (size=n) 和待分配的Partition排序。  
(2) 将第i个Partition分配到第  $(i \% n)$  个Broker上。  
(3) 将第i个Partition的第j个Replica分配到第  $((i + j) \% n)$  个Broker上

- 消息投递可靠性

- 发送成功就算成功投递; 不保证消息成功投递到broker
- Master-Slave 模型，只有当Master和Slave都收到消息才算投递成功。  
保证了最高的投递可靠性，降低了性能
- Master确认收到及表示成功投递；

- Partition ack

- 当ack=1，表示producer写partition leader成功后，broker就返回成功，无论其他的partition follower是否写成功
  - broker宕机导致partition的follower和leader切换，会导致丢数据
- 当ack=2，表示producer写partition leader和其他一个follower成功的时候，broker就返回成功，无论其他的partition follower是否写成功
- 当ack=-1[partition的数量]的时候，表示只有producer全部写成功的时候，才算成功，kafka broker才返回成功信息。
- ACK前需要保证有多少备份  
ISR(in-sync Replica)
  - 0: 不等待broker的消息确认 延迟最小，但可能丢数据
  - 1: leader 已经接受了数据的确认消息，Replica异步拉取消息，比较折中
  - -1: ISR列表中的所有Replica都需要返回确认信息
    - min.insync.replicas=1 至少有一个Replica返回成功，否则Producer异常

- Message状态

- kafka中，消息的状态保存在consumer中，broker只记录offset值。

- Message持久化

- Kafka会把消息持久化到本地本间系统，并且保持O(1)极高的效率。  
Kafka写入和读取都是按顺序操作的。  
可达到单机每秒10w数据
- Message有效期
  - Kafka会长久保留其中的消息，以便Consumer可以多次消费。  
可配置。
- Kafka吞吐量
  - kafka的高吞吐量体现在读写上，分布式并发的读写都非常快；  
写的性能体现在以O(1)的时间复杂度进行顺序写入；  
读的性能体现在以O(1)的时间复杂度进行顺序读取；  
增加topic的分区partition，和consumer group中的consumer，可提高并发性能
- Kafka delivery guarantee(message传送保证)
  - At most once , 最多1次，消息可能会丢失，不会重复传输
  - At least once , 最少1次，消息不会丢，可能会传送多次
  - Exactly once , 仅传输1次
- 批量发送
  - Kafka支持以消息集合为单位进行批量发送，以提高push效率。
- PUSH-AND-PULL
  - kafka中的Producer和Consumer采用PUSH-AND-PULL模式。  
Producer像broker push消息，  
Consumer从broker pull消息；  
二者异步处理
- Kafka集群中broker之间的关系
  - broker不是主从关系，各个broker在集群中的地位是相同的，可以随意增加或删除broker节点
- 负载均衡
  - kafka 0.8.x 使用metadata api来管理broker的负载
  - kafka0.7.x 使用Zookeeper实现负载
- 同步异步
  - Producer 采用异步PUSH方式，极大提高了Kafka系统的吞吐率  
可通过参数控制同步/异步方式
- 分区机制
  - Kafka的broker支持消息分区partition，Producer可以决定把消息发到哪个partition中；  
在一个partition中，message的顺序就是Producer发送消息的顺序，一个topic可以有多个partition，partition数量可配置  
partition可以设置Replic副本，副本保存在不同的broker上，第一个partition是leader，其他partition是follower；  
消息先写到partition leader上，再由leader push到其他follower上。
- 离线数据装载
  - kafka由于对可扩展的数据持久化的支持，它也适合想Hadoop或者数据仓库中进行数据装载
- 实时数据和离线数据
  - kafka即支持离线数据，也支持实时数据，因为kafka的message持久化到文件，并可以设置有效期；  
因此可以把kafka作为高效的存储来使用，可以作为离线数据；  
当作为分布式实时消息系统时，大多数情况下还是用于实时处理的。
- 插件支持

- 不少活跃社区已经开发出不少插件来扩展Kafka的功能
- 解耦和异步通信
  - 作为MQ, Producer和Consumer异步处理了消息
- 冗余
  - replica 有多个副本, 保证一个broker节点宕机后不影响整个服务
- 扩展性
  - broker节点可以水平扩展, partition也可以水平增加, partition replica也可以水平增加
- 峰值
  - 在访问量剧增的情况下, kafka水平扩展, 应用仍然需要继续发挥作用
- kafka文件存储机制
  - kafka名词解释
    - Broker: Kafka节点, 一个Kafka节点就是一个broker, 多个broker可以组成一个Kafka集群。
    - Topic: Kafka节点, 一个Kafka节点就是一个broker, 多个broker可以组成一个Kafka集群。
    - Partition: topic物理上的分组, 一个topic可以分为多个partition, 每个partition是一个有序的队列
    - Partition Replica: partition 分区副本, 平均分布在各个broker,可以指定 $\geq 1$ ,  $\leq$  broker数量的副本数, 第一个Replica会作为Leader Partition提供服务, 其他Replica作为follower; 选择follower时需要兼顾一个问题,就是新leader server上所已经承载的partition leader的个数,如果一个server上有过多的partition leader,意味着此server将承受着更多的IO压力.在选举新leader,需要考虑到"负载均衡",partition leader较少的broker将会更有可能成为新的leader.
      - 副本分配算法
        - 将所有N Broker和待分配的i个Partition排序.
        - 将第i个Partition分配到第 $(i \bmod n)$ 个Broker上.
        - 将第i个Partition的第j个副本分配到第 $((i + j) \bmod n)$ 个Broker上.
    - Segment: partition物理上由多个segment组成, 每个Segment存着message信息, 消息存储的文件; 包含.index和.log2部分
    - Producer: 消息的生产者, 想Kafka Leader Partition 发送消息;可以有多个; Producer将消息发送到哪个Broker: Producer发送消息到Broker是由生产者客户端决定的, 生产者客户端会获取所有Broker和Leader Partition信息, 根据partition负载均衡算法(或自己实现kafka.producer.Partitioner接口)利用key和partition 决定消息写入到哪个partition, 然后找到对应的Broker来发送消息 生产者负责选择将哪个记录分配给主题中的哪个分区。可以简单地平衡负载而以循环方式完成此操作, 也可以根据某些语义分区功能(例如基于记录中的某些键)完成此操作
    - Consumer: 订阅topic消费message, consumer作为一个线程来消费 由消费者维护topic partition的消费情况,即offset位置, 存储在Zookeeper和Kafka服务端
    - Consumer Group: 消费者组, 1个partition只能被同一个消费者组中的一个消费者消费, 其他消费者阻塞; 1个partition 可被同步的消费者组中的各1个消费者同时消费; 多个partition可被消费者组中相同数量的消费者消费;



当高并发下，可同时增加partition和消费者提高消息的消费效率；

- 持久化

- 消息会被持久化到系统文件(segment),只有被刷新到磁盘的消息才会被消费

- kafka中文件存储结构

- topic中partition存储分布:

- topic 在磁盘上是按partition保存的,

- partition 是一个文件夹目录;

- 以[topicName]-[partitionIndex]命名

- partitionIndex起始为0

- Partition是一个Queue的结构，每个Partition中的消息都是有序的，生产的消息被不断追加到Partition上，其中的每一个消息都被赋予了一个唯一的offset值。

- partition中文件存储方式:

- 每个partition(目录)相当于一个巨型文件被平均分配到多个大小相等segment(段)数据文件中。但每个段segment file消息数量不一定相等，这种特性方便old segment file快速被删除。

- 每个partition只需要支持顺序读写就行了，segment文件生命周期由服务端配置参数决定。

- partition中segment文件存储结构:

- producer发message到某个topic，message会被均匀的分布到多个partition上（随机或根据用户指定的回调函数进行分布），kafka broker收到message往对应partition的最后一个segment上添加该消息，当某个segment上的消息条数达到配置值或消息发布时间超过阈值时，segment上的消息会被flush到磁盘，只有flush到磁盘上的消息consumer才能消费，segment达到一定的大小后将不会再往该segment写数据，broker会创建新的segment。

- 每个part在内存中对应一个index，记录每个segment中的第一条消息偏移。

- segment file组成：由2大部分组成，分别为index file和data file，此2个文件一一对应，成对出现，后缀".index"和".log"分别表示为segment索引文件、数据文件。

- segment文件命名规则：partition全局的第一个segment从0开始，后续每个segment文件名为上一个全局partition的最大offset(偏移message数)。数值最大为64位long大小，19位数字字符长度，没有数字用0填充。

- 每个segment中存储很多条消息，消息id由其逻辑位置决定，即从消息id可直接定位到消息的存储位置，避免id到位置的额外映射。

- 如:

- 00000000000000000000.index

- 00000000000000000000.log

- 00000000000000000368769.index

- 00000000000000000368769.log

- kafka最佳实践

- Topic: 创建topic时，指定partition数量，最好 $\geq$ broker数，提高kafka效率，和broker使用率;

- 同时指定replica数量，replica最好 $>2$ , $\leq$ broker数量;

- replica=1时，表示只有1份数据; replica不能大于broker数量; replica $\geq 2$ 时,才能保证某个broker宕机时，不影响业务(当所有存replica的broker都宕机后，该partition不能正常工作)

- Kafka 消息存储可设置过期时间，默认7天

- Kafka提供的命令行大部分需要提供 bootstrap-server 或 zookeeper

- Kafka目录及配置,命令说明

- 默认端口: 9092

- 目录及配置

- bin/bin windows Kafka命令所在目录
  - --bootstrap-server or --zookeeper 是命令中必传项
  - kafka-server-start.sh [-daemon] server.properties [--override property=value]\*  
启动Server
    - 如:  
./bin/kafka-server-start.sh config/server.properties &
  - kafka-server-stop.sh 停止集群服务
    - 如:  
./bin/kafka-server-stop.sh
  - kafka-topics.sh kafka topic相关操作
    - --bootstrap-server [String ip:port,...] 指定kafka broker地址, 与zookeeper 其一必填
    - --zookeeper [String ip:port,...] 指定zookeeper地址 与--bootstrap-server其一必填
    - 子主题 11
    - --create 创建
    - --alter 修改
    - --delete 删除
    - --topic [String topicName] 指定topic名称
    - --partitions [int cnt] 指定分区数量, 默认为1  
--create --alter 可指定
    - --replication-factor [int cnt] 指定复制因子, 即副本数量,  
--create 时可指定
    - --describe 查看topic详情, 含partition,rsr , replica等
    - --list 查看topic列表
    - --if-not-exists 当topic不存在时执行  
--create可指定
    - --replica-assignment 手工指定分区副本  
这种方式根据分区号的数值大小按照从小到大的顺序进行排列, 分区与分区之间用逗号"," 隔开, 分区内多个副本用冒号":"隔开。并且在使用 replica-assignment 参数创建主题时不需要原本必备的 partitions 和 replication-factor 这两个参数。  
如:  
--replica-assignment 1:2,3:4,5:6  
topic有三个partition, 其中partition\_0的replica分布在broker1和2上, partition\_1的replica分布在broker3和4上, partition\_2的replica分布在broker5和6上。  
--create , --alter可指定
    - 如:  
> ./bin/kafka-topics.sh --create --bootstrap-server 127.0.0.1:9092 --topic test-topic --partitions 3 --replication-factor 3  
> Created topic test-topic.  
java.lang.IllegalArgumentException: Only one of --bootstrap-server or --zookeeper must be specified  
--create 创建topic  
--alter 修改topic partitions,replication-fator,配置等信息  
--delete 删除partitions  
--replication-factor 复制因子  
--partitions 分区数

```

--topic topic名称
--list 查看kafka topic列表
--describe 查看topic信息
> ./bin/kafka-topics.sh --list --bootstrap-server 127.0.0.1:9092
查看topic信息:
> ./bin/kafka-topics.sh --describe --bootstrap-server localhost:9092 --
topic test-topic
修改Topic分区:
> ./bin/kafka-topics.sh --alter --zookeeper localhost:22181 --topic test-
topic --partitions 2

```

- kafka-console-producer.sh kafka控制台生产者

- 如:
 

```

> ./bin/kafka-console-producer.sh --bootstrap-server 127.0.0.1:9092 --
topic test-topic

```

- kafka-console-consumer.sh kafka控制台消费者

- --group 指定消费者分组
- --from-beginning 从最早位置开始消费数据
- --offset [String offsetId] 指定消费位置id
- --partition [int pid] 指定监听的分区号
- 如:
 

```

> ./bin/kafka-con
sole-consumer.sh --bootstrap-server localhost:9092 --topic test-topic --
from-beginning --group 0

```

- kafka-consumer-groups.sh kafka消费者组操作

- --all-groups 应用到所有消费者组
- --group [String group] 指定要查看的组
  - ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 -
 -describe --offsets --group 1
- --describe 显示消费者组及消费情况  
需指定--all-groups 或--group 使用

```

./bin/kafka-consumer-groups.sh --bootstrap-server
localhost:9092 --describe --all-groups
GROUP      TOPIC      PARTITION  CURRENT-OFFSET LOG-
END-OFFSET LAG      CONSUMER-ID  HOST
CLIENT-ID
1          test-topic  0          85             109          24          -
-          -
TOPIC topic名字
PARTITION 分区id
CURRENT-OFFSET 当前已消费的条数
LOG-END-OFFSET 总条数
LAG 未消费的条数
CONSUMER-ID 消费id
HOST 主机ip
CLIENT-ID 客户端id

```

- --offsets 只显示消费者组分区消费情况  
仅在--bootstrap-server ,--describe下使用  
是--describe的默认子行为

- ```

      ■ ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --describe --all-groups --offsets

```
- `--list` 显示所有消费者组名称
    - ```

        ■ ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --list
  

          >
          0
          1
      
```
- `--members` 显示消费者组中成员的描述信息  
 仅在`--bootstrap-server` ,`--describe`下使用  
 不显示offset信息
    - ```

        ■ ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --describe --group 1 --members

```
- `--delete` 删除组
  - `--delete-offsets` 删除消费者组消费位置  
 同时只支持一个消费者和多个topic
    - ```

        ■ ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --delete-offsets --group 0 --topic test-topic execute

```
- `--execute` 执行操作: `--reset-offsets`
  - `--reset-offsets` 重置消费者组消费信息  
 可使用`--dry-run`或`--execute` 执行
  - `--all-topics` Consider all topics assigned to a group in the `reset-offsets` process.
  - `--topic [String topicName]` 指定topic名称, 用于delete 或 offset process  
 支持多个topic  
 需指定`--all-topics` 或 `--topic [topicName:partitionId,...]` 使用  
 . In `reset-offsets` case, partitions can be specified using this format: `topic1:0,1,2`, where 0,1,2 are the partition to be included in the process.
  - 重新指定消费者组消费位置
    - `--to-earliest` Reset offsets to earliest offset.
    - `--to-latest` Reset offsets to latest offset.
    - `--to-offset <Long: offset>` Reset offsets to a specific offset.
 需在`--reset-offsets` 下指定该参数
    - ```

        ■ ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --reset-offsets --group 0 --topic test-topic --to-earliest execute

```
- config 配置文件目录  
<http://kafka.apache.org/documentation/#configuration>
  - server.properties kafka broker(Server)配置
    - broker.id BrokerId
    - log.dirs 数据文件保存目录
    - zookeeper.connect zookeeper链接地址  
 格式:

hostname1:port1,hostname2:port2,hostname3:port3/chroot/path

- listeners 设置监听类型和端口

FORMAT:

listeners = listener\_name://[host.name]:port

host.name 为变量 或具体指

默认: PLAINTEXT://:9092

- num.partitions=1 设置默认分区数
- compress-type 压缩类型  
gzip snappy lz4 zstd
- offsets.topic.replication.factor topic 分区默认副本数
- offsets.topic.segment.bytes topic分区保存磁盘的segment文件大小
- producer.properties kafka-console-producer生产者配置
- consumer.properties kafka-console-consumer消费者配置
- libs jar包目录
- logs 日志目录
- site-docs 文档目录

#### ○ 常用命令

- 创建Topic:  
kafka-topic.sh --create --bootstrap-server 127.0.0.1:9092 --partitions 1 --replication-factor 1 --topic newTopic
- 查看topic信息:  
kafka-topic.sh --describe --bootstrap-server 127.0.0.1:9092 --topic newTopic
- 修改topic分区:  
kafka-topic.sh --alter --bootstrap-server 127.0.0.1:9092 --partitions 2 --topic newTopic
- 删除分区:  
kafka-topic.sh --delete --topic newTopic
- 查看topic消费情况:  
> ./bin/kafka-consumer-groups.sh --bootstrap-server localhost:9092 --describe --all-groups --offsets

#### ○ Zookeeper结构

- /brokers kafka实例目录  
/brokers/ids/ kafka实例列表  
/brokers/topics/ kafka下topic列表  
/brokers/topics/[topic] kafka下某个topic  
/brokers/topics/[topic]/partitions/ kafka下topic分区列表  
/brokers/topics/[topic]/partitions/0..n/state  
/brokers/topics/\_\_consumer\_offsets  
/consumers kafka所有消费者

#### ● 快速开始

<http://kafka.apache.org/quickstart>

- 1. 下载最新安装包并解压  
> tar -xzf kafka\_2.12-2.5.0.tgz  
> cd kafka\_2.12-2.5.0
- 2. 启动Kafka Server (需Java环境)  
依赖于zookeeper, 需先启动Zookeeper  
> bin/kafka-server-start.sh config/server.properties  
可使用已有Zookeeper服务,或启动Kafka自带的单节点Zookeeper服务

(启动Kafka自带Zookeeper单节点服务)

> bin/zookeeper-server-start.sh config/zookeeper.properties&

o 3. 创建topic

创建topic命令需指定broker地址或zookeeper地址,

java.lang.IllegalArgumentException: Only one of --bootstrap-server or --zookeeper must be specified  
topic名称,partition数量,partition replica数量

> ./bin/kafka-topics.sh --create --bootstrap-server 127.0.0.1:9092 --topic test-topic

> Created topic test-topic.

java.lang.IllegalArgumentException: Only one of --bootstrap-server or --zookeeper must be specified

--create 创建topic

--alter 修改topic partitions,replication-factor,配置等信息

--delete 删除partitions

--replication-factor 复制因子

--partitions 分区数

--topic topic名称

--list 查看kafka topic列表

--describe 查看topic信息

> ./bin/kafka-topics.sh --list --bootstrap-server 127.0.0.1:9092

查看topic信息:

> ./bin/kafka-topics.sh --describe --bootstrap-server localhost:9092 --topic test-topic

修改Topic分区:

> ./bin/kafka-topics.sh --alter --zookeeper localhost:22181 --topic test-topic --partitions 2

手工指副本分配:

--replica-assignment broker\_id\_for\_part1\_replica1 : broker\_id\_for\_part1\_replica2...

o 4. 创建生产者, 发送消息

> ./bin/kafka-console-producer.sh --bootstrap-server 127.0.0.1:9092 --topic test-topic

o 5. 启动消费者, 消费消息

> ./bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic test-topic --from-beginning --group 0

--from-beginning 从头开始消费

--group 指定Consumer Group

o 6. 创建kafka server集群

■ 1. 复制多个server.properties配置文件, 并修改以下配置:

broker.id=0

listeners=PLAINTEXT://:9093

logDir=/tmp/kafka-logs-1

■ 2. 分别启动3个Server

> bin/kafka-server-start.sh config/server-1.properties &

...

> bin/kafka-server-start.sh config/server-2.properties &

...

■ 3. 创建一个具有3分区,3复制因子的topic: topic-three

> ./bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --partitions 3 --replication-factor 3 --topic topic-three

> ./bin/kafka-topics.sh --describe --bootstrap-server localhost:9092 --topic topic-three

Topic: topic-three PartitionCount: 3 ReplicationFactor: 3 Configs: segment.bytes=1073741824

|                    |              |           |                 |            |
|--------------------|--------------|-----------|-----------------|------------|
| Topic: topic-three | Partition: 0 | Leader: 1 | Replicas: 1,2,0 | Isr: 1,2,0 |
| Topic: topic-three | Partition: 1 | Leader: 0 | Replicas: 0,1,2 | Isr: 0,1,2 |
| Topic: topic-three | Partition: 2 | Leader: 2 | Replicas: 2,0,1 | Isr: 2,0,1 |

- UI客户端
  - kafkatool

## RocketMQ

## RabbitMQ

## ActiveMQ

## 服务默认端口

---

**Mysql: 3306**

**Oracle: 1521**

**ProgreSQL: 5432**

**Redis: 6379**

**Mongo: 27017**

**zookeeper: 2181**

**kafka: 9092**

**Tomcat: 8080**

**Windows RDP: 3389**

## 分布式框架

---

## SpringCloud

- Springboot
  - <https://docs.spring.io/spring-boot/docs/2.3.1.RELEASE/reference/html/>
  - 配置
    - org.springframework.boot.context.config
    - .ConfigFileApplicationListener
    - 配置项 (1.5+和2.0+版本配置名称有区别)
      - <https://docs.spring.io/spring-boot/docs/2.3.1.RELEASE/reference/html/appendix-a-application-properties.html>
      - 核心配置
        - logging.config 日志记录配置文件的位置。例如，用于logback的 `classpath: logback.xml`。
        - logging.level.\* 日志级别严重性映射。例如，`logging.level.org.springframework = DEBUG`。

- spring.application.name 应用名称
- spring.autoconfigure.exclude 要排除的自动配置类。  
同 SpringBootApplication(exclude = [])
- spring.banner.image.location  
banner图片文件的位置 (也可以使用jpg或png)  
classpath:banner.gif
- spring.banner.location banner文字资源位置  
默认: classpath:banner.txt  
org.springframework.boot.SpringApplicationBannerPrinter
- spring.config.location 替换默认设置的配置文件位置,多个路径按,分割  
默认为: classpath:/,classpath:/config/,file:./,file:./config/
- spring.config.name 替换默认配置文件名  
默认为application  
ConfigFileApplicationListener
- spring.profiles  
用逗号分隔的概要文件表达式列表, 至少要匹配一个概要文件表达式才能包含在内的文档。
- spring.profiles.active 以逗号分隔的活动配置文件列表。可以被命令行开关覆盖
- spring.profiles.include  
无条件激活指定的逗号分隔的配置文件列表 (如果使用YAML, 则激活配置文件列表)
- JSON属性
  - spring.jackson.date-format 日期格式字符串或标准日期格式类名称。例如, yyyy-MM-dd HH: mm: ss。
  - spring.jackson.time-zone 格式化日期时使用的时区。例如, "America / Los\_Angeles"或" GMT + 10"
- 数据属性
  - spring.data.elasticsearch.client.reactive.endpoints  
要连接到的Elasticsearch端点的逗号分隔列表
  - spring.data.elasticsearch.client.reactive.password  
ES连接凭证密码
  - spring.data.elasticsearch.client.reactive.username  
ES连接用户名
  - spring.data.mongodb.authentication-database mongo认证数据库名称
  - spring.data.mongodb.auto-index-creation MongoDB是否创建索引
  - spring.data.mongodb.database  
MongoDB数据库名称
  - Neo4jredis
  - solr
- 服务器属性
  - server.address 服务绑定的地址
  - server.port 服务器http端口
  - server.compression.enabled 是否启用响应压缩。 false
  - server.compression.mime-types 以逗号分隔的应压缩的MIME类型列表。
  - server.error.path 错误控制器的路径。
- application.properties/.xml
  - 包含应用相关的配置项, 被应用程序上下文(ApplicationContext)加载, 会被后续的配置文件覆盖掉相同属性的配置
- bootstrap.properties



- 包含引导相关的外部配置项，被引导上下文(BootstrapContext)加载，BootstrapContext是ApplicationContext的父上下文，二者共用Environment且优先于application.properties加载。  
如: 依赖的外部的Zookeeper配置，spring.application.name server.port等配置
  - application-[profile].properties
    - 分环境应用程序上下文配置文件，当设置spring.profiles.active=xxx时，会加载application-xxx.properties
  - bootstrap-[profile].properties
  - application.yml
    - 与application.properties 作用相同
  - bootstrap.yml
    - 与bootstrap.properties 作用相同
  - 命令行传入变量
    - 命令行传入变量可覆盖配置文件中的变量
  - 系统环境变量
  - 加载顺序
    - 不同名称配置文件加载顺序
      - bootstrap
      - bootstrap-[profile]
      - application
      - application-[profile]
      - 后边加载的配置会覆盖前边的配置
    - 同名称不同后缀加载顺序
      - 0 = "properties"
      - 1 = "xml"
      - 2 = "yml"
      - 3 = "yaml"
  - 默认配置文件加载路径
    - classpath:/,classpath:/config/,file:/,file:/config/
    - 加载顺序为: 以上顺序的返序
    - ConfigFileApplicationListener#getSearchNames()
- 特性: <https://docs.spring.io/spring-boot/docs/2.3.1.RELEASE/reference/html/spring-boo-t-features.html>
- Maven依赖
  - SpringBoot依赖有2种方式
    - 1. 继承spring-boot-starter-parent
 

```
org.springframework.boot
spring-boot-starter-parent
2.1.3.RELEASE
<relativePath/>
```

改方法可以使用properties覆盖内部依赖版本，如:

```
<spring-data-releasetrain.version>Fowler-SR2</spring-data-releasetrain.version>
```

- 2. 在DependencyManagement中scope=import 依赖spring-boot-dependency pom

```
org.springframework.data  
spring-data-releasetrain  
Fowler-SR2  
import  
pom
```

```
org.springframework.boot  
spring-boot-dependencies  
2.1.3.RELEASE  
pom  
import
```

该方式不能使用properties形式覆盖原始依赖版本。要达到同样效果，需要在dependencyManagement中spring-boot-dependencies前添加pom依赖

- <!--SpringBoot依赖-->  
<dependency> org.springframework.boot  
spring-boot-dependencies  
\${spring-boot.version}  
pom  
import

- org.springframework.cloud  
spring-cloud-dependencies  
\${spring-cloud.version}  
pom  
import

- 所有springboot提供的starter 都以spring-boot-starter-[xx]开头
- server.port使用依赖

- org.springframework.cloud  
spring-cloud-context

- 主方法

- public static void main(String[] args) {  
SpringApplication.run(Example.class, args);  
}

- 自定义SpringApplication  
new  
SpringApplication(MySpringConfiguration.class).bannerMode(Banner.Mode.OFF)  
.run(args);
- 在启动期间运行的任务应由CommandLineRunner和ApplicationRunner组件执行，而不是使用Spring组件生命周期回调（例如）@PostConstruct。
- 访问应用程序参数:  
如果您需要访问传递给的应用程序参数，则SpringApplication.run(...)可以注入org.springframework.boot.ApplicationArgumentsBean。该ApplicationArguments接口提供对原始String[]参数以及已解析option和non-option参数的访问  
@Component  
public class MyBean {  
@Autowired  
public MyBean(ApplicationArguments args) {  
boolean debug = args.containsOption("debug");  
List<String> files = args.getNonOptionArgs();  
// if run with "--debug logfile.txt" debug=true, files=["logfile.txt"]  
}  
}
- 使用ApplicationRunner或CommandLineRunner
  - 如果启动后需要运行一些特定的代码SpringApplication，则可以实现ApplicationRunner或CommandLineRunner接口。这两个接口以相同的方式工作，并提供一个单一的run方法，该方法在SpringApplication.run(...)完成之前就被调用。  
这些CommandLineRunner接口提供了作为简单字符串数组的应用程序参数访问，而则ApplicationRunner使用ApplicationArguments了前面讨论的接口。以下示例显示了一个CommandLineRunnerwith run方法  
@Component  
public class MyBean implements CommandLineRunner {  
public void run(String... args) {  
// Do something...  
}  
}
- 外部化配置
  - 配置随机值  
RandomValuePropertySource是用于注射的随机值（例如，进入机密或试验例）是有用的。它可以产生整数，longs，uuid或字符串
    - my.secret=\${random.value}  
my.number=\${random.int}  
my.bignumber=\${random.long}  
my.uuid=\${random.uuid}  
my.number.less.than.ten=\${random.int(10)}  
my.number.in.range=\${random.int[1024,65536]}
  - 访问命令行属性
    - 默认情况下，SpringApplication将所有命令行选项参数（即以开头的参数-，例如--server.port=9000）转换为a property并将其添加到Spring Environment。如前所述，命令行属性始终优先于其他属性源。  
如果您不想将命令行属性添加到中Environment，则可以使用禁用它们SpringApplication.setAddCommandLineProperties(false)。
  - Application Property Files

- YAML Shortcomings
    - 无法使用@PropertySource注释加载YAML文件
  - org.springframework.context.Aware 接口
    - 可实现相关\*Aware接口，来获取相关配置或bean  
或使用@Autowired注入  
EnvironmentAware 获取Environment  
ApplicationContextAware 获取ApplicationContext  
BootstrapContextAware 获取BootstrapContext
  - 注解  
org.springframework.boot.autoconfigure
    - @SpringBootApplication 声明为一个SpringBoot应用  
默认会加载DataSourceAutoConfiguration.class
      - 排除自动配置的类:
1. @SpringBootApplication(exclude = {DataSourceAutoConfiguration.class})
  2. @SpringBootApplication(excludeName = "org.xx.xx")指定完整限定名
  3. 使用spring.autoconfigure.exclude属性来控制要排除的自动配置类的列表。
    - @SpringBootApplication注释可用于启用这三个功能，即：
      - @EnableAutoConfiguration： 启用Spring Boot的自动配置机制
      - @ComponentScan： 启用@Component(@Controller,@Service,@Repository)对应用程序所在的软件包的扫描（请参阅最佳实践）
      - @Configuration： 允许在上下文中注册额外的bean或导入其他配置类
    - @EnableAutoConfiguration： 启用Spring Boot的自动配置机制， 自动加载,注入Spring配置
    - @ComponentScan(basePackages = {,}) 指定Spring组件的扫描目录，包括扫描@Configuration  
不指定时，默认扫描Application所在包下的所有路径
    - Config
      - @Configuration 声明自动配置加载的注解
      - @ConfigurationProperties(prefix = "")  
将Spring配置文件属性绑定到JavaBean中
  4. 需使用@Component加入Bean扫描
  5. 在扫描不到时，在任何
    - @Configuration上使用@EnableConfigurationProperties(ValidatorConfig.class) 声明要自动配置的类
  6. 在@Bean上扫描，可从Environment中的配置注册到bean中；  
prefix: 配置属性前缀,按.分割，支持属性嵌套；  
Spring Boot提供了绑定@ConfigurationProperties类型并将其注册为Bean的基础架构；  
有时，带注释的类@ConfigurationProperties可能不适合扫描，例如，如果您正在开发自己的自动配置或想要有条件地启用它们。在这些情况下，请使用@EnableConfigurationProperties注释指定要处理的类型列表。可以在任何@Configuration类上完成此操作；  
验证：
 

```
@ConfigurationProperties(prefix="acme")
@Validated
public class AcmeProperties{
    @NotNull
    private InetAddress remoteAddress;
}
```

可使用@Validated添加验证

- 依赖:

```
<dependency> org.springframework.boot  
spring-boot-configuration-processor
```

- @Component

```
@ConfigurationProperties("acme")
```

```
public class AcmeProperties {}
```

或

```
@Configuration
```

```
@EnableConfigurationProperties(AcmeProperties.class)
```

```
@ConfigurationProperties("acme")
```

```
public class AcmeProperties {}
```

- 要使用配置属性扫描, 请将@ConfigurationPropertiesScan注释添加到您的应用程序。

通常, 它将添加到带有注释的主应用程序类中, @SpringBootApplication但可以将其添加到任何@Configuration类中。默认情况下, 将从声明注释的类的包中进行扫描。如果要定义要扫描的特定程序包

```
@SpringBootApplication
```

```
@ConfigurationPropertiesScan({ "com.example.app", "org.acme.another" })
```

```
public class MyApplication {
```

```
}
```

- 宽松的绑定

- acme.my-project.person.first-name

use in .properties and .yaml 中划线格式

- acme.myProject.person.firstName

驼峰格式

- acme.my\_project.person.first\_name

下划线格式

- ACME\_MYPROJECT\_PERSON\_FIRSTNAME 系统环境变量 大写格式

- 以上配置可以在properties,.yaml,环境变量,系统变量中使用

- 转换时间

- Spring Boot为表达持续时间提供了专门的支持。

可以使用该注解进行类型转换

```
@DurationUnit(ChronoUnit.SECONDS)
```

配置文件值可以是:

7. Long形式, 使用ms作为默认单位

8. ISO-8601格式

9. 值和单位耦合的可读性高的单位(如 10s表示10秒)

- @ConfigurationProperties("app.system")

```
public class AppSystemProperties {
```

```
@DurationUnit(ChronoUnit.SECONDS)
```

```
private Duration sessionTimeout = Duration.ofSeconds(30);
```

```
}
```

- 转换日期

- Spring Boot还可以使用java.time.Periodtype。

可以在应用程序属性中使用以下格式:

10. 常规int表示形式 (除非@PeriodUnit指定, 否则使用天作为默认单位)

11. 标准的ISO-8601格式使用java.time.Period

12. 值和单位对耦合的更简单格式 (例如, 1y3d表示1年零3天)

o 转换数据大小

- Spring Framework的DataSize值类型表示字节大小

@DataSizeUnit(DataUnit.MEGABYTES)

应用程序属性中的以下格式可用：

13. 常规long表示形式（除非@DataSizeUnit已指定a，否则使用字节作为默认单位）

2.值和单位耦合的更具可读性的格式（例如，10MB意味着10兆字节）

- @ConstructorBinding 构造函数绑定

@ConfigurationProperties("acme")

- @PropertySource("")

与@Configuration，

或 @ConfigurationProperties

一起使用

指定配置文件

- @Value (“\${?:default}”) 注入properties里面的配置项，及执行SpEL表达式

- SpEL 支持算数运算符，逻辑运算符，三目运算符(a?:b),

a?.b等

<https://www.jianshu.com/p/e0b50053b5d3>

- 通过@value("#{}")获取springcontext容器中的值的信息。

如果我们想通过@value获取spring容器中的值（包括bean和bean的属性值），我们可以通过

@value("#{bean名称}")或者@value("#{bean名称.属性名}",该属性要有setter方法)

可执行SpEL表达式

- 通过@value("\${}")获取properties文件中的值的信息。

可执行部分SpEL表达式

- @ImportResource 加载xml配置文件

- @Import 导入配置文件

- @EnableDiscoveryClient 开启服务发现，可用于zookeeper和eureka

org.springframework.cloud.client.discovery.EnableDiscoveryClient

- zookeeper注册中心依赖

org.springframework.cloud

spring-cloud-starter-zookeeper-discovery

- Feign(Ribbon+Hystrix)

- @EnableFeignClients 启用feign

org.springframework.cloud.netflix.feign.EnableFeignClients

- feign/hystrix依赖

org.springframework.cloud

spring-cloud-starter-feign

org.springframework.cloud

spring-cloud-starter-hystrix

- @EnableHystrix 启用断路器

- 同feign.hystrix.enabled=true

- hystrix.command.default.execution.timeout.enabled=true 设置Hystrix默认超时开启

- hystrix.command.default.execution.isolation.thread.timeoutInMilliseconds 设置

Hystrix超时时间

- @EnableCircuitBreaker 开启断路器,可服务熔断

- @FeignClient(value="serverName",fallback=XXX.class) 设置接口为Feign请求

- @HystrixCommand注解方法失败后，系统将切换到fallbackMethod方法执行。指定回调方

法

使用feign时, 无需设置改注解

- Eureka
  - @EnableEurekaClient配置本应用将使用服务注册和服务发现
  - @EnableEurekaServer 启动一个服务注册中心
- SpringMVC
  - @RestControllerAdvice
- @ControllerAdvice
  - 增强Controller处理, 可实现:

#### 14. Controller 错误全局处理

#### 15. Controller全局数据绑定

##### 3.Controller 全局数据预处理

- @ExceptionHandler
  - 设置异常处理方法, 作用于方法, 与@ControllerAdvice配合使用
- @InitBinder
  - 数据预处理, 作用于方法, 与

@ControllerAdvice配合使用

- @ModelAttribute
  - 数据绑定
- @RestController 声明Controller为RESTFul接口, 并且各方法的返回值会自动处理为

@ResponseBody

- @CrossOrigin 允许跨域
- @Controller 声明一个Controller类
- @RequestMapping

@PostMapping

@GetMapping

- @PathVariable 注入路径参数

@RequestParam 注入QueryString/Form参数

@RequestBody 注入RequestBody为对象

- @ResponseBody 设置返回值对象转为json
- SpringBoot设置时区
  - 1. 接口JSON设置时区

spring.jackson.date-format=yyyy-MM-dd HH:mm:ss

spring.jackson.time-zone=Asia/Shanghai

- 2. 项目环境设置时区(可尝试在SpringApplication.run()前设置)

@PostConstruct

void setDefaultTimezone() {

TimeZone.setDefault(TimeZone.getTimeZone("Asia/Shanghai"));

}

- JPA

- @Entity

@Table

- @Column 声明字段
- @Transient 声明不需要与数据库映射的字段, 在保存的时候不需要保存进数据库。
- 测试
  - @SpringBootTest(webEnvironment = RANDOM\_PORT)

@ActiveProfiles("test")

@Slf4j

public abstract class TestBase {

.....

}

- @Test声明一个方法为测试方法

- @Test

```
@Transactional
@WithMockUser(username = "user-id-18163138155", authorities = "ROLE_TEACHER")
void should_import_student_success() throws Exception {
    .....
}
```

- @Transactional被声明的测试方法的数据会回滚，避免污染测试数据。
- @WithMockUser Spring Security 提供的，用来模拟一个真实用户，并且可以赋予权限。

- Mybatis
  - @MapperScan Mybatis接口扫描
  - ```
org.mybatis.spring.boot
mybatis-spring-boot-starter
${mybatis.spring.boot.version}
```
- @Transactional 开启事务，非Mybatis
- Scheduling
  - @EnableScheduling 开启Scheduling
  - @Scheduled(cron = "{\$}") 开启定时任务
- Mongo
  - @EnableMongoRepositories(basePackages = {""}) 开启Mongo Repository扫描
- Swagger 自动生成接口文档
  - @Api(value="",description="",tags="") 声明一个API组
  - @ApiOperation(value="",notes="",httpMethod="") 声明一个API
- Lombok 自动生成Bean Get/Set
  - @Data
- @Setter
- @Getter
- @AllArgsConstructor
- @NoArgsConstructor
  - @Slf4j 自动生成slf4j日志对象

- Gateway(Zuul) 网关，使用各种过滤器实现
- Discovery
  - zookeeper
  - eureka
- Feign(Hystix+Ribbon)
- Spring-Retry(重试机制)

## RPC

- Dubbo
- Apache Thrift

## 分布式事务

- LCN
- Seata

## 定时任务

## 分布式定时任务



- ElasticJob(当当开源)

## A/B机打卡

## Oauth

---

### OAuth2.0

- OAuth2.0是OAuth协议的延续版本，但不向前兼容，2012年10月正式发布为RFC 6749
  - OAuth2.0是授权行业的标准协议，致力于简化客户端开发人员工作，同时为Web应用，桌面应用，移动电话等提供特定的授权流程。  
<https://oauth.net/2/>
  - 参与者
    - 服务提供方：用户使用服务提供方存储受保护的资源
    - 用户：存放在服务提供方的受保护资源的拥有者
    - 客户端：需要访问服务提供方资源的第三方应用。认证前，客户端需向服务提供者申请客户端认证
  - OAuth2.0 基本认证流程
    - 1. 通过AppId, AppSecret请求授权地址生成Auth Code，请求中可带state字段，scope类型
    - 2. 通过用户Auth Code和AppId,AppSecret，请求生成AccessToken (短期，有时效性，有权限范围scope)
    - 3. 通过AccessToken 获取用户OpenId，获取用户信息等其他API
  - OAuth2.0分为2个角色
    - Authorization Server负责获取用户授权并分发token
    - Resource Server 负责处理API Call
  - OAuth2.0的四种授权方式
    - 1. 授权码(authorization-code)
      - 指第三方应用先申请授权码，然后再根据授权码获得令牌  
authorize: response\_type=code  
token: grant\_type=authorization\_code
      - 如: A网站需要B网站的授权，
1. 访问 B网站授权接口获得授权码  
[https://b.com/oauth2/authorize?](https://b.com/oauth2/authorize?response_type=code&client_id=CLIENT_ID&redirect_uri=CALLBACK_URL&scope=read)  
response\_type=code&  
client\_id=CLIENT\_ID&  
redirect\_uri=CALLBACK\_URL&  
scope=read
  2. B网站跳转用户登录授权页，授权成功后跳转回redirect\_url A网站，同时返回授权码  
[https://a.com/callback?code=AUTHORIZATION\\_CODE](https://a.com/callback?code=AUTHORIZATION_CODE)
  3. A网站后端拿到code后，使用code和client\_id,client\_secret 申请令牌  
[https://b.com/oauth2/token?](https://b.com/oauth2/token?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&grant_type=authorization_code&code=AUTHORIZATION_CODE&redirect_uri=CALLBACK_URL)  
client\_id=CLIENT\_ID&  
client\_secret=CLIENT\_SECRET&  
grant\_type=authorization\_code&  
code=AUTHORIZATION\_CODE&  
redirect\_uri=CALLBACK\_URL
- 响应数据:
- ```
{
```

```
"access_token":"ACCESS_TOKEN",
"token_type":"bearer",
"expires_in":2592000,
"refresh_token":"REFRESH_TOKEN",
"scope":"read",
"uid":100101,
"info":{"...}
}
```

- 案例: 微信网页授权

o 2. 隐藏式(implicit)

- 允许直接向前端颁发令牌。这种方式没有授权码这个中间步骤, 所以称为 (授权码) "隐藏式" (implicit)。

authorize: response\_type=token

- 如:

4. A 网站提供一个 链接, 要求用户跳转到B网站, 授权用户数据给A完整使用。

<https://b.com/oauth2/authorize?>

response\_type=token&

client\_id=CLIENT\_ID&

redirect\_uri=CALLBACK\_URL&

scope=read

5. 用户跳转到B网站, 登录同意给予A网站授权。这时, B网站就会跳回redirect\_uri参数的指定网站, 并且把令牌作为URL参数, 传给A网站

[https://a.com/callback#token=ACCESS\\_TOKEN](https://a.com/callback#token=ACCESS_TOKEN)

注意, 令牌的位置是 URL 锚点 (fragment), 而不是查询字符串 (querystring), 这是因为 OAuth 2.0 允许跳转网址是 HTTP 协议, 因此存在"中间人攻击"的风险, 而浏览器跳转时, 锚点不会发到服务器, 就减少了泄漏令牌的风险。

这种方式把令牌直接传给前端, 是不安全的。因此, 只适用于一些安全要求不高的场景, 并且令牌有效期非常短, 通常就是会话期间 (session) 有效, 浏览器关掉, 令牌就失效。

o 3. 密码式(password)

- 如果你高度信任某个应用, RFC 6749 也允许用户把用户名和密码, 直接告诉该应用。该应用就使用你的密码, 申请令牌, 这种方式称为"密码式" (password)。

grant\_type=password

- 如:

6. A网站要求用户提供B网站的用户名和密码。拿到以后, A就直接向B请求令牌

<https://oauth.b.com/token?>

grant\_type=password&

username=USERNAME&

password=PASSWORD&

client\_id=CLIENT\_ID

7. B网站验证身份通过后, 直接给出令牌。注意, 这时不需要跳转, 而是把令牌放在JSON数据中, 作为Http响应, A以此拿到令牌

这种方式需要用户给出自己的用户名密码, 风险很大, 因此只适用于其他授权方式都无法采用的情况, 而且必须是用户高度信任的应用

o 4. 凭证式 (client credentials)

- 适用于没有前端的命令行应用, 即在命令行下请求令牌。

grant\_type=client\_credentials

- 如:

8. A应用在命令行向B发出请求

<https://oauth.b.com/token?>

grant\_type=client\_credentials&

client\_id=CLIENT\_ID&

client\_secret=CLIENT\_SECRET

9. B网站验证通过后，直接返回令牌

这种方式给出的令牌是针对第三方应用的，不是针对于用户，即有可能多个用户共享同一个令牌

10. A拿到令牌后，请求B接口时，都必须在请求头带认证字段"Authorization"，值为"Bearer token"

curl -H "Authorization: Bearer ACCESS\_TOKEN" \

["https://api.b.com"](https://api.b.com)

- 案例：微信公众号

- 更新令牌

- OAuth2.0允许用户更新令牌，具体方法：

- B网站颁发令牌时，一次颁发2个令牌，一个用户获取数据access\_token，一个用于刷新令牌refres\_token。

- 令牌到期前使用refresh\_token发送请求，更新令牌

- token: grant\_type:refresh\_token

- <https://b.com/oauth/token?>

- grant\_type=refresh\_token&

- client\_id=CLIENT\_ID&

- client\_secret=CLIENT\_SECRET&

- refresh\_token=REFRESH\_TOKEN

## OpenId

- 概念：去中心化的网上身份认证系统，

- 我们可以通过 URI（又叫 URL 或网站地址）来认证一个网站的唯一身份

- 原理简述：假设你已经拥有一个在A网站注册获得的OpenID帐号，B网站支持 OpenID帐号登录使用，而且你从未登录过。此时你在B网站的相应登录界面输入你的OpenID帐号进行登录的时候，浏览器会自动转向A网站的某个页面 进行身份验证。这时你只要输入你在A网站注册时候提供的密码登录A网站，对B网站进行验证管理（永久允许、只允许一次或者不允许）后，页面又会自动转到B 网站。

## 日常工具

---

### 开发

- Maven
- Git
- Document4j(WordToPdf)

### 工具类(Jar)

- Guawa

## JavaScript/ES6/jQuery

---

## Interview

---

## Http/Https原理和流程

- Http原理

- HyperText Transfer Protocol

- 应用层协议,基于TCP/IP协议,C/S架构,基于请求/响应范式的,明文传输

- 无状态协议: 客户端和服务端不需要建立长久连接

- 默认端口: 80

- TCP/IP 4层模型

- IP 网络层

- TCP 传输层

- TCP能保证数据包顺序传输

- TCP报文

- ACK、SYN和FIN这些大写的单词表示标志位, 其值要么是1, 要么是0

- ack、seq小写的单词表示序号

- ACK: 确认序号有效

- SYN: 发送新连接

- SYN=1 表示这是一个连接请求, 或连接接受报文

- SYN这个标志位只有在TCP建产连接时才会被置1, 握手完成后SYN标志位被置0。

- FIN: 释放连接

- TCP3次握手建立连接

- 第一次握手: 客户端发送SYN包(syn=1)到服务器, 并设置发送序号seq为X,进入SYN\_SEND状态, 等待服务器确认;

- SYN:1 seq:X

- 第二次握手: 服务器收到SYN包, 必须确认客户的SYN(ack=X+1), 同时自己也 发送一个SYN包 (syn=1), 即SYN+ACK包, 设置发送序号seq为Y 此时服务器进入SYN\_RECV状态;

- SYN:1 ACK:1 seq:Y ack: X+1

- 第三次握手: 客户端收到服务器的SYN+ACK包, 向服务器发送确认包ACK(ack=Y+1), 并设置发送序号seq为Z,此包发送完毕, 客户端和服务端进入ESTABLISHED状态, 完成三次握手。

- ACK: 1 ack: Y+1 SEQ:Z

- 握手过程中传送的包里不包含数据, 三次握手完毕后, 客户端与服务器才正式开始传送数据。理想状态下, TCP连接一旦建立, 在通信双方中的任何一方主动关闭连接之前, TCP 连接都将被一直保持下去。

- 原因: 3次握手完成两个重要的功能, 既要双方做好发送数据的准备工作(双方都知道彼此已准备好), 也要允许双方就初始序列号进行协商, 这个序列号在握手过程中被发送和确认。

- TCP4次挥手关闭连接

- 1. 主动方发送FIN报文, 并设置发送序号seq为u (等于前面已经传送过来的数据的最后一个字节的序号加1), 客户端进入FIN-WAIT-1 (终止等待1) 状态

- FIN: 1 seq:U

- 2. 被动方收到FIN后 置发送序号seq为V,发送确认序号ack=U+1,此时, 服务端就进入了CLOSE-WAIT (关闭等待) 状态

- ACK:1 seq:V ack:U+1

- 3. 被动方发送FIN+ACK, 置发送序号seq为W, 确认需要ack为U+1, 被动方进入LAST-ACK状态, 客户端进入FIN-WAIT-2 (终止等待2) 状态

- FIN:1 ACK:1 seq:W ack:U+1

- 4. 主动方收到被动方FIN+ACK后, 置发送需要seq为U+1,ack为W+1, 发送给被动方, 进入2MSL TIME-WAIT,然后连接断开。被动方收到ACK后断开连接

- ACK:1 seq:U+1 ack:W+1

- 主动方发送第四次挥手后，会等待2MSL时间  
MSL即Maximum Segment Lifetime，也就是报文最大生存时间

- 等待2MSL的意义？

第一，为了保证A发送的最有一个ACK报文段能够到达B。这个ACK报文段有可能丢失，因而使处在LAST-ACK状态的B收不到对已发送的FIN和ACK 报文段的确认。B会超时重传这个FIN和ACK报文段，而A就能在2MSL时间内收到这个重传的ACK+FIN报文段。接着A重传一次确认。

第二，就是防止上面提到的已失效的连接请求报文段出现在本连接中，A在发送完最有一个ACK报文段后，再经过2MSL，就可以使本连接持续的时间内所产生的所有报文段都从网络中消失

- TCP是全双工模式，当client发出FIN报文段时，只是表示client已经没有数据要发送了，client告诉server，它的数据已经全部发送完毕了；但是，这个时候client还是可以接受来server的数据；当server返回ACK报文段时，表示它已经知道client没有数据发送了，但是server还是可以发送数据到client的；当server也发送了FIN报文段时，这个时候就表示server也没有数据要发送了，就会告诉client，我也没有数据要发送了，如果收到client确认报文段，之后彼此就会愉快的中断这次TCP连接。
- 原因：由于TCP连接是全双工的，因此每个方向都必须单独进行关闭。这个原则是当一方完成它的数据发送任务后就能发送一个FIN来终止这个方向的连接。收到一个FIN只意味着这一方向上没有数据流动，一个TCP连接在收到一个FIN后仍能发送数据。首先进行关闭的一方将执行主动关闭，而另一方执行被动关闭。

#### ■ 提问

- 为什么连接的时候是三次握手，关闭的时候却是四次挥手？

- 3次握手 是为了保证client和Server都做好发送和接收数据的准备
- 4次挥手 是确认了Client和Server 双方都不再发送数据，都释放了连接
- 这是因为服务端的LISTEN状态下的SOCKET当收到SYN报文的连接请求后，它可以把ACK和SYN(ACK起应答作用，而SYN起同步作用)放在一个报文里来发送。但关闭连接时，当收到对方的FIN报文通知时，它仅仅表示对方没有数据发送给你了；但未必你所有的数据都全部发送给对方了，所以你可能未必会马上会关闭SOCKET,也即你可能还需要发送一些数据给对方之后，再发送FIN报文给对方来表示你同意现在可以关闭连接了，所以它这里的ACK报文和FIN报文多数情况下都是分开发送的。

- 为什么不能用两次握手进行连接？

- 3次握手完成两个重要的功能，既要双方做好发送数据的准备工作(双方都知道彼此已准备好)，也要允许双方就初始序列号进行协商，这个序列号在握手过程中被发送和确认。

现在把三次握手改成仅需要两次握手，死锁是可能发生的。作为例子，考虑计算机S和C之间的通信，假定C给S发送一个连接请求分组，S收到了这个分组，并发送了确认应答分组。按照两次握手的协定，S认为连接已经成功地建立了，可以开始发送数据分组。可是，C在S的应答分组在传输中被丢失的情况下，将不知道S 是否已准备好，不知道S建立什么样的序列号，C甚至怀疑S是否收到自己的连接请求分组。在这种情况下，C认为连接还未建立成功，将忽略S发来的任何数据分组，只等待连接确认应答分组。而S在发出的分组超时后，重复发送同样的分组。这样就形成了死锁。

类似SYN攻击

- 如果已经建立了连接，但是客户端突然出现故障了怎么办
  - 服务端有超时计时器维持，当超过维持时间后会断开连接
- 发送了FIN只是表示这端不能继续发送数据(应用层不能再调用send发送)，但是还可以接收数据

- 应用层如何知道对端关闭
  - 通常，在最简单的阻塞模型中，当你调用recv时，如果返回0，则表示对端关闭。在这个时候通常的做法就是也调用close，那么TCP层就发送FIN，继续完成四次握手。如果你不调用close，那么对端就会处于FIN\_WAIT\_2状态，而本端则会处于CLOSE\_WAIT状态。
  - [https://blog.csdn.net/qg\\_38950316/article/details/81087809](https://blog.csdn.net/qg_38950316/article/details/81087809)  
<https://www.cnblogs.com/kindnull/p/10307333.html>
- TCP/IP 与 Socket之间的关系
  - socket是对TCP/IP协议的封装，Socket本身并不是协议，而是一个调用接口(API)，通过Socket，我们才能使用TCP/IP协议
- Http工作过程  
(一次HTTP操作称为一个事务)
  - 1. URL地址解析:  
protocol ://[username:password@]hostname[:port] / path /  
[;parameters][?query]#fragment

DNS解析域名为IP

- 2. 封装HTTP请求数据包

把以上部分结合本机自己的信息，封装成一个HTTP请求数据包

- 3.封装成TCP包，建立TCP连接（TCP的三次握手）
- 4. 客户机发送请求命令

建立连接后，客户机发送一个请求给服务器，请求方式的格式为：请求类型、统一资源标识符(URL)、协议版本号，后边为MIME信息包括请求修饰符、客户机信息和可能的内容。

GET示例:

GET <http://baidu.com> HTTP/1.1

Connection: Keep-Alive

Accept: /

User-Agent: Microsoft-CryptoAPI/10.0

Host: baidu.com

POST示例:

POST <http://tracking.miui.com/track/v1> HTTP/1.1

gzip: 0

Content-Type: application/x-www-form-urlencoded

User-Agent: Dalvik/2.1.0 (Linux; U; Android 9; MIX 2 MIUI/V11.0.2.0.PDECNXM)

Host: tracking.miui.com

Connection: Keep-Alive

Accept-Encoding: gzip

Content-Length: 2626

...请求实体

- 5. 服务器响应

服务器接到请求后，给予相应的响应信息，其格式为一个状态行，包括信息的协议版本号、状态码，状态码描述，后边是MIME信息包括服务器信息、实体信息和可能的内容。

实体消息是服务器向浏览器发送头信息后，它会发送一个空白行来表示头信息的发送到此为结束，接着它就以Content-Type应答头信息所描述的格式发送用户所请求的实际数据

HTTP/1.1 200 OK

Server: NWS\_Oversea\_Domestic\_Mid

Connection: keep-alive

Date: Wed, 04 Dec 2019 02:39:19 GMT

Last-Modified: Wed, 04 Dec 2019 01:16:44 GMT

Content-Type: application/ocsp-response

Content-Length: 1574

...实体主体数据

#### - 6. 服务器关闭TCP连接

一般情况下，一旦Web服务器向浏览器发送了请求数据，它就要关闭TCP 连接，然后如果浏览器或者服务器在其头信息加入了这行代码

Connection:keep-alive

TCP连接在发送后将仍然保持打开状态，于是，浏览器可以继续通过相同的连接发送请求。保持连接节省了为每个请求建立新连接所需的时间，还节约了网络带宽。

在收到任意一方Connection:close 后关闭连接

#### - 响应码

##### - 1xx

- 100 Continue

服务器未拒绝

- 101 Switching Protocols

转换协议

##### - 2xx

- 202 Accepted 供处理的请求已被接受，但是处理未完成。

- 203 Non-authoritative Information 文档已经正常地返回，但一些应答头可能不正确，因为使用的是文档的拷贝。

- 204 No Content 没有新文档。浏览器应该继续显示原来的文档。如果用户定期地刷新页面，而Servlet可以确定用户文档足够新，这个状态代码是很有用的。

- 205 Reset Content 没有新文档。但浏览器应该重置它所显示的内容。用来强制浏览器清除表单输入内容。

- 206 Partial Content 客户发送了一个带有Range头的GET请求，服务器完成了它。

##### - 3xx

- 300 Multiple Choices 多重选择。链接列表。用户可以选择某链接到达目的地。最多允许五个地址。

- 301 Moved Permanently 所请求的页面已经转移至新的url。

- 302 Found 所请求的页面已经临时转移至新的url。

- 303 See Other 所请求的页面可在别的url下被找到。

- 304 Not Modified 未按预期修改文档。客户端有缓冲的文档并发出了一个条件性的请求（一般是提供If-Modified-Since头表示客户只想比指定日期更新的文档）。服务器告诉客户，原来缓冲的文档还可以继续使用。

- 305 Use Proxy 客户请求的文档应该通过Location头所指明的代理服务器提取。

- 306 Unused 此代码被用于前一版本。目前已不再使用，但是代码依然被保留。

- 307 Temporary Redirect 被请求的页面已经临时移至新的url。

##### - 4xx

- 400 Bad Request 服务器未能理解请求。

- 401 Unauthorized 被请求的页面需要用户名和密码。

- 402 Payment Required 此代码尚无法使用。

- 403 Forbidden对被请求页面的访问被禁止。

- 404 Not Found 服务器无法找到被请求的页面。

- 405 Method Not Allowed 请求中指定的方法不被允许。

##### - 5xx

- 500 Internal Server Error 请求未完成。服务器遇到不可预知的情况

- 501 Not Implemented 请求未完成。服务器不支持所请求的功能。

- 502 Bad Gateway 请求未完成。服务器从上游服务器收到一个无效的响应。

- 503 Service Unavailable 请求未完成。服务器临时过载或宕机。

- 504 Gateway Timeout 网关超时。

- 505 HTTP Version Not Supported 服务器不支持请求中指明的HTTP协议版本。

#### • Https原理

- Http+SSL/TLS
  - SSL: SecureSocketLayer
  - TLS: SSL3.0 Transport Layer Security
    - 默认端口号: 443
    - TLS 握手在 TCP握手之后进行
- 非对称加密
  - 使用非对称加密 (RSA) 的公钥对 (对称加密) 密钥进行加密  
对随机数-预主密钥加密
- 对称加密
  - 使用对称加密(AES,DES,3DES等)对传输数据进行加密  
使用(随机数1,随机数)
- 数字摘要
  - MD5, SHA1等
- 数字签名技术
  - 数字签名建立在公钥加密体制基础上, 是公钥加密技术的另一类应用。它把公钥加密技术和数字摘要结合起来, 形成了实用的数字签名技术。
- 数字证书内容
  - 包括了加密后服务器的公钥、权威机构的信息、服务器域名, 还有经过CA私钥签名之后的证书内容 (经过先通过Hash函数计算得到证书数字摘要, 然后用权威机构私钥加密数字摘要得到数字签名), 签名计算方法以及证书对应的域名。
- 问题
  - 如何保证服务器给客户端下发的公钥是真正的公钥, 而不是中间人伪造的公钥?
    - 1. 当客户端收到这个证书之后, 使用本地配置的权威机构的公钥对证书进行解密得到服务端的公钥和证书的数字签名, 数字签名经过CA公钥解密得到证书信息摘要。
    - 2. 用证书签名的方法计算一下当前证书的信息摘要, 与收到的信息摘要作对比, 如果一样, 表示证书一定是服务器下发的, 没有被中间人篡改过。因为中间人虽然有权权威机构的公钥, 能够解析证书内容并篡改, 但是篡改完成之后中间人需要将证书重新加密, 但是中间人没有权威机构的私钥, 无法加密, 强行加密只会导致客户端无法解密, 如果中间人强行乱修改证书, 就会导致证书内容和证书签名不匹配。
  - 优势
    - 信任主机的问题.
    - 防止通讯过程中的数据的泄密和被篡改
  - 劣势
    - 在黑客攻击、拒绝服务工具、服务器劫持方面起不到作用
    - SSL证书的信用链体系并不安全, 特别是在某些国家可以控制CA根证书的情况下, 中间人攻击一样可行
    - HTTPS连接缓存不如HTTP高效, 流量成本高
    - HTTPS连接服务器端资源占用高很多, 支持访客多的网站需要投入更大的成本
    - HTTPS协议握手阶段比较费时, 对网站的响应速度有影响, 影响用户体验
- SSL连接建立过程
  - 1. Client 向 Server发送 Hello , 随机码1,客户端支持的加密算法列表  
Client -> Hello ,random number1 ,cipher suites ->Server
  - 2. Server接收数据后响应client, 返回随机数2和匹配的加密算法  
Server -> random number2,matched cipher suites - client
  - 3. 随即 Server给Client发送数字证书报文。  
Server -> server certificate -> Client



- 4. 客户端解析证书，验证公钥是否有效，如颁发机构，过期时间等，若异常，则弹出警告提示证书存在问题，  
若正常，则生成随机值（预主密钥）
  - 5. 客户端通过认证后，通过随机数1，随机数2，预主密钥组装会话密钥，  
然后通过证书的公钥加密 预主密钥  
client -> assemble session secret key = random  
number1+random2+premaster;  
encrypt (premaster secret) with public key ->Server
  - 6. 传送加密信息：公钥加密的预主密钥
  - 7. 服务器用私密解密6中的预主密钥，然后 通过随机数1，随机数2,预主密钥 组装会话密钥
  - 8. 客户端通过会话密钥加密一条消息发送给服务端，主要验证服务端是否正常接收客户端加密数据
  - 9. 服务端通过会话密钥加密一条消息回传给客户端，如果客户端能正常接收则SSL层连接建立完成
- Https工作过程
  - 同Http工作过程，  
只是在TCP 握手后增加了TLS层握手
- Websocket原理
  - ws: 80  
wss: 443
  - Request
    - GET /chat HTTP/1.1  
Host: server.example.com  
Upgrade: websocket  
Connection: Upgrade  
Sec-WebSocket-Key: x3JJHMbDL1EzLkh9GBhXDw==  
Sec-WebSocket-Protocol: chat, superchat  
Sec-WebSocket-Version: 13  
Origin: <http://example.com>
  - Response
    - HTTP/1.1 101 Switching Protocols  
Upgrade: websocket  
Connection: Upgrade  
Sec-WebSocket-Accept: HSmrc0sMlYUkAGmm5OPpG2HaGWk=  
Sec-WebSocket-Protocol: chat
  - 是真正的全双工方式，建立连接后客户端与服务端是完全平等的，可以互相主动请求。而HTTP长连接基于HTTP，是传统的客户端对服务器发起请求的模式。
  - tomcat实现websocket
    - @ServerEndpoint(value = "/WSChat")
    - @OnOpen  
public void onOpen(Session session) {}
    - @OnClose  
public void onClose() {}
    - @OnMessage  
public void onMessage(Session session, String message) {}
  - js 实现websocket
    - websocket = new WebSocket(wsUri,"protocol");
    - websocket.onopen =function(event){};

- `websocket.onmessage =function(event){}`
- `websocket.onerror =function(event){};`
- `websocket.onclose =function(event){};`
- `Socket.readyState`
  - 只读属性 `readyState` 表示连接状态，可以是以下值：
    - 0 - 表示连接尚未建立。
    - 1 - 表示连接已建立，可以进行通信。
    - 2 - 表示连接正在进行关闭。
    - 3 - 表示连接已经关闭或者连接不能打开。
- `Socket.bufferedAmount`
  - 只读属性 `bufferedAmount` 已被 `send()` 放入正在队列中等待传输，但是还没有发出的 UTF-8 文本字节数。
- 浏览器网页同时最大发起http请求数
  - 15-20个

## 微信公众号/小程序开发

## 业务场景实践方案

---

### 高并发场景方案

### 秒杀场景方案

### 轮询

- ajax 轮询
- long poll
  - 发起请求后，服务端等待有数据后返回，然后再次发起请求
- websocket

## 压测相关

---

### Jemeter

### LoadRunner

### AWS

---

### Serverless

- Lambda

### DynamoDB

### DocumentDB

### Elemental MediaConverter

### S3

# 大数据

---

ElasticSearch

Hive

Spark

Scala

推送

---

友盟

极光

*XMind - Trial Version*