久

杨

珍

精

悟尽归原 几近于道。

create by Yangcl

2018-03-15

1 Mysql基础篇

2 Mysql高级特性

2.1数据库索引与高性能查询

2.1.1 MySQL索引5大原则

**A：最左前缀匹配原则(重要的原则)**

mysql会一直向右匹配直到遇到范围查询（>、<、between、like）就停止匹配。比如a = 1 and b = 2 and c > 3 and d = 4，如果建立（a,b,c,d）顺序的索引，d是用不到索引的，如果建立(a,b,d,c)的索引则都可以用到，a,b,d的顺序可以任意调整。 一个查询中，如果第一个查询条件为索引，同时因为使用了>、<、between或like导致他失效了，那么索引整体失效。

**B：=和in可以乱序**

比如a = 1 and b = 2 and c = 3 建立（a,b,c）索引可以任意顺序，mysql的查询优化器会帮你优化成索引可以识别的形式。

**C：尽量选择区分度高的列作为索引**

区分度的公式是count(distinct column)/count(\*)。表示字段不重复的比例，比例越大我们扫描的记录数越少，唯一键的区分度是1，而一些状态、性别字段可能在大数据面前区分度就是0，那可能有人会问，这个比例有什么经验值吗？使用场景不同，这个值也很难确定，一般需要join的字段我们都要求是0.1以上，即平均1条扫描10条记录。

**D：索引列不能参与计算，保持列“干净”**

比如from\_unixtime(create\_time) = ’2014-05-29’就不能使用到索引，原因很简单，b+树中存的都是数据表中的字段值，但进行检索时，需要把所有元素都应用函数才能比较，显然成本太大。所以语句应该写成create\_time = unix\_timestamp(‘2014-05-29’);

**E：尽量的扩展索引，不要新建索引**

比如表中已经有a的索引，现在要加(a,b)的索引，那么只需要修改原来的索引即可。

2.1.2 索引使用的典型场景(使用索引优化查询)

A：匹配全值

对索引中所有列都指定具体值，即对索引中的所有列都有等值匹配的条件。

#设置组合索引（rental\_date,inventory\_id,customer\_id）为唯一索引。

EXPLAIN

SELECT \* FROM

rental

WHERE rental\_date = '2005-05-25 17:22:10'

AND inventory\_id = 373

AND customer\_id = 343 ;



B：匹配值的范围查询

对索引值进行范围查找，这种情况最常见。

#设置索引idx\_fk\_customer\_id(customer\_id)

EXPLAIN

SELECT

\*

FROM

rental

WHERE customer\_id >= 373

AND customer\_id < 400 ;



**C：匹配最左前缀**

仅仅使用索引中的最左边列进行查询。比如组合索引（col1,col2,col3）能够被col1，col1+col2，col1+col2+col3的等值查询利用到的。

#创建索引idx\_payment\_date(payment\_date,amount,last\_update);

EXPLAIN

SELECT \* FROM

payment

WHERE payment\_date = '2006-02-14 15:16:03'

AND last\_update = '2006-02-15 22:12:32' ;



从结果可以看出利用了索引，但又row为182行，所有只使用了部分索引。换一种方式，如下脚本，将不适用索引，请注意标红色的amount条件

EXPLAIN

SELECT \* FROM

payment

WHERE **amount** = 3.98

AND last\_update = '2006-02-15' ;



从结果看出，这次查询没有利用索引，进行了全表查找。因为我们所创建的索引中没有包含amount参数。

**D：所查询列全为索引**

当查询列都在索引字段中。即select中的列都在索引中。

#创建索引idx\_payment\_date(payment\_date,amount,last\_update);

EXPLAIN

SELECT

last\_update

FROM

payment

WHERE payment\_date = '2005-08-19 21:21:47'

AND amount = 4.99 ;



Extra内容显示为Using index,说明不需要通过索引回表，Using index就是平时说的覆盖索引扫描（即找到索引，就找到了要查询的结果，不用再回表查找了）。

**E：匹配列前缀**

仅仅使用索引的第一列，并且只包含索引第1列的开头部分进行查找。

#创建索引idx\_title\_desc\_part(title(10)，description(20));

EXPLAIN

SELECT

title

FROM

film\_text

WHERE title LIKE 'african%' ;



type=range，范围扫描，基于索引做范围扫描，为诸如BETWEEN，IN，>=，LIKE类操作提供支持。

为提高like的查询效率，请注意创建索引的代码中红色的部分，title(10)大意为匹配前10个字符。大表查询中极少使用LIKE这个关键字。

**F：索引部分等值匹配，部分范围匹配**

EXPLAIN

SELECT

inventory\_id

FROM

rental

WHERE rental\_date = '2006-02-14 15:16:03'

AND customer\_id >= 300

AND customer\_id <= 400 ;



type=ref，说明使用了索引。

**G：列名是索引，column\_name is null，使用索引**

EXPLAIN

SELECT

\*

FROM

payment

WHERE rental\_id IS NULL ;



2.1.3索引存在但不能使用索引的典型场景

**A：以%开头的like查询**

EXPLAIN

SELECT

\*

FROM

actor

WHERE last\_name LIKE '%NI%' ; **#百分号“%”只能在最右的情况索引才不会失效！**



**B：数据类型出现隐式转化，不会使用索引**

EXPLAIN

SELECT

\*

FROM

actor

WHERE last\_name = 1 ;



last\_name为varchar类型存储，但匹配值为int类型的数字，此时索引失效；正确如下：WHERE last\_name = '1' ;

**C：组合索引，不满足最左原则，不使用符合索引**

详见2.1.1 A中的说明。

**D：用索引比全表扫描还慢，则不要使用索引**

如查询以“S”开头的标题的电影，返回记录比例比较大，mysql预估索引扫描还不如全表扫描。

**E：用or分割条件，若or前后只要有一个查询条件的列没有索引，就都不会用索引**

EXPLAIN

SELECT

\*

FROM

payment

WHERE customer\_id = 203

OR amount = 3.96 ;



2.1.4 mysql explain结果中type的含义

|  |  |
| --- | --- |
| type值 | 含义 |
| system | CONST的特例，当表上只有一条元组匹配 |
| const | WHERE条件筛选后表上至多有一条元组匹配时，比如WHERE ID = 2 （ID是主键，值为2的要么有一条要么没有） |
| eq\_ref | 参与连接运算的表是内表（在代码实现的算法中，两表连接时作为循环中的内循环遍历的对象，这样的表称为内表）。  基于索引（连接字段上存在唯一索引或者主键索引，且操作符必须是“=”谓词，索引值不能为NULL）做扫描，使得对外表的一条元组，内表只有唯一一条元组与之对应。 |
| ref | 可以用于单表扫描或者连接。参与连接运算的表，是内表。基于索引（连接字段上的索引是非唯一索引，操作符必须是“=”谓词，连接字段值不可为NULL）做扫描，使得对外表的一条元组，内表可有若干条元组与之对应。 |
| ref\_or\_null | 类似REF，只是搜索条件包括：连接字段的值可以为NULL的情况，比如 where col = 2 or col is null |
| range | 范围扫描，基于索引做范围扫描，为诸如BETWEEN，IN，>=，LIKE类操作提供支持 |
| index\_scan | 索引做扫描，是基于索引在索引的叶子节点上找满足条件的数据（不需要访问数据文件） |
| all | 全表扫描或者范围扫描：**不使用索引**，顺序扫描，直接读取表上的数据（访问数据文件） |
| unique\_subquery | 在子查询中，基于唯一索引进行扫描，类似于EQ\_REF |
| index\_subquery | 在子查询中，基于除唯一索引之外的索引进行扫描 |
| index\_merge | 多重范围扫描。两表连接的每个表的连接字段上均有索引存在且索引有序，结果合并在一起。适用于作集合的并、交操作。 |
| ft | FULL TEXT，全文检索 |

你定义的索引在查询时是否生效取决于这里展示的type的值。

2.2数据库索引深层原理

2.1节从实用性角度总结了在开发中用到的索引知识点，尤其在重构某个功能或者项目模块的时候显得尤为重要。但是并没有从一个深层的角度来认识索引的本质。

这一节中将汇集所有优秀的文章，加以总结：什么是索引的本质。

2.2.1索引的数据结构及算法基础

**A：索引的本质**

MySQL官方对索引的定义为：索引（Index）是帮助MySQL高效获取数据的数据结构。提取句子主干，就可以得到索引的本质：索引是一种数据结构。数据库查询是数据库的主要功能之一，最基本的查询算法是顺序查找（linear search）时间复杂度为O(n)，显然在数据量很大时效率很低。优化的查找算法如二分查找（binary search）、二叉树查找（binary tree search）等，虽然查找效率提高了。但是各自对检索的数据都有要求：二分查找要求被检索数据有序，而二叉树查找只能应用于二叉查找树上，但是数据本身的组织结构不可能完全满足各种数据结构（例如，理论上不可能同时将两列都按顺序进行组织）。所以，在数据之外，数据库系统还维护着满足特定查找算法的数据结构。这些数据结构以某种方式引用（指向）数据，这样就可以在这些数据结构上实现高级查找算法。这种数据结构就是索引。



上图展示了一种可能的索引方式。左边是数据表，一共有两列七条记录，最左边的是数据记录的物理地址（注意逻辑上相邻的记录在磁盘上也并不是一定物理相邻的）。为了加快Col2的查找，可以维护一个右边所示的二叉查找树，每个节点分别包含索引键值和一个指向对应数据记录物理地址的指针，这样就可以运用二叉查找在O(log2n)的复杂度内获取到相应数据。虽然这是一个货真价实的索引，但是实际的数据库系统几乎没有使用二叉查找树或其进化品种红黑树（red-black tree）实现的，原因会在下文介绍。

**B：B-Tree**

为了描述B-Tree，首先定义一条数据记录为一个二元组[key, data]，key为记录的键值，对于不同数据记录，key是互不相同的；data为数据记录除key外的数据。那么B-Tree是满足下列条件的数据结构：

1. d>=2，即B-Tree的度；
2. h为B-Tree的高；
3. 每个非叶子结点由n-1个key和n个指针组成，其中d<=n<=2d；
4. 每个叶子结点至少包含一个key和两个指针，最多包含2d-1个key和2d个指针，叶结点的指针均为NULL；
5. 所有叶结点都在同一层，深度等于树高h；
6. key和指针相互间隔，结点两端是指针；
7. 一个结点中的key从左至右非递减排列；
8. 如果某个指针在结点node最左边且不为null，则其指向结点的所有key小于v(key1)，其中v(key1)为node的第一个key的值。
9. 如果某个指针在结点node最右边且不为null，则其指向结点的所有key大于v(keym)，其中v(keym)为node的最后一个key的值。
10. 如果某个指针在结点node的左右相邻key分别是keyi和keyi+1且不为null，则其指向结点的所有key小于v(keyi+1)且大于v(keyi)。



由于B-Tree的特性，在B-Tree中按key检索数据的算法非常直观：首先从根节点进行二分查找，如果找到则返回对应节点的data，否则对相应区间的指针指向的节点递归进行查找，直到找到节点或找到null指针，前者查找成功，后者查找失败。B-Tree上查找算法的伪代码如下：

BTree\_Search(node, key) {

if(node == null) return null;

foreach(node.key)

{

if(node.key[i] == key) return node.data[i];

if(node.key[i] > key) return BTree\_Search(point[i]->node);

}

return BTree\_Search(point[i+1]->node);

}

data = BTree\_Search(root, my\_key);

关于B-Tree有一系列有趣的性质，例如一个度为d的B-Tree，设其索引N个key，则其树高h的上限为logd((N+1)/2)，检索一个key，其查找结点个数的渐进复杂度为O(logdN)。从这点可以看出，B-Tree是一个非常有效率的索引数据结构。

https://www.cnblogs.com/tgycoder/p/5410057.html

**C：B+Tree**

2.2.2数据库事物隔离级别

https://www.cnblogs.com/wy697495/p/11025929.html

https://baijiahao.baidu.com/s?id=1611918898724887602&wfr=spider&for=pc

http://blog.codinglabs.org/articles/index-condition-pushdown.html

3 MyCat In Action

3.1 数据库集群基本术语

3.1.1 数据切分

Mycat的核心功能，通过某种特定的条件，将放在同一个数据库中的数据分散存放在多个数据库主机中，以达到分散到多台设备负载的效果。

数据切分有两种切分模式，一种是按照不同的表，将数据切分到不同的数据库主机中，这种切分方式称为数据的垂直切分，也叫纵向切分；另一种

则是根据表中的数据逻辑关系，将同一个表中的数据按照某种特定条件(比如：租户id)拆分到多个数据库主机中，这种方式称为数据的水平切分，也

叫做横向切分，当单表数据量达到800万条以上的时候，通常会对该表做横向切分。

字符串枚举分片

Java高并发底层原理与实际应用

JMM: Java Memory Model

1) 主内存、工作内存与堆栈的关系 【理论基础】

JVM将内存组织为**主内存**和**工作内存**两个部分。主内存主要包括本地方法区和堆。每个线程都有一个工作内存，工作内存中主要包括两个部分，一个是**属于该线程**私有的栈和对主存部分变量拷贝的寄存器(包括程序计数器PC和cup工作的高速缓存区)。



1. 所有的变量都存储在主内存中(虚拟机内存的一部分)，对于所有线程都是共享的。
2. 每条线程都有自己的工作内存，工作内存中保存的是主存中某些变量的拷贝。
3. 线程对变量的所有操作都必须在工作内存中进行，而不能直接读写主内存中的变量。
4. 线程之间无法直接访问对方的工作内存中的变量，线程间变量的传递均需要通过主内存来完成。

简而言之【工作内存：强调私有性；主内存：强调共享性】。进程代表一个软件，操作系统会为每一个运行的软件创建一个进程，比如：一个Tomcat进程代表了它在操作系统中运行的标识，这个Tomcat启动了一个线程池，包含200个线程，并发处理客户端传来的请求。每一个线程中会包含若干对象、字段、静态变量等等。

**堆区(heap)：**

存储的全部是对象，每个对象都包含一个与之对应的class的信息。(class的目的是得到操作指令，注意：一个class可以创建多个对象)。

jvm只有一个堆区被所有线程共享，堆中不存放基本类型和对象引用，只存放对象本身。

**方法区**：

又叫静态区，跟堆一样，被所有的线程共享。方法区包含所有的class和static变量。

方法区中包含的都是在整个程序中永远唯一的元素，如class，static变量。https://www.cnblogs.com/cmfwm/p/7671188.html

**栈区**：

每个线程包含一个栈区，栈中只保存基础数据类型8个和自定义对象的引用(不是对象)，对象都存放在堆区中。

每个栈中的数据(原始类型和对象引用)都是私有的，其他栈不能访问。

栈分为3个部分：基本类型变量区、执行环境上下文、操作指令区(存放操作指令)。

JAVA寄存器

所有进程都使用寄存器，Java开发组决定Java只使用四个寄存器，这是因为如果使用的寄存器数多于处理器端口数，那么处理器的效率将严重地降低。Java虚拟机使用下列寄存器管理系统堆栈：

**程序记数寄存器**：跟踪程序执行的准确位置

**堆栈指针寄存器**：指示操作栈项

**框架寄存器**：指向当前执行的环境

**变量寄存器**：指向当前执行环境中第一个本地变量

堆栈分离的本质：

在现代计算机系统中，读写速度分级：硬盘<内存<CPU；CPU的读写速度远高于内存，导致CPU会出现大量空闲时间（因为在等待从内存中读取数据），因此在内存和CPU之间增加了一套缓存机制，一般打开CPU-Z会显示你的CPU有3级缓存，6M左右。**栈**保存的是若干个对象在**堆**中的内存地址，当一个线程在运行的时候，当前这个线程会根据栈中保存的地址去堆中拷贝这个对象的副本中的字段和字段的值到工作内存中，降低CPU空闲等待时间，从而提升系统性能。

JVM规范定义了线程对内存间交互操作：

**交互规则**：

1. 锁定：Lock，作用于主内存中的变量，把一个变量标识为一条线程独占的状态。
2. 读取：Read，作用于主内存中的变量，把一个变量的值从主内存传输到线程的工作内存中。
3. 加载：Load，作用于工作内存中的变量，把read操作从主内存中得到的变量的值放入工作内存的变量副本中。
4. 使用：Use， 作用于工作内存中的变量，把工作内存中一个变量的值传递给执行引擎。
5. 赋值：Assign，作用于工作内存中的变量，把一个从执行引擎接收到的值赋值给工作内存中的变量。
6. 存储：Store，作用于工作内存中的变量，把工作内存中的一个变量的值传送到主内存中。
7. 写入：Write，作用于主内存中的变量，把store操作从工作内存中得到的变量的值放入主内存的变量中。
8. 解锁：Unlock，作用于主内存中的变量，把一个处于锁定状态的变量释放出来，之后可被其它线程锁定。

**并且这8种操作必须遵循以下规则**：

1、不允许read和load、store和write操作之一单独出现。即不允许一个变量从主内存被读取了，但是工作内存不接受，或者从工作内存回写了但是主内存不接受。

2、不允许一个线程丢弃它最近的一个assign操作，即变量在工作内存被更改后必须同步改更改回主内存。

3、工作内存中的变量在没有执行过assign操作时，不允许无意义的同步回主内存。

4、在执行use前必须已执行load，在执行store前必须已执行assign。

5、一个变量在同一时刻只允许一个线程对其执行lock操作，一个线程可以对同一个变量执行多次lock，但必须执行相同次数的unlock操作才可解锁。

6、一个线程在lock一个变量的时候，将会清空工作内存中的此变量的值，执行引擎在use前必须重新read和load。

7、线程不允许unlock其他线程的lock操作。并且unlock操作必须是在本线程的lock操作之后。

8、在执行unlock之前，必须首先执行了store和write操作。

可见，锁操作都是围绕主内存来进行。在将变量从主内存读取到工作内存中，必须顺序执行read、load；要将变量从工作内存同步回主内存中，必须顺序执行store、write。并且这8种操作必须遵循以下规则：

2) JMM 【理论基础】

JVM是整个计算机虚拟模型，属于语言级的内存模型，它确保在不同的编译器和不同的处理器平台之上，通过插入特定类型的 **Memory Barrier**（内存屏障）来禁止特定类型的编译器重排序和处理器重排序，为上层提供一致的内存可见性保证；JMM 决定一个线程对共享变量(实例域、静态域和数组)的写入何时对其它线程可见（线程的可见性）。

由于JVM运行程序的实体是线程，而每个线程创建时JVM都会为其创建一个工作内存（也称为：**栈空间**），工作内存是每个线程的私有数据区域，而Java内存模型中规定所有变量都存储在**主内存**，主内存是共享内存区域，所有线程都可访问，但线程对变量的操作(读取赋值等)必须在栈空间中进行，首先要将变量从主内存拷贝到自己的栈空间，然后对变量进行操作，操作完成再将变量写回主内存，不能直接操作主内存中的变量，各个线程中的工作内存，储存着主内存中的变量副本拷贝，因为不同的线程无法访问对方的工作内存，因此线程间的通讯(传值) 必须通过主内存来完成,其简要访问过程如下图：

**JMM关于同步规定**：

1. 线程解锁前，必须把共享变量的值刷新回主内存。
2. 线程加锁前，必须读取主内存的最新值到自己的工作内存。
3. 加锁解锁是同一把锁。

线程A和线程B之间需要通信，必须经过3个步骤：

1. 线程A和线程B均从主内存中读取同一个共享变量到自己的工作内存中，各自进行计算。
2. 线程A把自己工作内存中更新过的共享变量副本刷新到主内存中。
3. 线程B到主内存中读取线程A之前更新过的共享变量。

3) Happen-Before规则【学术讨论】

Java的并发采用的是共享内存模型，使用happen-before规则实现共享变量的同步操作(简称hb规则)。该规则定义了Java多线程操作的有序性和可见性，防止了*编译器重排序* 对程序结果的影响。按照官方的说法：当一个变量被多个线程读取并且至少被一个线程写入时，如果读操作和写操作没有HB关系，则会产生数据竞争问题。要想保证操作B的线程看到操作A的结果(无论A和B是否在一个线程)，那么在A和B之间必须满足HB原则，如果没有，将有可能导致重排序。当缺少HB关系时，就可能出现重排序问题。

**HB有哪些规则**：(http://ifeve.com/java-使用-happen-before-规则实现共享变量的同步操作/)

1. 程序次序规则：一个线程内，按照代码顺序，书写在前面的操作先行发生于书写在后面的操作。
2. 锁定规则：在监视器锁上的解锁操作必须在同一个监视器上的加锁操作之前执行。
3. volatile变量规则：对一个变量的写操作先行发生于后面对这个变量的读操作。
4. 传递规则：如果操作A先行发生于操作B，而操作B又先行发生于操作C，则可以得出操作A先行发生于操作C。(*这个规则至关重要，如何熟练的使用传递规则是实现同步的关键*。)
5. 线程启动规则：Thread对象的start()方法先行发生于此线程的每一个动作。
6. 线程中断规则：对线程interrupt()方法的调用先行发生于被中断线程的代码检测到中断事件的发生。
7. 线程终结规则：线程中所有的操作都先行发生于线程的终止检测，我们可以通过Thread.join()方法结束、Thread.isAlive()的返回值手段检测到线程已经终止执行。
8. 对象终结规则：一个对象的初始化完成先行发生于他的finalize()方法的开始。

4) JMM内存交互3个基本特性【学术讨论】

原子性：

原子性即一个操作或者多个操作，要么全部执行并且执行的过程不会被任何因素打断，要么就都不执行。原子是世界上的最小单位，具有不可分割性。在 Java 中，对基本数据类型的变量的读取和赋值操作是原子性操作，即这些操作是不可被中断的，要么执行，要么不执行。

可见性：

可见性是指当多个线程访问同一个变量时，一个线程修改了这个变量的值，其他线程能够立即看得到修改的值。JMM 是通过在线程 A 变量工作内存修改后将新值同步回主内存，线程 B 在变量读取前从主内存刷新变量值，这种依赖主内存作为传递媒介的方式来实现可见性。

有序性：

**线程内**，从某个线程的角度看方法的执行，指令会按照一种叫 " 串行 "（as-if-serial）的方式执行，此种方式已经应用于顺序编程语言。

**线程间**，这个线程 " 观察 " 到其他线程并发地执行非同步的代码时，由于指令重排序优化，任何代码都有可能交叉执行。

唯一起作用的约束是：对于同步方法，同步块（synchronized 关键字修饰）以及 volatile 字段的操作仍维持相对有序。JMM保证有序性的手段是：禁止指令重排序。

JMM规定的这三个基本特性，是Java多线程的核心原理。

5) 指令重排序 与 内存屏障【理论基础】

指令重排序

**多线程环境中**线程交替执行，由于编译器优化重排的原因，两个线程中使用的变量能否保证一致性是无法确定的，结果无法预测。计算机在执行程序时，为了提高性能，编译器和处理器常常会做指令重排，一般分为以下三种重排方式：



单线程环境里面，不存在指令重排序所造成的的问题，程序最终执行结果和代码顺序执行的结果一致。在多核心CPU，多线程的环境下，处理器在进行重新排序的时候，必须要考虑指令之间的数据依赖性。

**数据依赖性**

数据依赖性是一种因果关系。如右图所示，源代码经过重排序后，可能出现的执行顺序如下：

上述源代码有三种可能出现的重排序方式，分别为：1234（排序不变）、2134和1324。由于数据依赖性的存在，第四种情况由于y需要x变量先赋值计算完成，所以不可能排到x变量参与计算的代码之前。直观感受这种情况：1243，也是违背数据依赖性原则的，所以也不会出现。

**结果无法预测** 代码如下：

int a = 0;

int b = 0;

int x = 0;

int y = 0;

|  |  |
| --- | --- |
| 线程1 | 线程2 |
| x = a | y = b |
| b = 1 | a = 2 |
| x = 0 y = 0 | |

如果编译器对这段代码进行执行重排优化后，可能出现下列情况：

|  |  |
| --- | --- |
| 线程1 | 线程2 |
| b = 1 | a = 2 |
| x = a | y = b |
| x = 2 y = 1 | |

内存屏障 - Memory Barrier

**指令重排序是编译器的一种优化策略，但有时候我们需要禁止他，此时需要使用内存屏障。**Memory Barrier又称内存栅栏，是一个 CPU 指令。可以分为：LoadLoad屏障、LoadStore屏障、StoreStore屏障和 StoreLoad屏障。内存屏障有两个作用：

特性1：保证特定操作的执行顺序。

特性2：保证某些变量的内存可见性（**volatile的内存可见性**）。

这其中**特性1**：阻止屏障两侧的指令重排序，插入一条 Memory Barrier 会告诉编译器和 CPU，不管什么指令都不能和这条 Memory Barrier 指令重排序；也就是说：可以通过插入内存屏障来禁止在内存屏障前后的指令被重排序优化。**特性二**：强制把写缓冲区/高速缓存中的脏数据等写回主内存，让缓存中相应的数据失效。如一个Write-Barrier（写入屏障）将刷出所有在Barrier之前写入cache的数据，因此任何CPU上的线程都能读取到这些数据的最新版本，让每个线程栈重新去堆中再次拷贝变量副本。

内存屏障阻碍了CPU采用优化技术来降低内存操作延迟，因此必定会带来性能损失。Java中对内存屏障的使用常见的有 volatile 和 synchronized 关键字修饰的代码块，还可以通过 Unsafe 这个类来使用内存屏障。

volatile【实际运用】

volatile是Java虚拟机提供的**轻量级**同步机制，主要解决两个问题：【**保证可见性**】和【**禁止指令重排序**】，不保证原子性。是乞丐版的synchronized，阉割了线程安全性。其最底层在操作CPU的内存屏障。

通过前面对JMM的铺垫，各个线程对主内存中共享变量的操作都是各个线程各自拷贝到自己的工作内存操作后再写回主内存中的。这就可能存在一个线程A修改了共享变量X的值还未写回主内存中时，另外一个线程B又对内存中的一个共享变量X进行操作，但此时A线程工作内存中的共享比那里X对线程B来说并不不可见。这种工作内存与主内存同步延迟现象就造成了可见性问题。禁止指令重排序则解决了顺序性问题。

可见性-验证代码

public class MyData {

// 变量未添加volatile关键字，没有可见性

public int number = 0;

// public volatile int number = 0;

public int getNumber() {

return number;

}

public void setNumber(int number) {

this.number = number;

}

}

public class VolatileMain {

public static void main(String[] args) {

MyData data = new MyData();

new Thread( () -> {

System.out.println(Thread.currentThread().getName() + " 进入");

// 线程等待3秒

try {TimeUnit.SECONDS.sleep(3);} catch (InterruptedException e) {e.printStackTrace();}

data.setNumber(10); // 设置一个新值 10

System.out.println(Thread.currentThread().getName() + " 更新 number 值为：" + data.getNumber());

} , "线程A" ).start();

while (data.getNumber() == 0) {

// main 线程在此处等待，直到number值不在等于0。但因为线程A修改number的值后，main线程没有得到通知，此处会一直循环。

}

// main 线程“想”打印出来，但打印不出来。

System.out.println(Thread.currentThread().getName() + " mission is over");

}

}

代码解读：

在VolatileMain.java中，线程A和Main线程同时操作一个MyData()对象；线程A在等待3秒后，将number设置成10。但由于JMM内存交互的可见性特点，number字段没有添加volatile关键字，故线程A改变了number的值但是没有通知其他线程，造成其他线程无限等待。

1. 线程A和Main线程同时操作MyData对象；
2. 线程A启动，栈中记录了一个对象引用地址：MyData data。Main线程进行了相同的操作。
3. Main线程开始执行自旋锁：while，Main线程在此处等待，直到number值不在等于0
4. 3秒钟过后，线程A开始在拷贝MyData的副本到【工作内存】中，同时开始计算number的值，变为10。
5. 线程A将新值写入到主内存的堆空间的目标对象中，MyData对象的值得到更新，线程执行完成并退出。
6. 由于没有通知其他线程，主内存发生改变，所以Main线程继续等待，直到外部强制其退出。



当我们为MyData.java添加的number字段添加了volatile关键字后，这个问题得到解决：



不保证原子性

验证代码

public class MyData {

// 变量未添加volatile关键字，没有可见性

public volatile int number = 0;

public void addNumber() {

this.number ++;

}

public int getNumber() {

return number;

}

public void setNumber(int number) {

this.number = number;

}

}

public class VolatileMain {

public static void main(String[] args) {

MyData data = new MyData();

for(int i = 1 ; i <= 20 ; i ++) {

new Thread( () -> {

for(int n = 0 ; n < 1000 ; n ++) {

data.addNumber();

}

} , "线程-" + i ).start();

}

while (Thread.activeCount() > 2) {

// 等待上面20个线程计算完成，在用Main线程获取最终结果

Thread.yield();

}

// 由于期待值是：20000，但由于volatile并不具有原子性，故这里打印出正确数值的概率极小。

System.out.println(Thread.currentThread().getName() + " finally number value is : " + data.getNumber());

}

}

运行结果如下：



底层原因分析

原子性即一个操作或者多个操作，要么全部执行并且执行的过程不会被任何因素打断，要么就都不执行。但是volatile不保证这种特性，究其原因需要从底层字节码的角度来分析，在eclipse中启用javap –c后，得到MyData.java（已经去掉和此方法无关的字节码）中addNumber()方法的Java字节码如下：

Compiled from "MyData.java"

public class com.matrix.sxt.e02.MyData {

public volatile int number;

public void addNumber();

Code:

0: aload\_0

1: dup

2: getfield #12 // Field number:I

5: iconst\_1

6: iadd

**7: putfield** #12 // Field number:I

10: return

}

aload\_0：从局部变量0中装载引用类型值。在源码中：public volatile int number = 0; 这一步初始化了number=0的这个变量。

dup：复制栈顶部一个字长内容。JMM规定栈负责运行，会加载类中的实例变量和8种基本类型数据到栈内，代码中number被声明为了int类型，正好是实例变量又是8种基本类型之一。

number++ 一共进行了4步操作：

getfield：拿到原始值number。由于没有加锁所以多个线程间会抢夺，都会从主内存中拿到原始值：0。其中：Field number:I，这里的I代表number是 int类型数据。

iconst\_1：将拿到的值压入栈（当int取值-1~5时，JVM采用iconst指令将常量压入栈中）。

iadd：int型，加1次“1”；因为number++ 等价于number = number + 1。如果这里是number = number + 2，那么这里的语义解释为：int型，加1次“2”。这一步是在各自的工作内存中执行，不在主内存中执行。

**putfield**：写回主内存。

在多线程环境下，无法保证原子性的原因就出现在了这4步操作上。因为线程在CPU中以纳秒为单位在急速切换，这四步每一步都可能在当前这个线程执行的时候被CPU挂起。在iadd和putfield这两步，当前线程在自己的工作内存中加1后，写回主内存的过程中会出现脏数据覆盖。比如：线程A计算完值为1，但在putfield执行之前被挂起；线程B计算完=1，写入主内存，线程B再次计算=2，其他线程依然挂起，线程B写入主内存。此时线程A被唤醒，写入主内存，number又变为1。

解决方案

解决这个问题最简单的方式是使用AtomicInteger，原子整形会进一步引申出CAS。代码修改如下：

public class MyData {

// 变量未添加volatile关键字，没有可见性

public volatile int number = 0;

public AtomicInteger atomicInteger = new AtomicInteger();

public void addNumber() {

this.number ++;

}

public void atomicAdd() {

atomicInteger.getAndIncrement();

}

public int getNumber() {

return number;

}

public AtomicInteger getAtomicValue() {

return atomicInteger;

}

}

public class VolatileMain {

public static void main(String[] args) {

MyData data = new MyData();

for(int i = 1 ; i <= 20 ; i ++) {

new Thread( () -> {

for(int n = 0 ; n < 1000 ; n ++) {

data.addNumber();

data.atomicAdd(); // 调用原子整形

}

} , "线程-" + i ).start();

}

while (Thread.activeCount() > 2) {

// 等待上面20个线程计算完成，在用Main线程获取最终结果

Thread.yield();

}

// 由于期待值是：20000，但由于volatile并不具有原子性，故这里打印出正确数值的概率极小。

System.out.println(Thread.currentThread().getName() + " finally number value is : " + data.getNumber());

System.out.println(Thread.currentThread().getName() + " atomic integer value is : " + data.getAtomicValue());

}

}

运行结果如下：

DCL双端检锁机制单例优化 – 实例

public class SingletonVolatile {

private static **volatile** SingletonVolatile instance = null;

private SingletonVolatile() {

System.out.println(Thread.currentThread().getName() + " SingletonVolatile私有构造函数实例化");

}

// DCL Double check lock 双端捡锁

public static SingletonVolatile getInstance() {

if(instance == null) {

synchronized (SingletonVolatile.class) {

if(instance == null) {

instance = new SingletonVolatile(); // 双端捡锁问题点

}

}

}

return instance;

}

public static void main(String[] args) {

for (int i = 1; i <=1000; i++) {

new Thread(() ->{ SingletonVolatile.getInstance(); } , "当前线程：" + String.valueOf(i) ).start();

}

}

}

双端检锁机制不一定线程安全，原因是有指令重排的存在，加入volatile可以禁止指令重排从而达到线程安全。原因在于某一个线程在执行到第一次检测，读取到的instance不为null时，instance的引用对象可能没有完成初始化。instance = new SingletonVolatile(); 这段代码的执行可以分解为以下3个步骤：

1. memory=allocate(); 分配对象内存空间
2. instance(memory); 初始化对象
3. instance=memory; 设置instance的指向刚分配的内存地址,此时instance!=null

步骤2和步骤3不存在数据依赖关系。而且无论重排前还是重排后程序执行的结果在单线程中并没有改变，因此这种重排优化：1->3->2是允许的。但是指令重排只会保证串行语义的执行一致性(单线程) 并不会关心多线程间的语义一致性，所以当一条线程访问instance不为null时，由于instance实例未必完成初始化，也就造成了线程安全问题。

原子引用

原子引用是对volatile的进一步完善和封装，他提供了volatile不具备的原子性，但没有使用线程锁synchronized，所以系统并发性不受影响。

Compare and Swap(CAS)【实际运用】

CAS：CPU原子操作

即：**比较并交换**。多个线程工作内存中的“**一个**”变量A(堆中的副本拷贝)和主内存中的这个变量A进行比较，如果期望值与真实值一致，那么更新为当前线程指定的值；CAS只能操作一个变量，是一条CPU并发原语，它的功能是判断内存某个位置的值是否为预期值，如果是则更新为新的值，这个过程是原子的。

**如何保证原子性**？

CAS并发原语体现在Java语言中是rt.jar包中的sun.misc.Unsafe.java（*oracle未公布rt.jar中的jdk包和sun包下的源代码*）类中的各个方法。调用Unsafe类中的CAS方法，JVM会帮我实现CAS汇编指令，这是一种完全依赖于硬件功能，通过它实现了原子操作。关于**系统原语**：原语属于操作系统用途范畴，是由若干条指令组成，用于完成某个功能的一个过程，并且原语的执行必须是连续的，在执行过程中不允许中断（原语的执行必须是连续的，就像国家主席车队出行会出现交通管制，绝对不允许社会车辆插队）。也就是说CAS是一条原子指令，不会造成所谓的数据不一致的问题。

比如本段落中的“Unsafe类中的CAS方法”具体指：public final native boolean compareAndSwapInt(Object o, long offset, int expected, int x); 他是一个由native修饰的本地方法，该方法的实现位于：unsafe.cpp中，其汇编指令如下：



这其中：Atomic::cmpxchg…事实上在汇编层面，只要是以Atomic开头的，都是原子操作，不允许插队。

AtomicInteger.java源代码解析

public class CasDemo {

public static void main(String[] args) {

AtomicInteger ai = new AtomicInteger(5); // 初始化的值为5(否则默认为0)，此值会放到主物理内存(堆中)

ai.getAndIncrement(); // 进行自增加1操作后，变为6。此方法极具CAS代表性，后面细聊。

boolean flag1 = ai.compareAndSet(6, 2020); // 堆中的变量为6，线程期望主内存的值为6，那么更新为新的值：2020

boolean flag2 = ai.compareAndSet(6, 2021); // 堆中的变量为2020，线程期望主内存的值为6，由于不一致，所以更新为2021的时候失败。

System.out.println(flag1 + " 当前值：" + ai.get());

System.out.println(flag2 + " 当前值：" + ai.get());

}

}

运行如下：



如果线程的期望值和主物理内存的真实值一致，那么则修改更新值；如果不一致，那么修改失败，重新获得主物理内存的真实值。CAS的底层原理是自旋锁和Unsafe.java。

打开AtomicInteger.java底层源代码，基本是对Unsafe类的封装。

**public class** AtomicInteger **extends** Number **implements** java.io.Serializable {  
 **private static final long serialVersionUID** = 6214790243416807050L;

**private volatile int value**; // 提供可见性与顺序性保障

**private static final** Unsafe **unsafe** = Unsafe.getUnsafe(); // 提供原子性保障

**private static final long valueOffset**;  
**static** {  
 **try** {

// 使用Unsafe类获取一个字段的内存偏移量（即内存地址），同时赋值给valueOffset字段。此处为AtomicInteger.java的value字段

// 后面会通过获取valueOffset这个内存地址，操作该内存地址所对应的值。

**valueOffset** = **unsafe**.objectFieldOffset(AtomicInteger.**class**.getDeclaredField(**"value"**));  
 } **catch** (Exception ex) {  
 **throw new** Error(ex);  
 }  
}

……

}

当查看AtomicInteger.java的源码的时候会发现，他首先获取了Unsafe类，然后使用volatile关键字修饰**一个**私有实例变量。AtomicInteger拥有JMM规定的3个基本特性：可见性、有序性和原子性；这其中可见性和有序性由volatile关键字提供保障，而原子性则由Unsafe类来完成。其中valueOffset是内存地址偏移量，他是Unsafe类要操作的指针。Unsafe类相当于一个后门，基于该类可以操作特定内存的数据，其内部方法均为native修饰，可以像C语言的指针一样直接操作物理内存。也就是说Unsafe类中的方法都直接调用操作系统底层资源执行相应任务。

AtomicInteger.java中的getAndIncrement()方法，非常具有CAS的代表性，开始进一步拆解：

这里的代码解析要结合多线程来理解。我们假设有2个线程A和B，这两个线程启动后，同时操作

（AtomicInteger ai = new AtomicInteger(5)）同一个AtomicInteger()对象：ai，并且调用ai.getAndIncrement()方法。线程A先执行，它在操作ai对象+1时，ai对象使用自旋锁，

**第1步**：v = getIntVolatile(o , offset); 获取当前对象（即：ai）在主内存中的当前值作为“期望值”；**第2步**：调用compareAndSwapInt方法，他的第二个参数是AtomicInteger的类变量value在堆中的地址，可以直接拿到最新值。但是这两步并不是原子操作，所以第1步中拿到值后，线程A可能被挂起，线程B会整个执行完成，从而将堆内存中的value值变成6。这时线程A继续执行第2步，

compareAndSwapInt方法从堆中再次取出value的值，发现是6，不是线程A期望的值5，同时返回

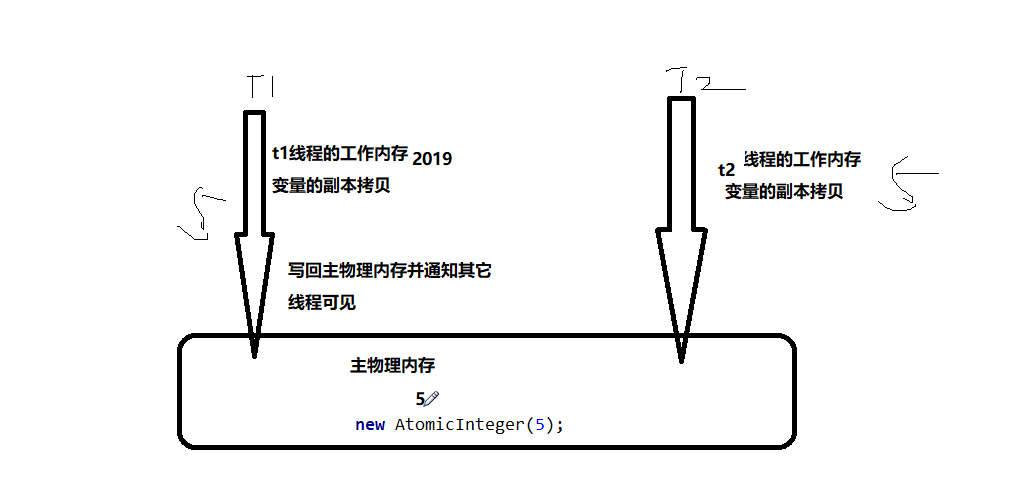
false，!false=true，故自旋锁do-while继续生效执行，线程A再次通过getIntVolatile方法，从堆中拿到最新的值6，再次调用compareAndSwapInt进行比较，发现value在堆中的值和线程A的期望值是一样的，那么将执行原子操作：v+1。

**注意**：

1. compareAndSwapInt方法是原子性的。
2. getIntVolatile方法和compareAndSwapInt方法前两个参数是一样的；都是在通过操作当前对象的类变量所在的内存地址的方式来获取值。他们都在尝试获取AtomicInteger的类变量private volatile int value的值。new AtomicInteger(5)的时候会在堆中为value分配一个地址，并指定值为5。



https://blog.csdn.net/u012516166/article/details/80529948



自旋锁

ABA问题

原子引用更新

ABA问题如和解决

Unsafe【实际运用】

https://www.jianshu.com/p/db8dce09232d

https://www.cnblogs.com/suxuan/p/4948608.html

synchronized-线程锁

Java Virtual Machine实际应用

Html/Css/Javascript 篇

Servlet/J2EE篇

实用代码篇

* 1. Java代码段

1.1 十六进制和八进制

* 1. Javascript代码段
  2. Mysql代码段

实际问题收集

1 SpringMvc线程安全问题



springmvc是单例模式的框架,但它是线程安全的,因为springmvc没有成员变量,所有参数的封装都是基于方法的,属于当前线程的私有变量. 因此是线程安全的框架

关于java成员变量和局部变量

https://blog.csdn.net/weixin\_37012881/article/details/82699089



实例变量属于某个对象的属性，必须创建了实例对象，其中的实例变量才会被分配空间，才能使用这个实例变量。

https://blog.csdn.net/yong\_zi/article/details/81285049







2 ThreadLocal

https://www.jianshu.com/p/3c5d7f09dfbd

3 HashMap算法

4 Zookeeper

https://blog.csdn.net/qq\_36864672/article/details/78596499

5 Redis

https://mp.weixin.qq.com/s/f9N13fnyTtnu2D5sKZiu9w

6 SPI 技术

https://mp.weixin.qq.com/s/VrqOpyTTIzlyMAtJFda9fw

7 乐观锁和悲观锁



https://segmentfault.com/a/1190000015239603

JVM类的双亲加载

代码样式

|  |  |  |
| --- | --- | --- |
| 方法名 | 返回值 | 功能描述 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
| 方法名 | 功能描述 |
|  |  |
|  |  |