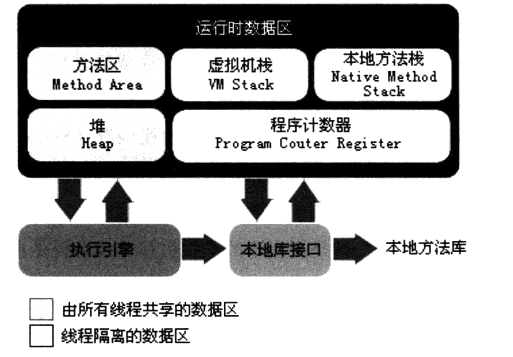
# JVM篇

## 1.1 Java内存和常见异常

运行时数据区域，如下图所示：



**程序计数器：**

* 用来记录指令的执行
* 协助完成线程的切换（时间分片）
* 线程私有的，伴随着线程生死
* 可能的异常未内存溢出异常

**Java虚拟机栈：**

* 线程私有的
* Java方法的执行会在虚拟机栈中形成一个栈帧
* 可能的异常为内存溢出和栈溢出

**堆：**

* 存放对象的实例
* GC区
* 线程共享的区域

本地方法栈：

方法区：

## 1.2 垃圾回收

### 1.2.1 对象死亡的判断

#### 1.2.1.1 引用计数算法

循环引用问题

#### 1.2.1.2 根搜索算法

#### 1.2.1.3 四种引用

#### 1.2.1.4 死亡标记与自救

1、 首先是无根的

2、 进行第一次标记

3、 进入队列

4、 Finalize() ，可以进行自救 oblect = this

### 垃圾收集

#### 标记清除算法

#### 复制算法

常用于新生代

#### 标记整理算法

常用于老年代

#### 分代收集算法

新生代：复制算法

老年代：标记整理

### 1.2.1 常见的垃圾收集器

## 1.3 内存分配

内存特点：

老年代

新生代

### 1.3.1 优先在Eden区分配

* 担保机制下向老年代的转移
* FullGC和MinorGC的区别

### 1.3.2 老年代的特点

* 大对象分配内存时会直接进入老年代
* 长期存活的对象会进入老年代（对象年龄）

### 1.3.3 空间分配担保

长期存活在新生代的大对象，需要老年代提供分配担保进入老年代

# 多线程篇

## 2.1 并发编程的挑战

### 上下文切换的原理

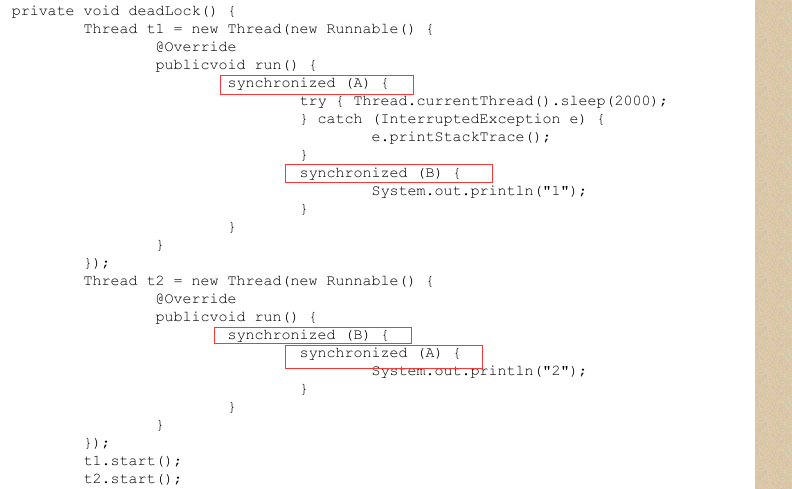
Cpu划分时间片

### 减少上下文切换的方法

* 无锁并发编程
* **使用CAS**
* 使用协程

### 死锁

代码案例：



原因：

**你中有我，我中有你**

**措施：**

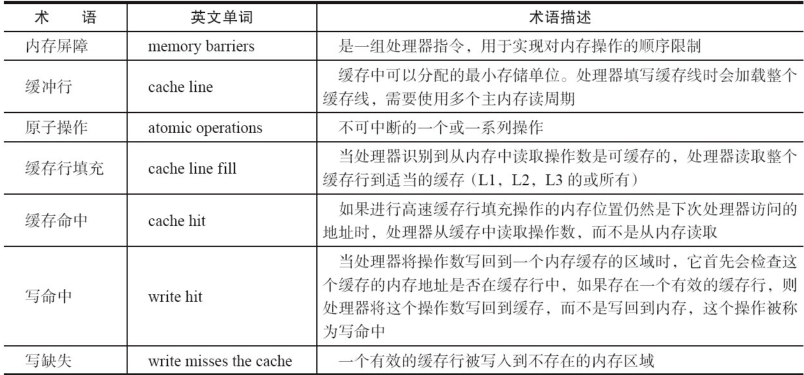
* 使用超时锁，锁的自动释放
* 数据库使用锁，保证获取锁与释放锁在同一个连接里面
* 避免消费一个资源的线程获取多个锁（一个锁只占用一个资源）

## 底层实现原理

### Volatile：

* 1、将缓存的数据写回内存中；2、是其他CPU缓存失效

CPU术语定义：



线程03

线程02

线程01

**缓存失效策**略

缓存

缓存

缓存

总线锁会所著所有资源，消耗较大

内存

### Synchronized：

不同情况下使用syn关键字时具体的锁对象：

1. 同步方法：调用当前方法的对象
2. 同步代码快：括号里面的参数
3. 静态同步方法：Class对象

几种锁类型：

偏向锁（锁的重入）：实现原理，对象头部记录偏向锁的ID.偏向锁的释放。适用于一个资源被同一个线程多次访问的情况

轻量级锁：线程快速切换，代码块执行花费时间短的

重量级锁：追求量，代码快执行时间较长

Java对象的头存储结构：

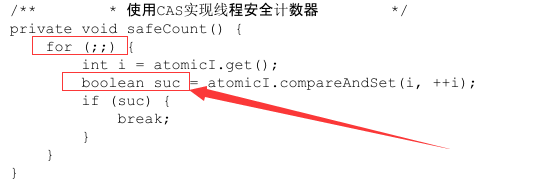
对象的hashcode值、分代年龄值、锁标记

**原子操作的实现原理：**

1. **内存总线锁（当资源数据不在同一个缓存行的时候）**
2. **缓存锁**

**JAVA中实现锁：**

1. **加锁**
2. **通过CAS操作（ABA问题，解决方案是加上一个版本号码；自旋时间过长）**



## 内存模型

通信

线程

线程

同步

缓存

缓存

内存

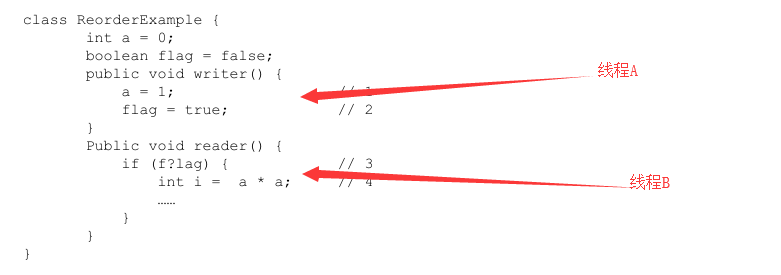
**核心问题：线程间的通信和数据同步**

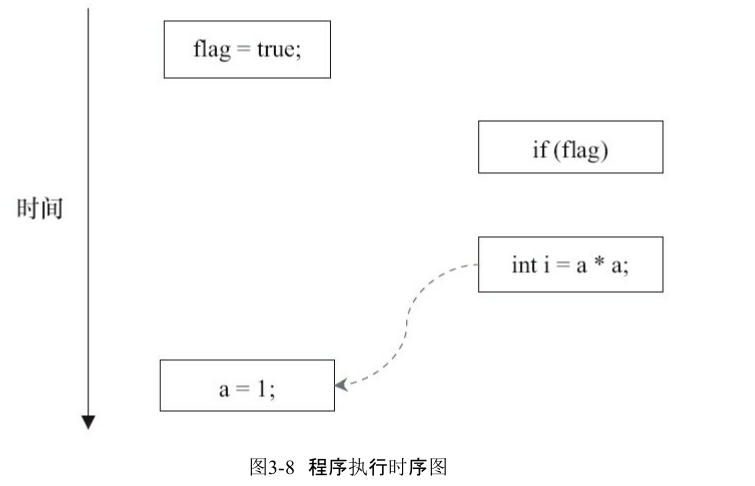
### 2.3.1 指令重排序

重要的是指令重排序不会导致结果的改变

Happen-before原则

**重排序对多线程的影响：**





### 2.3.1 顺序一致性

线程A

线程B

内存中顺序执行

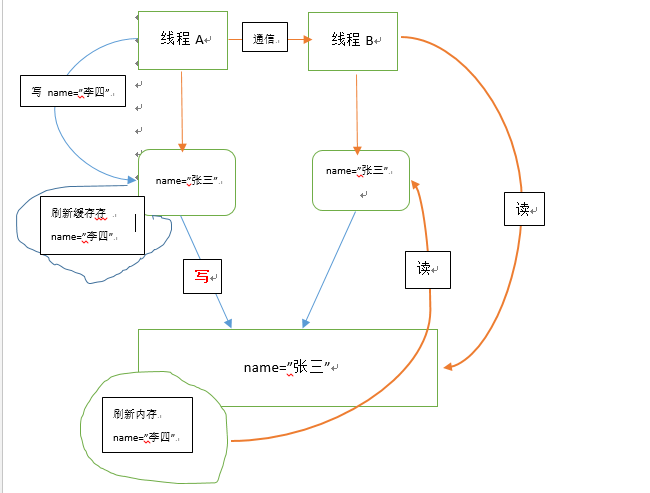
### 2.3.3 Volatile的内存语义

首先遵循happens-before原则。即保证锁的释放和锁的获取之间的内存可见性。在volatite变量中，意味着对volatitle变量的写可以让任意线程（读这个volatitle变量的线程）看到最新的写入结果。

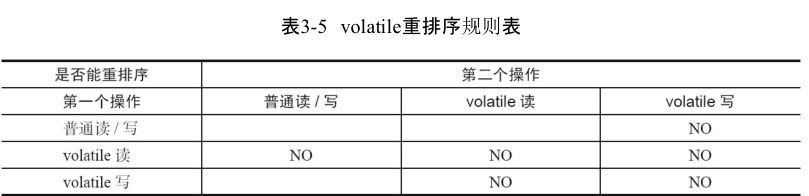
内存图解：

volatitle String name =”张三”

**volatitle 的读写内存模型：**



关于内存屏障的知识点：



### 2.3.4 锁的内存语义

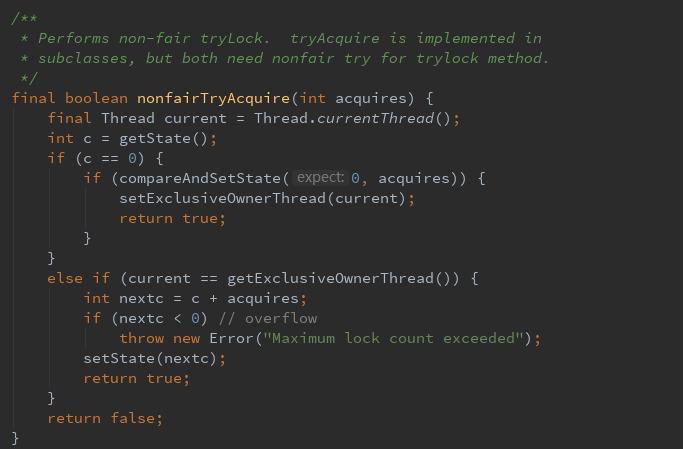
锁的获取与释放，对应着数据的写入与读取

### 2.3.5 ReentrantLock源码分析

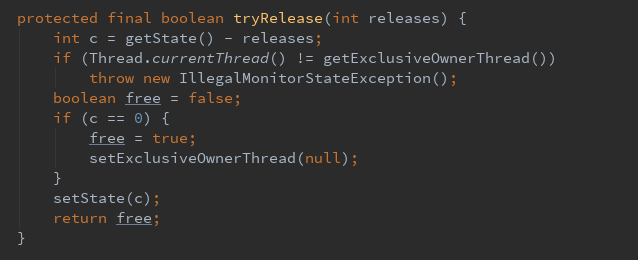
使用模型：

public class ReentrantLockExample {  
 ReentrantLock reentrantLock =new ReentrantLock();  
  
 private String appName = "安卓";  
 public static void main(String[] args) {  
  
 }  
  
 public void write(String name) {  
 reentrantLock.lock();  
 try {  
 this.appName = name;  
 } finally {  
 reentrantLock.unlock();  
 }  
 }  
  
 public String read() {  
 reentrantLock.lock();  
 try {  
 return this.appName;  
 } finally {  
 reentrantLock.unlock();  
 }  
 }  
  
}

**加锁：**



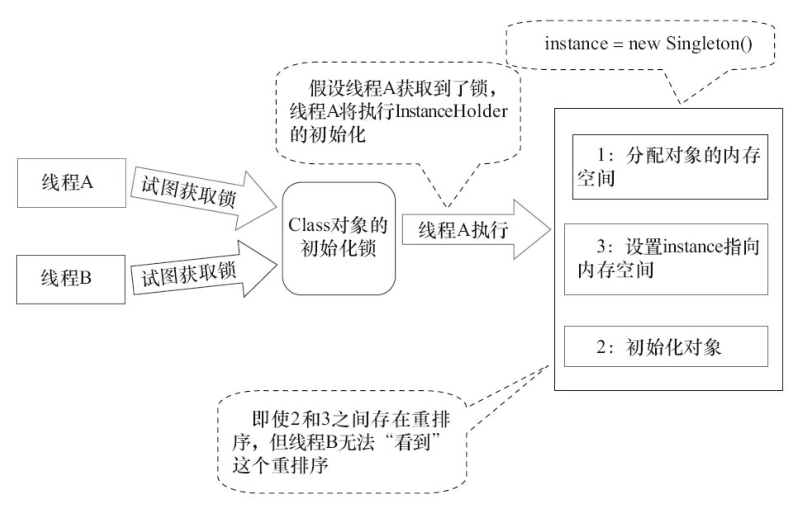
**解锁：**



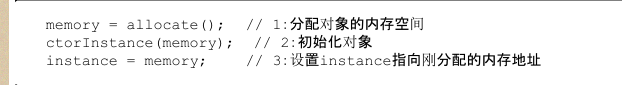
### 2.3.6 双重检查

涉及到的知识点：

1. 类初始化的时候会加锁

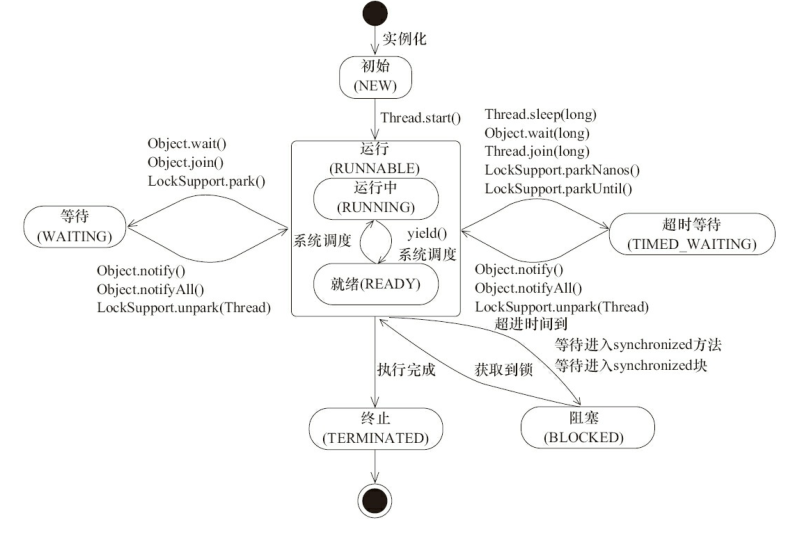


1. Object object = new Object();是经历了三个步骤



## 2.4 并发编程的基础

### 2.4.1 线程的状态转换



1、New 一个线程

2、Start一个线程（是否获得锁/获得锁后是否可以拿到时间分片）

3、2满足的情况下，调用wait()会释放锁，调用sleep()会睡眠但不释放锁

### 线程的相关知识点

* 守护线程
* 线程的终止策略
* 线程的中断

**中断的实战代码：**

public class Interrupted {  
  
 public static void main(String[] args) throws Exception{  
 Thread sleep = new Thread(new SleepThread(),"SleepThread");  
 Thread work = new Thread(new WorkThread(),"WorkThread");  
 sleep.setDaemon(true);  
 work.setDaemon(true);  
 sleep.start();  
 work.start();  
 Thread.*sleep*(2000);  
 sleep.interrupt();  
 work.interrupt();  
 */\*\*  
 \* SleepThread:false  
 \* WorkThread:true  
 \*/* System.*out*.println(sleep.getName()+":"+sleep.isInterrupted());  
 System.*out*.println(work.getName()+":"+work.isInterrupted());  
 }  
 private static class SleepThread implements Runnable {  
 @Override  
 public void run() {  
 try {  
 Thread.*sleep*(1000);  
 } catch (InterruptedException e) {  
 e.printStackTrace();  
 }  
 }  
 }  
 private static class WorkThread implements Runnable {  
 @Override  
 public void run() {  
 while(true){  
 }  
 }  
 }  
}

# 网络通信

# 3.1.1 TCP/UDP

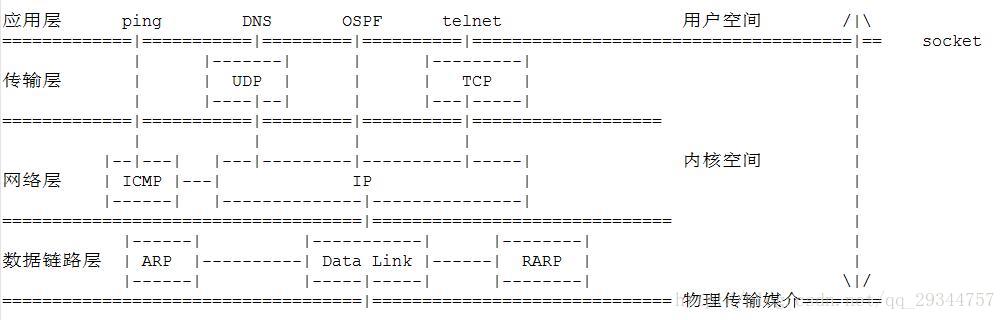
基本区别：

基于连接与无连接  
TCP要求系统资源较多，UDP较少；   
UDP程序结构较简单   
流模式（TCP）与数据报模式(UDP);   
TCP保证数据正确性，UDP可能丢包   
TCP保证数据顺序，UDP不保证

UDP应用场景：

面向数据报方式  
网络数据大多为短消息   
拥有大量Client  
对数据安全性无特殊要求  
网络负担非常重，但对响应速度要求高

### 3.1.2 TCP/IP协议四层模型



应用层(HTTP请求报文)

传输层（UDP/TCP+HTTP请求报文）

网络层（IP+UDP/TCP+HTTP请求报文）

数据链路层（MAC+IP+UDP/TCP+HTTP请求报文）

物理层

**IP:源IP地址，目标IP地址**

**TCP: 目标端口**

**TCP的握手与挥手（面试重点）**

经典好文：<https://juejin.im/post/5c6fbf54f265da2db718216a>

### 3.1.2 流量控制



接收者

发送端

缓冲区

滑动窗口实现了TCP流控制。首先明确滑动窗口的范畴：TCP是双工的协议，会话的双方都可以同时接收和发送数据。TCP会话的双方都各自维护一个发送窗口和一个接收窗口。各自的接收窗口大小取决于应用、系统、硬件的限制（TCP传输速率不能大于应用的数据处理速率）。各自的发送窗口则要求取决于对端通告的接收窗口，要求相同。

滑动窗口解决的是流量控制的的问题，就是如果接收端和发送端对数据包的处理速度不同，如何让双方达成一致。接收端的缓存传输数据给应用层，但这个过程不一定是即时的，如果发送速度太快，会出现接收端数据overflow，流量控制解决的是这个问题。

**滑动机制**

* 发送窗口只有收到发送窗口内字节的ACK确认，才会移动发送窗口的左边界。
* 接收窗口只有在前面所有的段都确认的情况下才会移动左边界。当在前面还有字节未接收但收到后面字节的情况下，窗口不会移动，并不对后续字节确认。以此确保对端会对这些数据重传。
* 遵循快速重传、累计确认、选择确认等规则。
* 发送方发的window size = 8192;就是接收端最多发送8192字节，这个8192一般就是发送方接收缓存的大小。

<https://www.cnblogs.com/woaiyy/p/3554182.html>

### 3.1.3 Socket编程模型

### 3.1.4 Http协议

1、Http请求

请求头

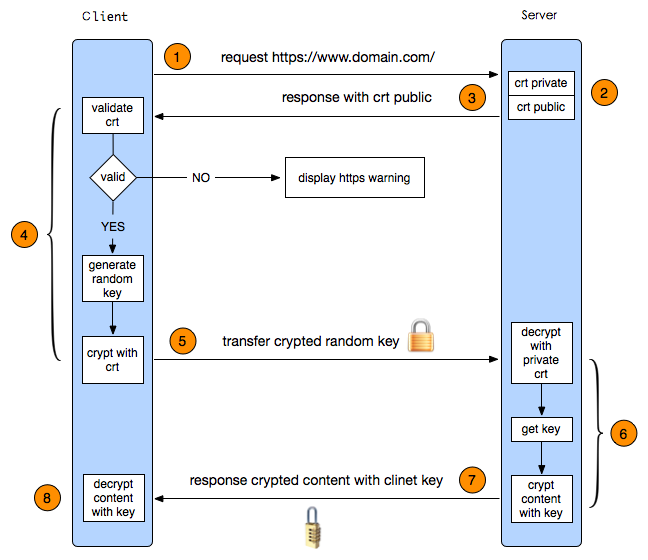
请求内容

请求体

2、Https请求

3、Http的无状态特性

4、Https通信的加密算法



|  |  |
| --- | --- |
|  | -> 客户端向服务端发送请求  -> 服务端返回数字证书  -> 客户端用自己的CA[主流的CA机构证书一般都内置在各个主流浏览器中]公钥去解密证书,  如果证书有问题会提示风险  -> 如果证书没问题客户端会生成一个对称加密的随机秘钥然后再和刚刚解密的服务器端的公  钥对数据进行加密,然后发送给服务器端  -> 服务器端收到以后会用自己的私钥对客户端发来的对称秘钥进行解密  -> 之后双方就拿着这个对称加密秘钥来进行正常的通信 |

5、Cookie与Session

### 3.1.5 Web攻击手段

## 3.2 IO

一些基本概念：

**同步与异步（客户端）**

* 同步： 同步就是发起一个调用后，被调用者未处理完请求之前，调用不返回。
* 异步： 异步就是发起一个调用后，立刻得到被调用者的回应表示已接收到请求，但是被调用者并没有返回结果，此时我们可以处理其他的请求，被调用者通常依靠事件，回调等机制来通知调用者其返回结果。

**阻塞和非阻塞（服务端）**

* 阻塞： 阻塞就是发起一个请求，调用者一直等待请求结果返回，也就是当前线程会被挂起，无法从事其他任务，只有当条件就绪才能继续。
* 非阻塞： 非阻塞就是发起一个请求，调用者不用一直等着结果返回，可以先去干其他事情。

### 3.2.1 BIO

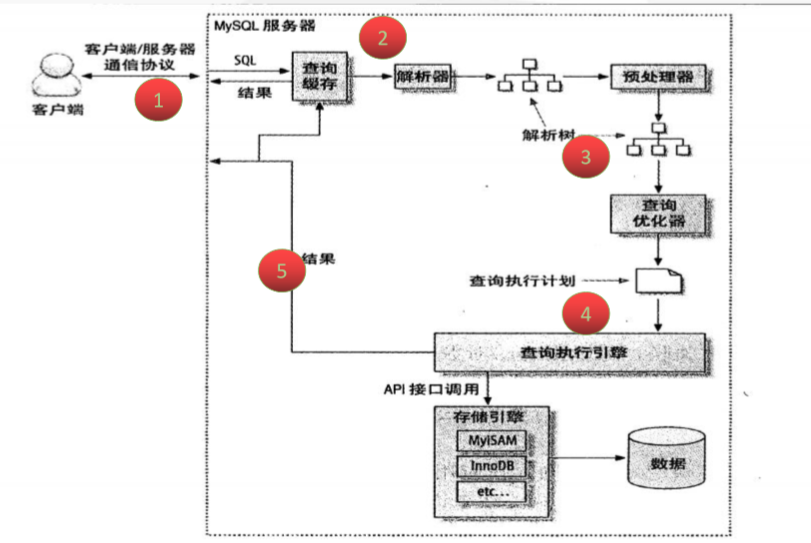
### 3.2.2 NIO

### 3.2.3 AIO

# 4. 数据库

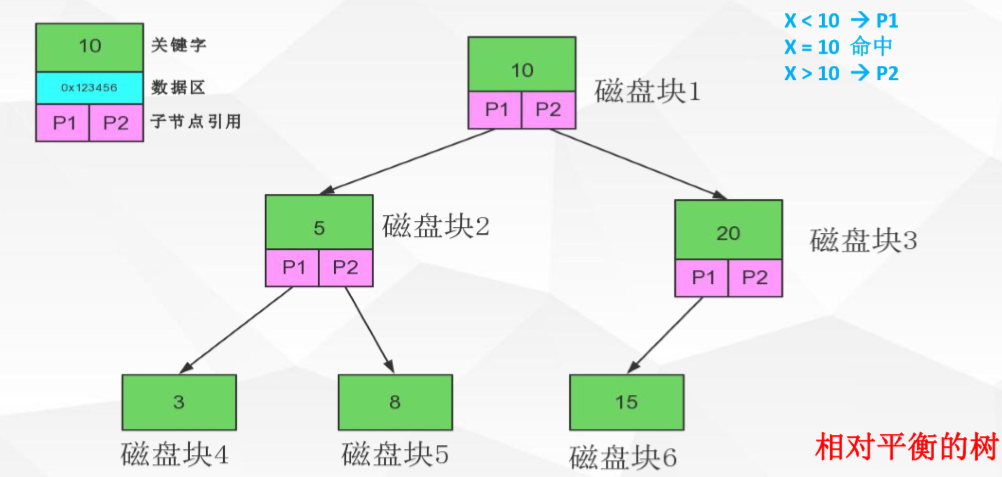
## 4.1 Mysql





### 4.1.1 Mysql的索引

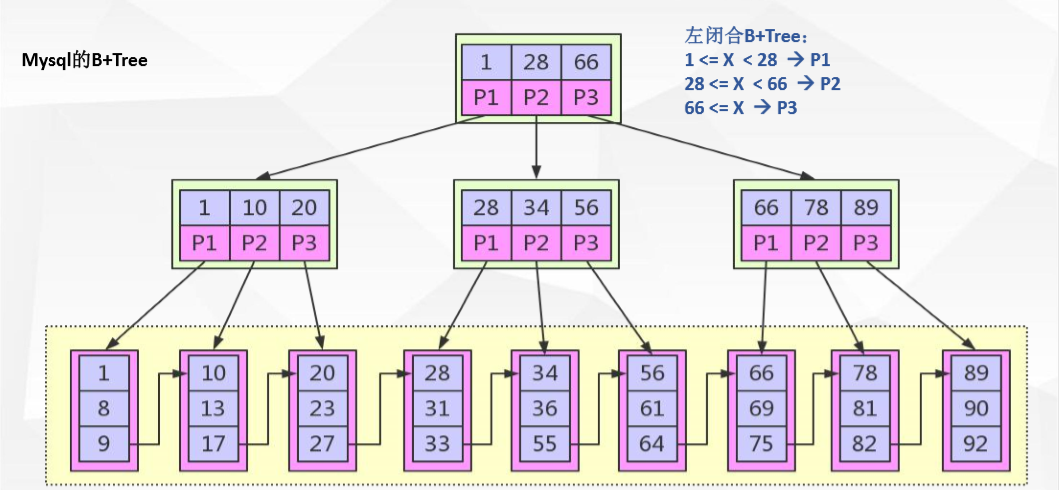
二叉查找树，结构如下：



每个节点的组成：

* 1. 关键字
  2. 数据区域
  3. 子节点引用

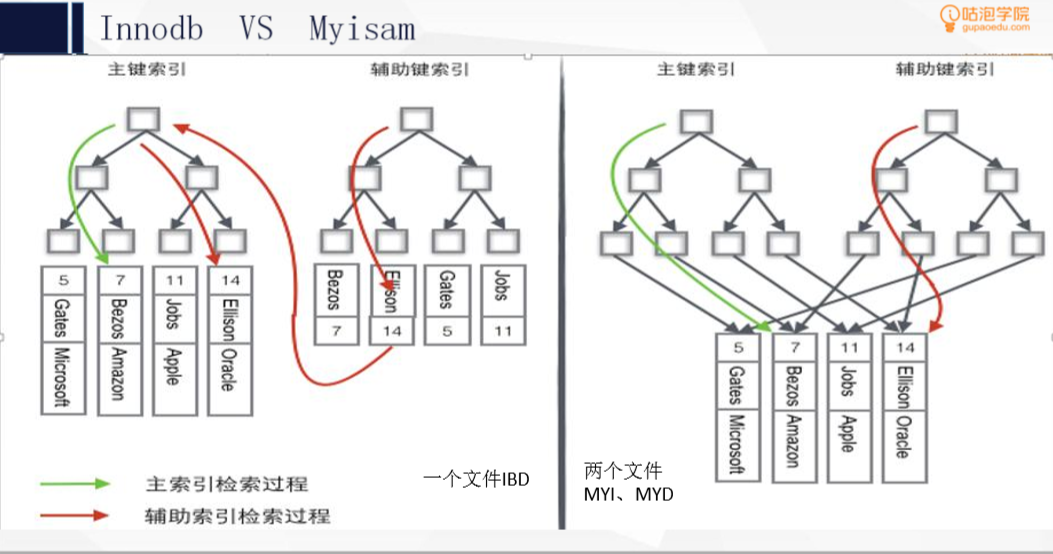
加强的多路平衡树：



特点：

1. 采用左闭合策略
2. 非叶子节点不保存数据，只保存子节点指针和数据区间关键字
3. 真正的数据保存在叶子节点
4. 叶子节点是顺序的，并且有引用关系

**InoDB的二级索引**



**索引的一些注意点：**

1. **作为索引的字段尽量离散**
2. **最左匹配原则**
3. **联合索引**

### 4.1.2 Mysql的锁

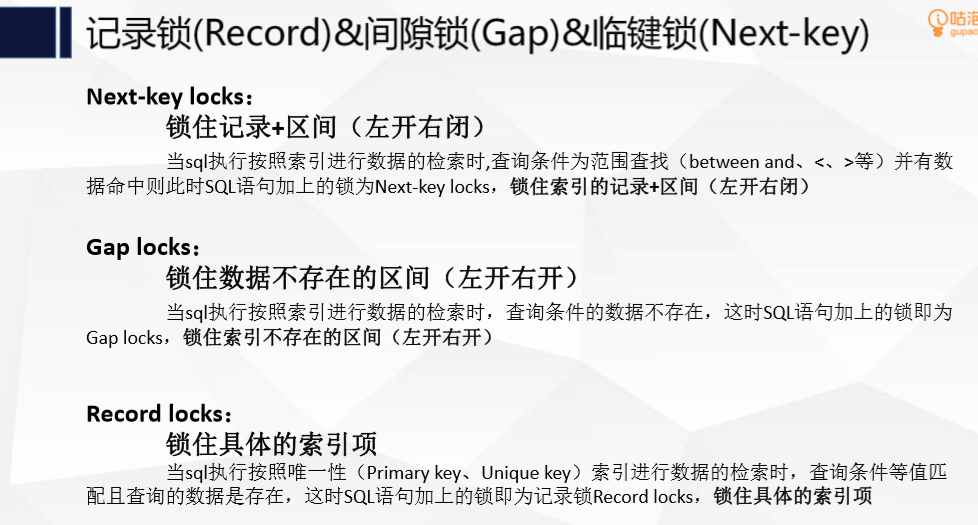
**InnoDB引擎的锁类型**

1. **共享锁**
2. **排他锁**
3. **意向共享锁**
4. **意向排他锁**
5. **自增锁**

**InnoDB行锁的实现：**

通过对索引加锁来实现。

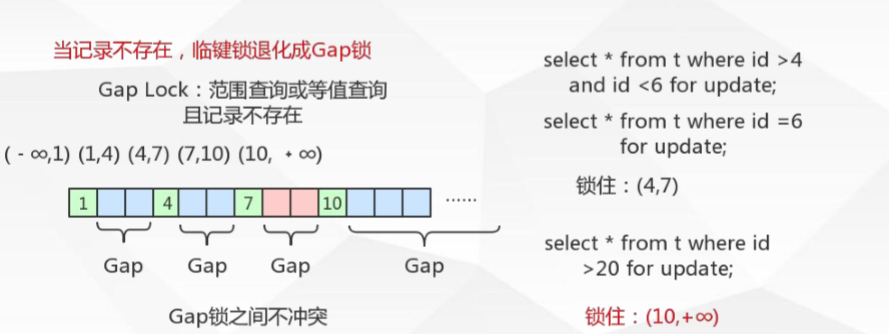
如果不走索引的时候会走表锁



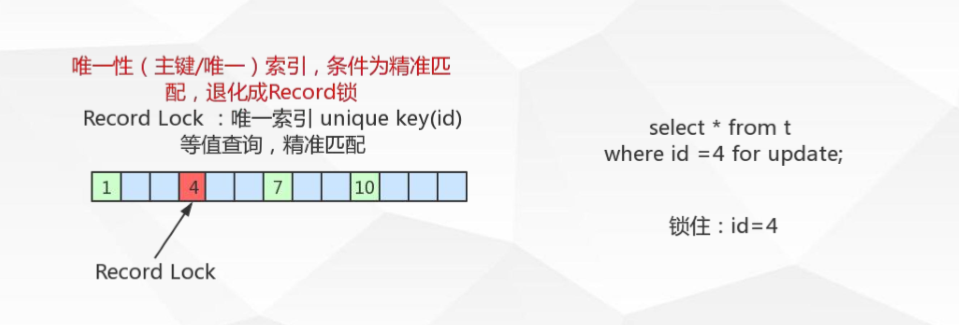
临键锁：



间隙锁：



记录锁：



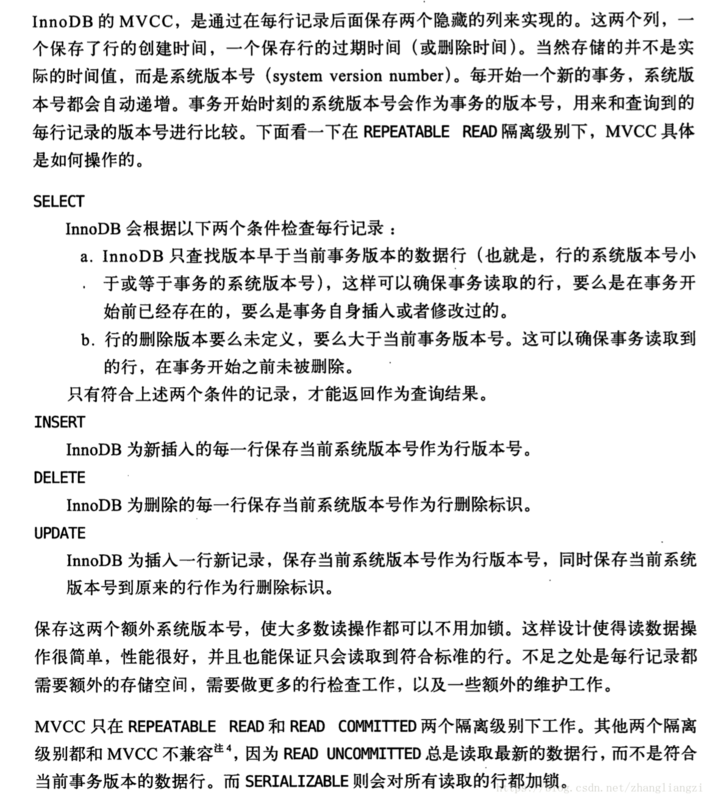
### 4.1.3 MMVC(多版本并发控制)

**普通话解释： 并发访问(读或写)数据库时，对正在事务内处理的数据做 多版本的管理。以达到用来避免写操作的堵塞，从而引发读操 作的并发问题。**

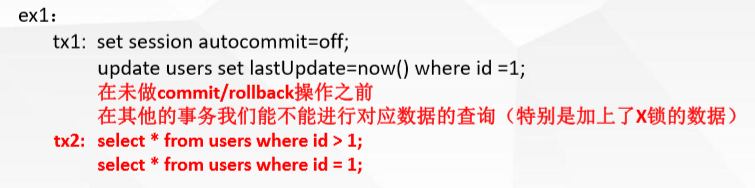


数据行的版本号：全局事物唯一ID

删除版本号：全局事物唯一ID



**假设当前事物版本号码为88**



**事物1**

**Update操作：**

1. 复制原始记录，版本控制号码设置成为 88重新插入
2. 将原始记录的删除版本号设置为88

**事物2**

**快照读：**

SQL读取的数据是快照版本，也就是历史版本，普通的SELECT就是快照读

innodb快照读，数据的读取将由 cache(原本数据) + undo(事务修改过的数据) 两部分组成

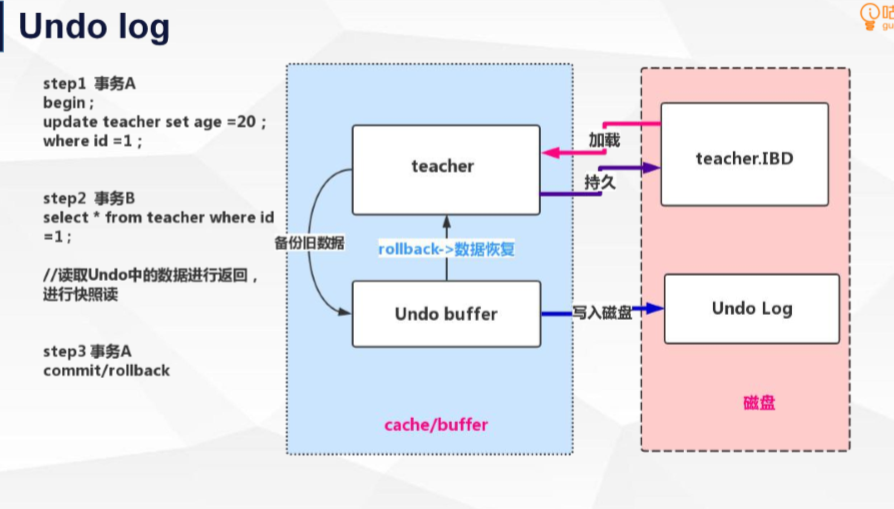
**当前读：**

SQL读取的数据是最新版本。通过锁机制来保证读取的数据无法通过其他事务进行修改

UPDATE、DELETE、INSERT、SELECT … LOCK IN SHARE MODE、SELECT … FOR UPDATE都是当前读

### 4.1.4 日志文件

undo日志用于存放数据修改被修改前的值



用来做数据恢复，主要记录逻辑操作（到每一行的操作）

## 4.2 Redis

## 4.3 MongoDB

# 5.分布式

## 序列化

* Java本身提供的系列化
* 基于Xml的序列化

### Java本身自带的序列化

1. ByteArrayOutputStream byteArrayOutputStream = new ByteArrayOutputStream();  
   try {  
    ObjectOutputStream objectOutputStream = new ObjectOutputStream(byteArrayOutputStream);  
    objectOutputStream.writeObject(person);  
    byteArrayOutputStream.toByteArray();  
     
   } catch (IOException e) {  
    e.printStackTrace();  
   }

<https://www.jianshu.com/p/61f5038df016>

需要关注的点：

* 序列化实现深度克隆
* Transient声明避免序列化
* 实体需要实现writeObject(ObjectOutputStream obj)、readObject( ObjectInputStream objectInputStream )）的方法
* 静态属性的序列化：序列化保存的是对象的状态，静态变量属于类的状态，因此 序列化并不保存静态变量。

### 5.1.2 Xml序列化

核心辅助累 XStream

### 5.1.3 Json实现序列化

### 5.1.4 Hession序列化

### 5.1.5 ProtoBuf序列化

1、BitMap/BitSet的算法原理

Int 4\*8=32

## 5.2 RPC

远程过程调用

### 5.2.1 RMI

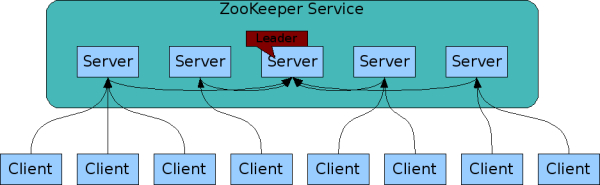
原理：服务端和客户端，服务端注册到一个网络

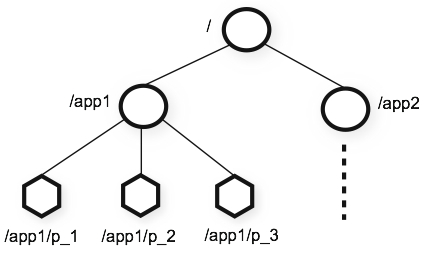
### 5.3.1 Zookeeper

分布式锁

联想在进程中的多线程共享资源

#### 架构及其数据存储：





zookeeeper的集群

处理事务请求（增 删 改）

Marter

事物请求

数据同步

**slave**

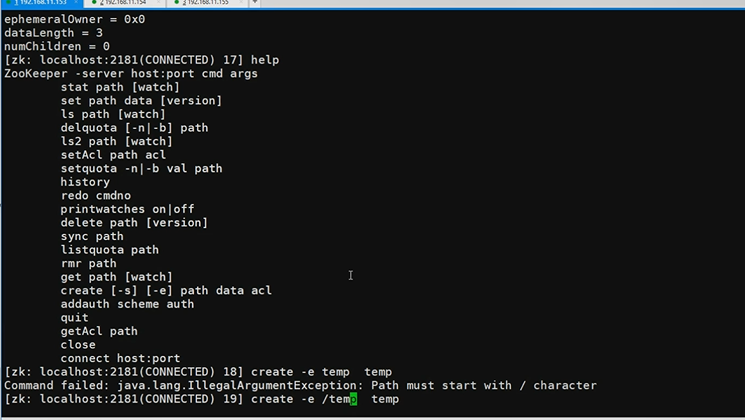
Observer

slave

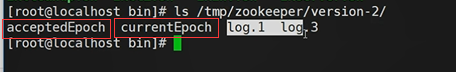
数据同步： 2pc 最红一致性

Zookeeper的客户端可以查看相关的数据存储信息

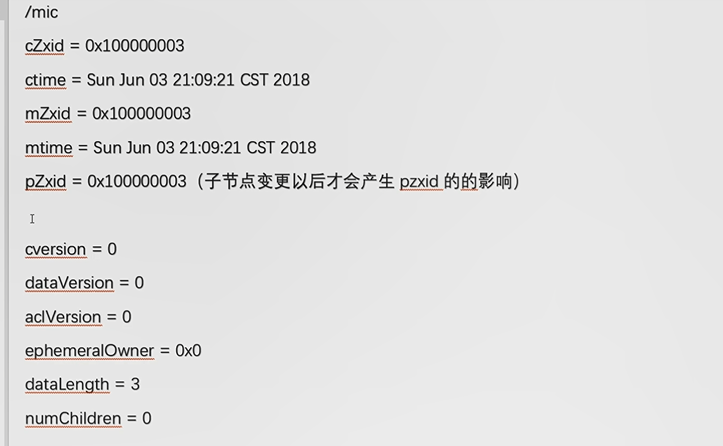
* 有序节点
* 临时节点和持旧节点
* 同级节点的唯一性
* 树形的存储结构



DataDir：



Zookeeper中文件得具体内容



#### Observera集群监控

ZK

#### 应**用模型**

文件存储模型

OrderService1

OrderService2

OrderService3

#### ZK设计实现分析

* 1. 单点故障问题
  2. 数据同步问题（改进版的2pc）
  3. 数据一致性问题
  4. 事物机制（过半节点ACK即可）
  5. 选举机制

2PC模型

1

1

参与者2

协调者

参与者1

yes

yes

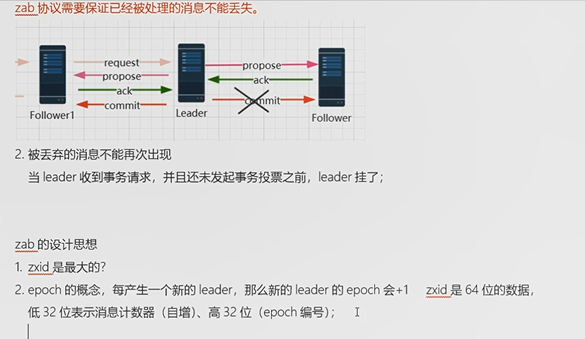
Commit

Commit

#### 崩溃恢复

1. Leader挂掉后的机制

Zxid最大的处理方法



**选举Leader的比较依据：**

Myid/zxid/epoch

**Zk节点的状态：**

**LOKING/LEADING/FLOWING/OBSERVERING**

#### 实现分布式锁

惊群效应

Zookeeper

ClientA

获得锁的节点

/Lock

**监听**

ClientB

**监听**

ClientC

**监听**

ClientD

#### 

惊群效应

Zookeeper

ClientA

获得锁的节点

/Lock

**监听**

ClientB

/Lock

**监听**

ClientC

/Lock

**监听**

ClientD

/Lock

**可以具体参考Java多线程编程中的实现**

**具体代码：**<https://www.cnblogs.com/wuzhenzhao/p/9996522.html>

# 6. 大数据专题

常用命令：

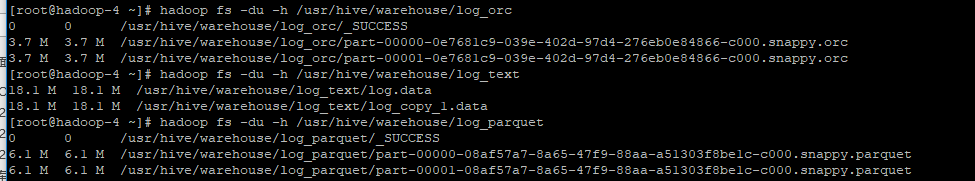
hadoop dfs -du -h 查看各级文件大小

## 6.1 数据存储格式

<https://www.cnblogs.com/eRrsr/p/6098454.html>

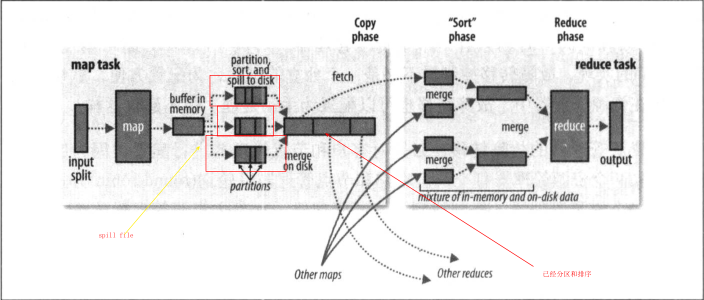
<https://www.cnblogs.com/xiohao/p/6421478.html>

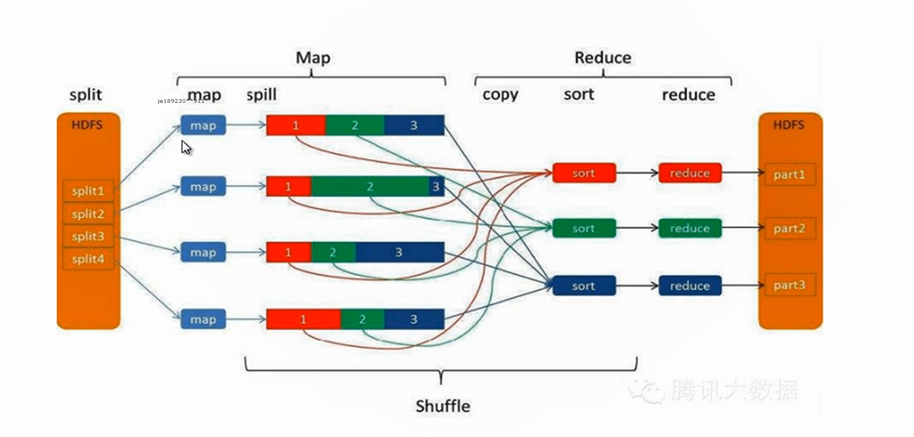
<https://www.jianshu.com/p/694f044d1c34>



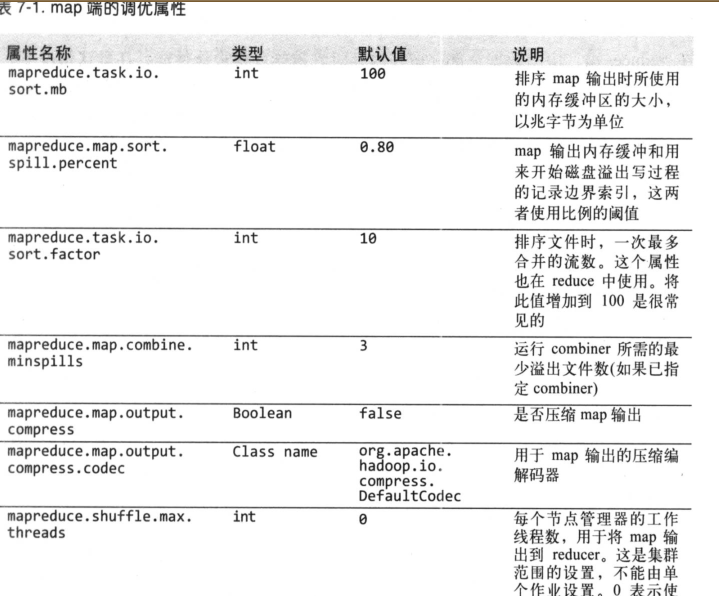


## 6.2 MR原理

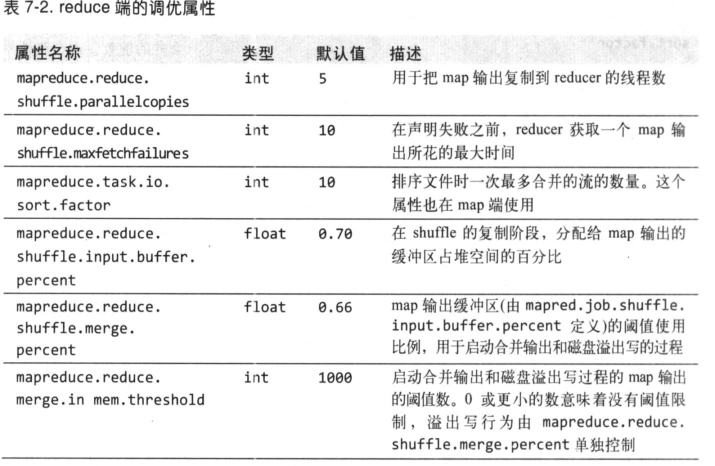




MAP端调优参数



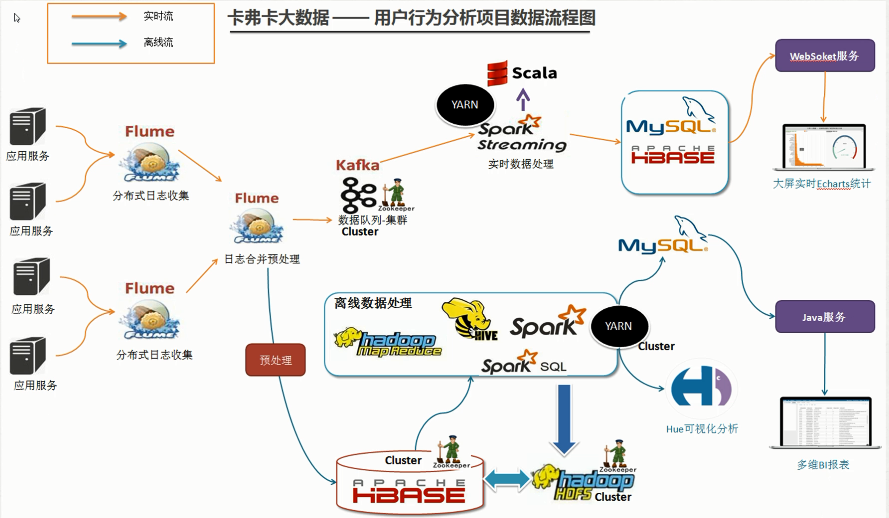
REDUCE端调优



### 考察点

* Map端和Reduce端的调优原理
* Reduce什么时候开始从Map端拉去数据

# 实战课程



## 离线数据采集

常见的Sqoo/Flume

## 实时数据采集

前端埋点