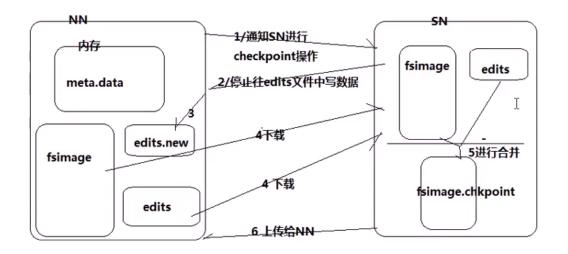
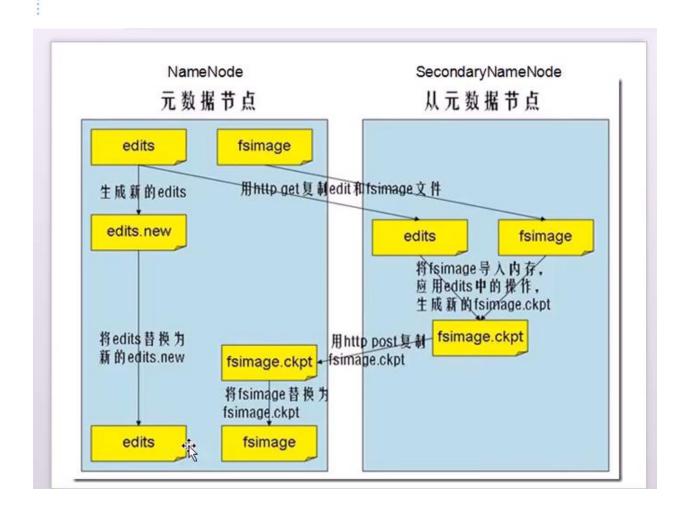


- HDFS的实现思想: 1、hdfs是通过分布式集群来存储文件,,为客户端提供了一个便捷的访问方式,就是一个虚拟的目录结构
  - 2、文件存储到hdfs集群中去的时候是被切分成block的
  - 3、文件的block存放在若干台datanode节点上
  - 4、hdfs文件系统中的文件与真实的block之间有映射关系,由namenode管理
  - 5 每一个block在集群由全左键多个刨木。打外身可以掉点数据的可靠性。还可以掉点法间的去叶鼻

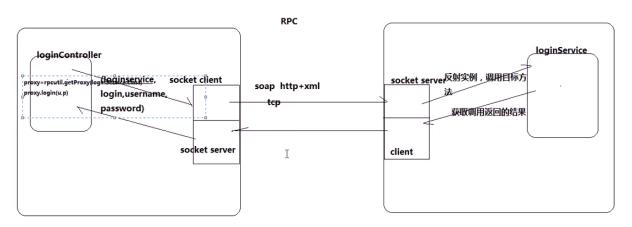


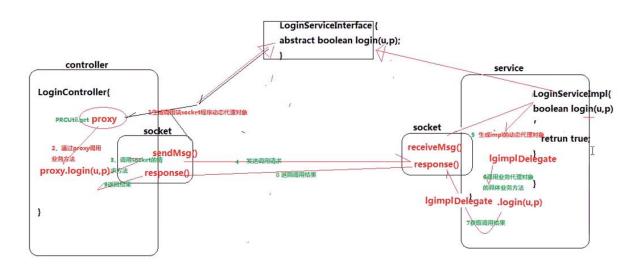
## NN的职责:

- 1、维护元数据信息
- 2、维护hdfs的目录树
- 3、响应客户端的请求



## 远程过程调用的底层实现机制





```
DFSClient.class
                                                                                 DFSInputStr...
                                                                                                               BlockReader.... 

WCMapper.java 
WCReducer.java
.1 //map 和 reduce 的数据输入输出都是以 key-value对的形式封装的
.2 //默认情况下,框架传递给我们的mapper的输入数据中,key是要处理的文本中一行的起始偏移量,这一行的内容作为value
3 public class WCMapper extends Mapper<LongWritable, Text, Text, LongWritable>{
4
.5
               //mapreduce框架每读一行数据就调用一次该方法
.60
               @Override
7
              protected void map(LongWritable key, Text value, Context context)
                                 throws IOException, InterruptedException {
.8
                        //具体业务逻辑就写在这个方法体中,而且我们业务要处理的数据已经被框架传递进来,在方法的参数中 key-value
9
10
                        //key 是这一行数据的起始偏移量 value 是这一行的文本内容
1
12
                        //将这一行的内容转换成string类型
13
                        String line = value.toString();
4
15
                        //对这一行的文本按特定分隔符切分
                        String[] words = StringUtils.split(line, " ");
6
17
                        //遍历这个单词数组输出为kv形式 k: 单词 v : 1
8
9
                        for(String word : words){
0
1
                                 context.write(new Text(word), new LongWritable(1));
12
13
                        }
4
5
FileSystem.c...
                                                                                      DFSInputStr...
                           DistributedF...
                                                        DFSClient.class
                                                                                                                    BlockReader....

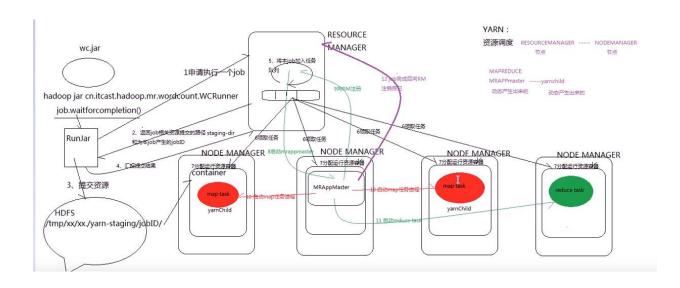
    WCReducer.java 
    ™
    3
    WCReducer.java 
    ™
    3
    WCReducer.java 
    ™
    S
    WCReducer.java 
    W
                                                                                                                                                 J WCMapper.java
    5 import org.apache.hadoop.io.LongWritable;
    6 import org.apache.hadoop.io.Text;
    7 import org.apache.hadoop.mapreduce.Reducer;
    9 public class WCReducer extends Reducer<Text, LongWritable, Text, LongWritable>{
 10
 11
 12
 13
                  //框架在map处理完成之后,将所有kv对缓存起来,进行分组,然后传递一个组<key,valus{}>,调用一次reduce方法
 14
                  //<hello,{1,1,1,1,1,1.....}>
 15⊕
                  @Override
 16
                  protected void reduce(Text key, Iterable<LongWritable> values,Context context)
 17
                                     throws IOException, InterruptedException {
 18
 19
                            long count = 0;
 20
                            for(LongWritable value:values){
  21
  22
                                     count += value.get();
 23
                            }
  24
 25
 26
 27
 28
```

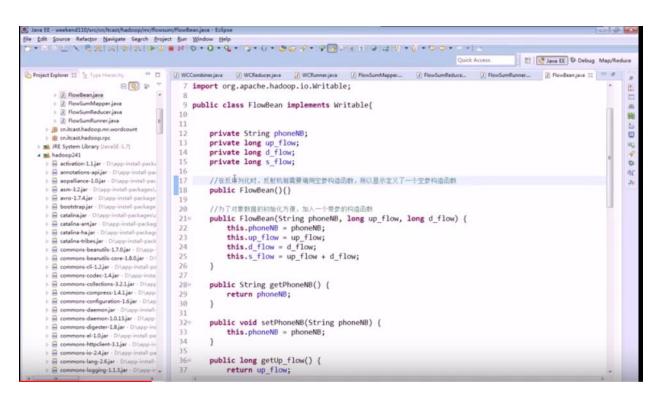
context.write(key, new LongWritable(count));

```
Quick Access
                                                                       ☐ ☐ Java EE 🦻 Debi

☑ SolrDao.java

           DFSClient.class
                       DFSInputStr...
                                   BlockReader....
                                                                         20
 219
        public static void main(String[] args) throws Exception {
 22
23
           Configuration conf = new Configuration();
24
25
           Job job = Job.getInstance(conf);
26
 27
           //本job使用的mapper和reducer的类
 28
            job.setMapperClass(WCMapper.class);
 29
            job.setReducerClass(WCReducer.class);
 30
31
32
           //指定reduce的输出数据kv类型
33
            job.setOutputKeyClass(Text.class);
 34
            job.setOutputValueClass(LongWritable.class); ]
 35
 36
37
            job.setMapOutputKeyClass(Text.class);
            job.setMapOutputValueClass(LongWritable.class);
38
40
            wcjob.setOutputKeyClass(Text.class);
41
            wcjob.setOutputValueClass(LongWritable.class);
42
43
            //指定mapper的输出数据kv类型
44
            wcjob.setMapOutputKeyClass(Text.class);
45
            wcjob.setMapOutputValueClass(LongWritable.class);
46
47
48
            //指定要处理的输入数据存放路径
49
            FileInputFormat.setInputPaths(wcjob, new Path("/wc/srcdata/"));
50
51
            //指定处理结果的输出数据存放路径
52
            FileOutputFormat.setOutputPath(wcjob, new Path("/wc/output/"));
53
            //将job提交给集群运行
54
55
            wcjob.waitForCompletion(true);
56
57
58
        }
```





```
ile Edit Source Refactor Navigate Search Project Run Window Help
               ☐ Java EE Debug Map/Reduce
                                                      C D D WCCombiner java D WCReducer java D WCRinner java D FlowSumMapper... D FlowSumReduce... D FlowSumReduce... D FlowSumReduce...
  Project Explorer $2 12 Type Hierarchy
                                         B 1 1 7
                                                                    47
                                                                                        this.s_flow = s_flow;
              FlowBeanjava
                                                          .
                                                                    48

⇒ II FlowSumMapper.java

                                                                     49
          | FlowSumReducer java | FlowSumReducer java | FlowSumReducer java | FlowSumRunner java | Flow
                                                                     50
                                                                                                                                                                                                                                                                           180
                                                                     51
                                                                     52
                                                                                                                                                                                                                                                                           0
                                                                                  //将对象数路序列化到波中
              @ cnitcast.hadoop.rpc
                                                                     530
       > mi JRE System Library (JavaSE-1.7)
                                                                                  public void write(DataOutput out) throws IOException {
                                                                   654
          > activation-1.1 jar - Dt.app-install-packu
            armotations-apijar - D\app-install-par
appaliance-1.0/ar - D\app-install-par
                                                                    56
                                                                                         out.writeUTF(phoneNB);
                                                                                         out.writeLong(up_flow);
out.writeLong(d_flow);
             asm-3.2.jar - Driapp-install-packages),
avvo-1.7.4.jar - Driapp-install-package
                                                                     58
                                                                     59
                                                                                         out.writeLong(s_flow);
             Bootstrapjar - Or\app-install-package
                                                                    60
              catalina.jar - Di\app-install-packages\r
                                                                    61
                                                                                 }
          > ☐ catalina-antjar - D\app-install-packag
> ☐ catalina-hajar - D\app-install-package
                                                                     63
           3 Gatalina-tribes jar - Disapprinstall-pack
                                                                   64
65
66**
                                                                                  //从数据读中反序列出对象的数据
          commons-bearutils-1.7.0 jar - D: App-
                                                                                   //从数据读中读出对象字段时,必须额序列化时的顺序保持一致
             commons-cli-1.2 jar - Dhapp-install-pe
Commons-codec-1.4 jar - Dhapp-install
                                                                                  @Override
                                                                                  public void readFields(DataInput in) throws IOException {
                                                                  68
69
70
           a commons-collections-3.2.1 jar - Drusps
                                                                                         phoneN8 = in.readUTF();
           ⇒ a commons-configuration-1.6 jar - Dt/ap
                                                                                         up_flow = in.readLong();
            commons-daemon.jar - D/app-install-
                                                                                         d flow = in.readLong();
                                                                                         s_flow = in.readLong();
             a commons-digester-1.8 jar - D/upp-inc
             a commons-el-1.0 jar - D\app-install-pa
           Commons-httpclient-3.1 jar - Dt.app-in-

commons-io-2.4 jar - Dt.app-install-pa

commons-lang-2.6 jar - Dt.app-install-
                                                                                  }
                                                                    76 }
          ▶ 🖨 commons-logging-1.1.3 jer - Drupp-ir 🐷
         F annotations-apiliar - Driapp-install-pas
            aopaliance-1.0jar - Dt\app-install-pac
                                                                87=
-88
89
                                                                                  @Override
            asm-3.2 jar - Ohapp-install-packagesh
                                                                                  public String toString() {
             avro-1.7.4jar - Dhapp-install-package
            Dootstrapjar - D/sop-install-package
                                                                  90
91
                                                                                  return "" + up_flow + "\t" +d_flow +[ "\t" + s_flow;
            atalinajar - Dhapp install packagestr
          F atalina-antjar - Drupp-imtall-packag
            🖨 catalina-hajar - Dr/app-install-package
                                                                   92
            atalina-tribes jar - D\app-install-pack
         → G commons-bearutils-1.7.0 jar - D(upp- 94 }

    Java EE - weekend110/src/cn/hcast/hadoop/mr/flowsum/FlowSumMapper java - Eclipse

 Rie Edit Source Refactor Navigate Search Project Run Window Help
ロマニーション N 表示 スタッス IP D M N ウェウェス・フェウェク・プログラック・プランス 日 マコグラット・ファント
                                                                                                                                                                                                                          ☐ Java EE Debug Map/Reduce
                                                                                                                                                                                            Quick Access
                                                     🗎 🗓 WCCombinerjava 🗓 WCReducerjava 🖟 WCRunnerjava 🖟 FlowSumMapper... 🗵 🖟 FlowSumReduce... 🖟 FlowSumRunner...
  Project Explorer 11 1 Type Hierarchy
                                                                                                                                                                                                                                     (J) FlowBeamjava - #
                                         B 1 7
                                                                     12 * FlowBean 是我们自定义的一种数据类型,要在hadoop的各个节点之间传输,应该遵循hadoop的序列化机制
                                                                     13 * 敦必须实现hadoop相应的序列化接口
                                                           E
     in hadoop-04
                                                                     14 * @author duanhaitao@itcast.cn
                                                                                                                                                                                                                                                                            10
     is hadoop-06-hbase
                                                                    16 */
                                                                     17 public class FlowSumMapper extends Mapper<LongWritable, Text, Text, FlowBean>{
                                                                                                                                                                                                                                                                            # hiveudf
     Mucene0830
     storm-comp
                                                                   20
21
                                                                                   //重到日志中的一行数据,切分音个字题,抽取出我们需要的字题: 手机号,上行波量, 然后封禁城kv发送出去
   # S weekend110
                                                                                   @Override
                                                                                                                                                                                                                                                                            er
                                                                                  protected void map(LongWritable key, Text value,Context context)
                                                                                                 throws IOException, InterruptedException {
          # @ cn.itcast.hadoop.mr.flowsym.
             > [] FlowBeanjava
                                                                                          //拿一行数据
             ► D FlowSumMapper.java
                                                                                          String line = value.toString();
                                                                    27
              ▶ ☐ FlowSumRunnerjava
                                                                                         String[] fields = StringUtils.split(line, "\t");
                                                                    28
          m cnitcast.hadoop.rpc
                                                                                          //查别我们盘要的定题
         mi JRE System Library (Invalid-1.7)
                                                                                          String phoneNB = fields[1];

→ mi hadoop241.

            activation-1.1 jar - Dr. app-install-packz
annotations-apijar - Dr. app-install-packz
appallance-1.0 jar - Dr. app-install-packz
                                                                                          long u_flow = Long.parseLong(fields[7]);
                                                                                          long d_flow = Long.parseLong(fields[8]);
             asm-3.2jar - D\app-install-packages\
            avro-1.7.4 jar - Dr\app-install-package
                                                                                          context.write(new Text(phoneNB), new FlowBean(phoneNB,u_flow,d_flow));
                                                                    36
             B bootstrapjar - Dr\app-install-package
                                                                    37
           > A catalinajar - Drapp-install-packagestul
          s actalina-antjar - Driapp-install-packag
             catalina-tribes.jar - D\app-install-pack
commons-bearutils-1.7.0.jar - D\app-
```

```
ile Edit Source Refactor Navigate Search Project Run Window Help
     ☐ Java EE Debug Map/Reduce
                                                                                                                     Quick Access
Project Explorer 22 12 Type Historich
                                🐃 🔯 🖟 WCCombinerjava 🖟 WCReducerjava 🖟 WCRunnerjava 🖟 FlowSumMapper... 🖟 FlowSumReduce... 🖫 🖟 FlowSumRunner... 🖟 FlowBeanjava
                        B 3 2 "
  iii demo-hadoop
                                             import org.apache.hadoop.io.Text;
  iii hadoop-04
                                          6 import org.apache.hadoop.mapreduce.Reducer;
  m hadoop-06
                                                                                                                                                                     100
⊳ 🧺 hadoop-06-hbase
                                             public class FlowSumReducer extends Reducer<Text, FlowBean, Text, FlowBean>{
                                                                                                                                                                     S
  iii hadoop-07
                                                                                                                                                                     0
> Sal hiveudf
 ≥ [a] lucene0830
                                                                                                                                                                     100
                                                  //框架每传递一组数据<1387788654,{flowbean,flowbean,flowbean,flowbean....}>调用一次我们的reduce方法
 storm-comp
                                                  //reduce中的业务逻辑数是遍历values,然后进行累加求和再输出
 weekend110
                                        139
                                                  @Override
                                                                                                                                                                     es:
                                                  protected void reduce(Text key, Iterable<FlowBean> values,Context context)

→ ⊕ cn.itcast.hadoop.hdfs

                                                           throws IOException, InterruptedException {
     ■ Enitcasthadoop.mr.flows
       ▶ ☑ FlowBeanjava
                                                      long up_flow_counter = 0;

→ FlowSumMapper.java

                                                      long d_flow_counter = 0;
       > [] FlowSumReducerjava

| [] FlowSumRunner.java
| [] cn.itcast.hadoop.mr.wordcount
|
                                                     for(FlowBean bean : values){
       @ cn.itcast.hadoop.rpc
    M JRE System Library [JavaSE-1.7]
                                                           up_flow_counter += bean.getUp_flow();
                                                           d_flow_counter += bean.getD_flow();
   # mi hadoop241
      activation-1.1 jar - D\app-install-packs
annotations-api,jar - D\app-install-pas
       aopaliance-1.0 jar - D/app-instali-pac
       asm-3.2.jar - D(app-install-packages)
       avro-1.7.4 ar - O\app-install-parkage
                                         28
                                              context.write(key, new FlowBean(key.toString(), up_flow_counter
       bootstrapjar - D/Japp-install-pockage
       atalinajar - Drupp-install-packages\u
                                        30
31
                                                 }
      atalina-hajar - Dhapp-install-package
      a catalina-tribes.jar - Dr.app-install-pack
commons-bearutils-1.7.0.jar - Dr.app-
      a commons-beanutils-core-1.8.0 jar - DA ...
```

```
Dava EE - weekend110/src/cn/rcast/hadoog/mr/flowsum/FlowSumRunner.java - Eclipse
 File Edit Source Refactor Navigate Search Project Run Window
 ☐ Java EE Debug Map/Reduce
                                                                                                                       Quick Access
                                 🐃 🗅 🕧 WCCombinerjava 🗓 WCReducerjava 🗓 WCRunnerjava 🗓 FlowSumMapper... 🗓 FlowSumMeduce... 🗓 FlowSumMunner... 🖫 🎚 FlowSumMunner... 🖫 🗓 FlowSumMunner...
 Project Explorer 11 1/2 Type Hierarchy
                          B 1 2 "
                                           13 public class FlowSumRunner extends Configured implements Tool{
                                                                                                                                                                        1
                                                                                                                                                                        iii hadoop-04
                                           15e
                                                    @Override
                                                                                                                                                                        201
   a hadoop-06
                                          -16
                                                   public int run(String[] args) throws Exception {
                                                                                                                                                                        166
  ≥ 🥁 hadoop-06-hbase
                                           17
                                                         Configuration conf = new Configuration();
                                                                                                                                                                        0
                                           18
  > Sel hiveudf
                                           19
                                                        Job job = Job.getInstance(conf);
                                                                                                                                                                        4
                                           20
  storm-comp
                                           21
                                                        job.setJarByClass(FlowSumRunner.class);
                                                                                                                                                                        ø
                                           22
  eg.
                                                         job.setMapperClass(FlowSumMapper.class);
                                                                                                                                                                        Ju

→ El cn.itcast.hadoop.hdfs

                                           24
                                                        job.setReducerClass(FlowSumReducer.class);

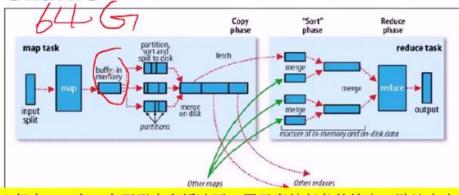
■ cnitcasthadoop.mr.flowsum

                                           25
        FlowBeanjava
                                           26
                                                         job.setMapOutputKeyClass(Text.class);
         ▶ I HowSumMapper.java
                                           27
                                                        job.setMapOutputValueClass(FlowBean.class);
         ▶ II) FlowSumReducerjava
                                           28
         ▶ ☐ FlowSumRunner.java
                                                         job.setOutputKeyClass(Text.class);
                                           29
      > 10 cn/tcasthadoop.mr.wordcount
                                                         job.setOutputValueClass(FlowBean.class);
       enitcasthadoop.rpc
      M JRE System Library (JavaSE-3.7)
                                                         FileInputFormat.setInputPaths(job, new Path(args[0]));
    # mi hadoop241
        activation-1.1 jar - Dr. app-install-packs
                                                        FileOutputFormat.setOutputPath(job, new Path(args[1]));
                                           33
       annotations-apijar - Dhapp-install-pai
                                           34
        appalliance-1.0 jar - Ditapp install-pac
                                           35
        asm-3.2.jar - Ottapp-install-packagest-
                                           36
                                                        return job.waitForCompletion(true)?0:1;
        🖨 avro-1.7.4jar - D/Japp-Install-package
                                           37
                                                   }
        bootstrapjar - D/Japp-install-package
                                           38
       acatalina.jar - DNapp-install-packages\u.
catalina-ant.jar - DNapp-install-packag
                                                    public static void main(String[] args) throws Exception {
       ☐ catalina-ha.jar - Dr.app-install-package
☐ catalina-tribes.jar - Dr.app-install-pack
                                          41
                                                         int res = ToolRunner.run(new Configuration(), new FlowSumRunner(), args);
       commons-bearutils-1.7.0 jar - D/upp-
commons-bearutils-core-18.0 jar - D/u
                                                         System.exit(res);
```

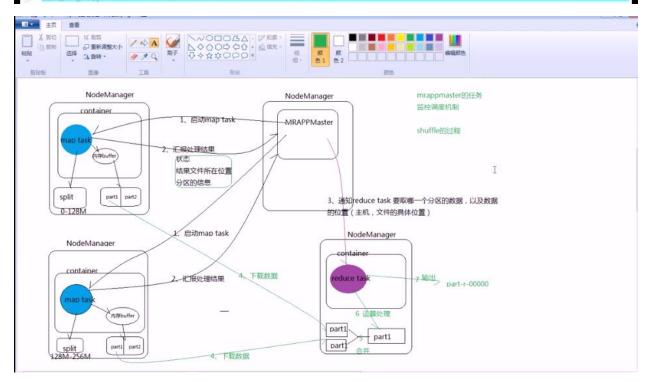
```
//设置我们商定义的分组逻辑定义
job.setPartitionerClass(AreaPartitioner.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(FlowBean.class);
//设置reduce的任务并发数,应该跟分组的数量保持一致
job.setNumReduceTasks(6);
3 import java.util.HashMap;
4
5 import org.apache.hadoop.mapreduce.Partitioner;
 7 public class AreaPartitioner<KEY, VALUE> extends Partitioner<KEY, VALUE>{
9
      private static HashMap<String,Integer> areaMap = new HashMap<>();
10
11-
      static{
          areaMap.put("135", 0);
12
          areaMap.put("136", 1);
13
14
         areaMap.put("137", 2);
          areaMap.put("138", 3);
15
16
         areaMap.put("139", 4);
17
18
19
20
21
22
23⊕
      @Override
      public int getPartition(KEY key, VALUE value, int numPartitions) {
24
25
          //从key中重到手机号,查询手机归属地字典,不同的省份近回不同的组号
26
27
          int areaCoder = areaMap.get(key.toString().substring(0, 3))==null?5:areaMap.get(key.toSt
28
29
          return areaCoder;
30
31
32 }
```

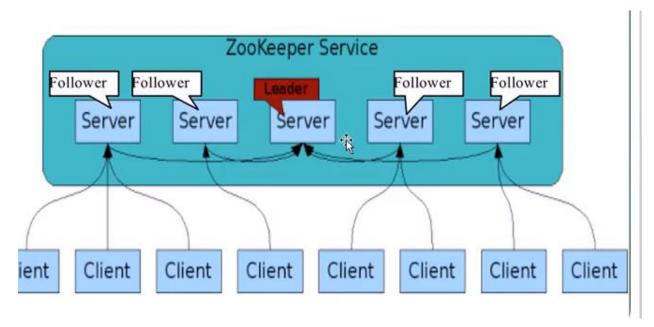


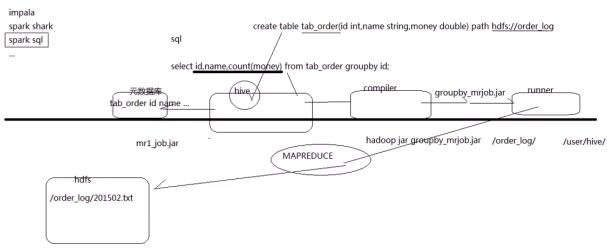
www.itcast.cn

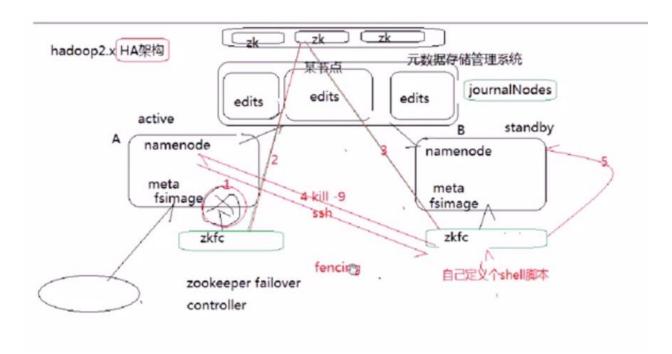


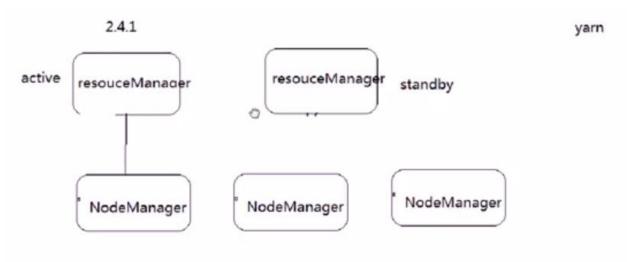
- 1.每个map有一个环形内存缓冲区,用于存储任务的输出。默认大小100MB(io.sort.mb属性),一旦达到阀值0.8(io.sort.spill.percent),一个后台线程把内容写到(spill)磁盘的指定目录(mapred.local.dir)下的新建的一个溢出写文件。
- 2.写磁盘前,要partition,sort。如果有combiner,combine排序后数据。
- 3. 等最后记录写完,合并全部溢出写文件为一个分区且排序的文件。
- 1.Reducer通过Http方式得到输出文件的分区。
- 2.TaskTracker为分区文件运行Reduce任务。复制阶段把Map输出复制到Reducer的内存或磁盘。一个Map任务完成,Reduce就开始复制输出。
- 3.排序阶段合并map输出。然后走Reduce阶段。











npase		列族 baseinfo	列族 extrainfo
	000001	name:fengjie age:18 career:huyou	hobbies:chuiniu married:true
	000002	name:furong age:28 xueli:benke jiguan:	hubei