

MapReduce 手记 ①

by 水星へノ訪客

MapReduce Errors

(2011/7/7)You cannot use abstract class to configure jobs

Here I met a problem as the following:

Exception in thread "main" java.lang.RuntimeException: java.lang.InstantiationException

it is the InstantiationException of java, because i add the code :

job.setInputFormatClass(FileInputFormat.class);

job.setOutputFormatClass(FileOutputFormat.class);

here, FileInputFormat and FileOutputFormat are abstract class which cannot be instantiated

from web

He doesn't even need a MultipleFileInputFormat. He just have to provide the type of the files, you have to decide whether these are sequencefiles or textfiles. As you should see for yourself- fileinputformat is a abstract class. So it can't be instantiated either. AND you don't need to put the previous job outputs into the same directory. Just call another addInputPath for the other directory. – Thomas Jungblut Apr 14 at 20:23

(2011/7/11)dead of hadoop process

when I just run hadoop job, it crashes---because not all nodes are normal, I do tests in one machine, so every kind of node is in this one machine, when some one is not normal, the job will not be done. Just stop all nodes and restart them until everything turn to be okay.

(2011/7/11)wrong key

[root@localhost hadoop-0.20.2]# bin/hadoop jar testelp-2.jar jobs.ConvertFormat input output
11/07/07 15:53:27 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.

11/07/07 15:53:28 INFO mapred.JobClient: Running job: job_201107071551_0001

11/07/07 15:53:29 INFO mapred.JobClient: map 0% reduce 0%

11/07/07 15:53:45 INFO mapred.JobClient: map 100% reduce 0%

11/07/07 15:53:59 INFO mapred.JobClient: Task Id : attempt_201107071551_0001_r_000000_0,
Status : FAILED

java.io.IOException: wrong key class: org.apache.hadoop.io.Text is not class util.AxiomKey

problems here

if you want to write a class, you should extends "SequenceFileOutputFormat<TripleSource, Triple>"like this:

just to rewrite RecordWrite
the above is not required!!

i made a PROBLEM here, the input <key,value>pair of reducer should be reduce(T key, Iterable<K> values, Context ctxt),
but not reduce(T key, K values, Context ctxt)
such small problem

(2011-7-19)output directory already exists

Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory input already exists
at
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:123)

you should make sure that the output directory the running hadoop process will use does not exist, that's it. So if you want to use the same name dir, you should delete pre-dir firstly. As follows:

(2011-7-23) comparator 1

something wrong with WritableComparable's compareTo if you do not compare all the properties

of such classes

which implements WritableComparable:

comapreing all the properties means that you should make sure EVERY paroperty's bigness,

not put all the other cases

just in one case. you should compare from first to last.

【2011-7-29】*** Do not try to save iterators of MR

I found a wired problem, in reduce work, if you iterate the iterable<T> values, the iterator or reference will indicate the first one element or last one , but works well in the middle.

DO NOT TRY TO SAVE ITERABLE<T> REFENCE!!! use the most primitive properties of Variables.

(2011-8-3)

When we implement this interface WritableComparable<T>, we should have T(), this constructor in this class, whatever.

【2011-8-6】***go back to the problem of 2011-7-29 【tri-star】

i asked the question on hadoop forums, and got the answer as follow:

我刚开始用 MR 编程时也犯过这种错误。Hadoop 权威指南已经说过，不能直接存储 map 或 reduce 函数的参数，因为这些引用会被重复利用，每一次调用这两个函数，参数所指向的值已经改变了。如果要保存一个值，比如 Text value, 就必须用拷贝构造函数来新建立一个保存。Text store = new Text(value)。

你如果直接保存它们，那最后它们都指向最后一个输入记录。

(2011-8-8)

hadoop 在 map 以后是先将 reduce 记录分到不同的 reduce task (一个 tasktracker 可能能执行有多个或一个 reduce task)，再进行规约，所以一定要先进行分割(partitioner)，让相同 key 值的记录到一个 reduce task 里去。

(2011-8-26)

Today i put the programe test in real enviroment.

(2011-8-28)

I have met a problem that i found there are some problems in the hadoop net.

- 1: it seems there are only two nodes
- 2: i cannot access the dfs.

(2011-8-28) bin/hadoop namenode -format

I encountered the following problems:

Bad connection to FS. command aborted.

maybe the namenode is not started, you can use following command:

bin/hadoop namenode -format

(2011-8-31)

When i runed a job, program did not print any log information.

****this is log4j configuration problem.**

you can modify the log4j.properties file.

【2011-9-3】 ****same keys to different node-must setPartitionerClass ?**

We go back to log-2011-8-8

We go back to log-2011-8-8

hadoop 在 map 以后是先将 reduce 记录分到不同的 reduce task (一个 tasktracker 可能能执行有多个或一个 reduce task), 再进行规约, 所以一定要先进行分割(partitioner), 让相同 key 值的记录到一个 reduce task 里去。

I fond a dicription from web:

setPartitionerClass 和 setNumReduceTasks: 设置 Partitioner, 默认为 HashPartitioner, 其根据 key 的 hash 值来决定进入哪个 partition, 每个 partition 被一个 reduce task 处理, 所以 partition 的个数等于 reduce task 的个数

this means that if the seted number of reduce tasks is more than correct case, hadoop will partitionate same keys to different reduce-tasks, but if you set setPartitionerClass, the seted number of reduce tasks will be trival.

1- if you set mapNumberTasks 0, no one record will be fed in.

2- if number of reductasks is bigger than number of different keys, same keys will be translate to different reduce tasks. For some mr tasks, it is ok, but for tasks like inferencing on Ontology, same keys should be merged in one reduce task which may cause unballence of system.

3- but i do not know what will happen if the number of reductasks is lower than the number of different keys.

(2011-9-4) wait for a while to setup HDFS!!!

1- Today, i am writting a data generator.

2- i met a problem which showed information as follow:

(hadoop)could only be replicated to 0 nodes, instead of 1

i found answers from web and there are many causes for that, two of them happened in my case are:

1) hdfs is still in safe mode;

2) some datanodes have not been in connection with namenode, so you may wait for some seconds.

(2011-9-5) scalability

I met the problem always waiting for me, scalability.

i should have a plan to sovlе it before October.

1- ensure the format of axioms.

2*- catch the method to process compressed data (can be later)

3- compress input data through creating dictionary of uris.

4- redesign the algorithm of reasoning.

today i runed an 1 million scalble data, but had a problem which showed that hadoop gives up procesing when the time of waiting for the reponse of a task is beyound the time limit.

the reason for this is one node running a long time.

Tonight will be a battle!

a- find the problems, between map and reduce process;

b- make clear of partitioner

c- redisgn the algorithm and data structures.

d- how to work on compressed data

>>>>>>>>about b:

1- combiner and partitioner are both for the decreament of amount of data to translate between nodes.

2- For partitioner, i found

the use of it is showed in map-end,

i suppose that:

1- if a job is not set one partitioner, mapper will still create as many sections as the number of ReduceNum, so without partitioner, records having the same key maybe written into different sections;

2- at least one partitioner is needed, although there would be many Reduce Tasks which make mapper create the same amount of sections as the number of reduce tasks, those records sharing the same key would be written in same section.

3- one section for one reduce task, on reduce step, one reduce task will read data from all sections mapped to it.

4- the default number of reduce tasks is 1!!!

(2011-9-6)

mapred.tasktracker.reduce.tasks.maximum is in mapred-site.xml

确实原来只有一个 reducer 在跑，现在有多个，但是多个任务同样造成失败。所以仍旧是 reducer 算法的问题。不是 reducer 太少的问题，而是一个 reducer 确实在十分钟内处理不完。。

我打算强制离开远程界面，看看再次登录是什么效果。。

=====

It showed that:

1- [about coding and testing] it is not a good way to strightly finish all coding and then do testing on each module.

2- [about rewriting] rewriting does not really means the rewriting of every line of code, you can use copy of moudles to rewrite.

【2011-9-16】还是有一个问题你没有解决：

就是为什么当 reduce 的个数超过键值个数时会产生同键记录分离的现象？

(2011-9-16)

首先测试压缩的问题。

I will learn how to use compressor?

*if i compress multi-files?

*if i compress the directory?

****you can use *tar* command to pack files, while use *gzip* to compress files

gzip can only compress single file instead of multi-files, so you'd tar them first.

*****FileInputFormat can only read first level directory, but if it is tar file, you can put files in directories.

*****if a file is a tar file, logs will contain binary info

(2011-9-25/26/27/28)极端封装

在编写 0.2.50 的时候，我将各个规则的执行封装到一个统一的过程中去了，这样极端的做法使得后来对于单个过程的特殊处理变得极为麻烦。应该在事先发现共同行为与特殊行为的情况，封装不一定带来灵活。

(2011-11-3/4)

if your key or value has parameters like "long...", RecordReader will throw NoSuchMethod exception. you should add one more construction.

(2011-11-3/4) Configuration order

```
/*-----*/
Configuration conf = new Configuration();
conf.setInt("maptasks", numMapper);
conf.set("output.fileName", "step-0-initialize");
Job job = new Job(conf);
/*-----*/
job got all above configurations.
while,
/*-----*/
Configuration conf = new Configuration();
conf.setInt("maptasks", numMapper);
Job job = new Job(conf);
conf.set("output.fileName", "step-0-initialize");
```

```
/*****  
job didn't get the "output.fileName" configuration.
```

【2011-11-11/12/18/21】**** No Write Out!!! 1

这两天遇到一个巨大的问题;

reducer 明明输出了数据, 但是其实并没有写出。

当我将 nummapper & numreducer 的数量改过之后, 又正确了。究竟是怎么回事?
输出时一定要检查输出文件。

【2011-11-11/12/18/21】*** iterator

不要直接用 mapred 的引用

包括类的内部传出的引用。

即使这样写也是错误的: BytesWriteble b = new BytesWriteble(itr.next().getBytes());

(2011-10-11/12) BAD NETWORK!!

--if there is no such dir, the pscp will create file named as this non-exist dir.

--i met a problem:

hdfs.DFSCClient: Exception in createBlockOutputStream java.io.IOException:Bad connect ack
with firstBadLink 10.3.8.243:50010

**you just need to close the firewall on node with ip 10.3.8.248.

--It showed that node3 always broke down while runing a job, i thought it was the interface of
switcher that led to thus problem, but i was wrong.

the net situation of node3 performed quite bad, so i restarted it, it works better, i tried to
update the configuration of it to see if it works.

不是因为 datanode 到 namenode 没有设置无密码的原因。也不是因为防火墙的原因。

*****确实是有些节点重新启动了一下就变好了!

//问题解决。

【2011-11-19/24/26/28】**** No Write Out!!! 2 ✓

自己写了一个 outputformat

reducer 的数量设定成一个时，所有输出正确写出了；

但是当 reducer 的数量设定成超过一个时，不是所有输出记录都被写出，

每一个 reducer 都会写到不同文件中，如果往同一个文件中写，会覆盖。
reducer 的输出不能合并。

一个做法是修改函数，使得对于某个 reducer,新建一个文件。

这些不同的文件最后都在输出文件下

【2011-12-10】**** add many bits 1

I have met a problem now.

hadoop writer , reader seems to add many bits.

--是不是一开始输入进来的时候？不是

--是 setCapacity?是，不是 setSize 的原因

【2011-12-13】**** add many bits 2

I have met a problem now.

hadoop writer , reader seems to add many bits.

--是不是一开始输入进来的时候？不是

--是 setCapacity?是，不是 setSize 的原因

form web:

你自己想想，如果只写那 17 个 byte，那如何做反序列化？很显然要将这 17 这个长度值也序列化才行，所

以那 8 个字节代表一个 long 变量，值就是这个 byte 数组的大小。

但不是这个原因。

【2011-12-28】**** add many bits 3

关于写出去的字节数组变长的事情，检查一下是否是 writer 的原因。即使是原带的也存在此问题。

【2011-10-3】BAD NETWORK!!

I have met a problem after configuring of a ubuntu processor : when i try to put files from local file system to hdfs, hadoop throws a Exception as following:

```
*****
```

```
org.apache.hadoop.ipc.RemoteException: java.io.IOException: File /user/lighttpd/f.jar could only be replicated to 0 nodes, instead of 1
```

```
*****
```

this exception happened when i access the host in remote form, but in local host is ok.

there are some reasons and solutions on web:

>>>>

1- maybe the fire wall is on

so i try to shut down the fire wall using command:

```
***sudo ufw disable
```

which is not working.

maybe i should open fire wall to perticular port like 22 for ssh protecol:

so i change the rules of fw using command:

```
***sudo ufw allow ssh %or sudo ufw allow 22
```

which does not work either.

>>>>

2- if you are in localhost mode, maybe its about /etc/hosts

modify it adding localhost

>>>>

3- i delete all files in /tmp/

and reformat namenode as

```
***bin/hadoop namenode -format
```

which does not work.

>>>>

4- to start namenode singlely

```
**bin/hadoop-daemon.sh start namenode
```

which dosn't work.

>>>>

MapReduce Manual

(2011-7-19) self4j & log4j

java.lang.NoClassDefFoundError: org/apache/log4j/LogManager

because hadoop framework has log4j itself, so you just use self4j which is just an api, but if you want to test your program with JUnit and MRUnit, you should import log4j packet or jar file.

(2011-7-20) MRUnit & JUnit

I have met many problems when I use JUnit and MRUnit to do test.

a> version, you should get 0.20.2 or higher version, instead of that, program will throw NoSuchMethod Exception of MockContext, higher versions take it down.

b> In map or reduce task, you should not read job's configuration, otherwise, it will throw NoSuchPointer Exception. In this case, you copy ordinary map or reduce code and rewrite it using fake configuration. That'll be ok.

Actually, if you do not want to rewrite previous mapreduce tasks, you can use `driver.withConfiguration` & `driver.withCounter` to achieve it.

(2011-7-22) MRUnit & JUnit 1

Every time, you use MRUnit `mapDriver`, it will achieve just one pair,

If you write

```
for(Pair<Key, Value> pair : input) mapDriver.withInput(pair);
```

it will just achieve the last one element in input list.

but for `mapreduceDriver`, all be ok,

```
for(Pair<Key, Value> pair : input) mapDriver.withInput(pair);
```

it will achieve all elements as input.

(2011-7-22)Comparator

if you use your own defined class as key or value(especially key), you should write method `"compareTo"`, otherwise, mr process will not group all the keys those are same as you think.

(2011/7/7) about dir

if you write /input, which means absolute directory, while input means related directory

/input = hdfs://localhost:9000/input

input = /user/root/input

(2011/7/7) ssh & dir & jar & log & safemode

必须在运行程序前先启动 sshd 服务以及 hadoop 进程, 这样才能进一步对分布式文件系统进行访问, 以及运行 hadoop 任务。如果没有进行配置, 可以访问 dfs 文件系统。

1, 启动 sshd 服务, service sshd start (service sshd stop)

2, 启动 hadoop 进程, bin/start-all.sh (bin/stop-all.sh)

访问文件系统的时候最好加上前缀 hdfs://localhost:9000/tmp, 如果不加直接是/tmp, 那么文件系统实际访问的是用户的根目录, 即/user/root/tmp

假如不启动 ssh 服务, 就直接启动 hadoop 进程, 那么连接会被拒绝, 如下示:

```
Starting namenode, logging to
/root/mr/hadoop-0.20.2/bin/../logs/hadoop-root-namenode-localhost.localdomain.out
localhost: ssh: connect to host localhost port 22: Connection refused
localhost: ssh: connect to host localhost port 22: Connection refused
starting jobtracker, logging to
/root/mr/hadoop-0.20.2/bin/../logs/hadoop-root-jobtracker-localhost.localdomain.out
localhost: ssh: connect to host localhost port 22: Connection refused
```

namenode 进入安全模式 bin/hadoop dfsadmin -safemode

namenode 退出安全模式 bin/hadoop dfsadmin -safemode leave

=====

hadoop 运行不需要将它的核心包打进去, 只需要将外部包打进去

=====

hadoop 有自己的日志系统, 所以只要用 slf4j 来写日志就行, 至于日志的输出 hadoop 默认输出到输出文件夹中。

hadoop 控制台只会给出主类日志信息

(2011.7.6) logs 1

where is the log files? how to look at log files?
in the hadoop/logs/userlogs

(2011.7.17) counter

you cannot use configuration to count something, because all mappers and reducers work at the same time,

hadoop provides Counter as one property of a job,

you can firstly create a counter of a job in map or reduce task then counts,

after that, out of job side, you get all counters of mapper tasks and reducer tasks, then count them all.

---in map or reduce task, you can create a counter as follow

```
context.getCounter("RDFS derived triples", "subclass of resource").increment(1);
```

---out of job side, you can get counter as follow

```
job.getCounters().findCounter("RDFS derived triples", "subclass of resource").getValue();
```

**type of counter numbers is long

(2011.7.17) configuration

---job sets one property

```
job.getConfiguration().setInt("reasoner.step", ++step);
```

---in map or reduce task, you can get property as follows

```
context.getConfiguration().getInt("reasoner.step", 0)
```

it is not helpful to change this, for every node has a copy

(2011.7.17) logs 2

-if you want to search map or reduce task's logs, got to logs/userlogs

-you should use `log.info(****)`, so that you can watch your information,

`log.debug` make use only when program turns wrong.

(2011-7-19) directory

if you just define a file path like :

```
new Path("dir");
```

the dfs will treat it as the path mapping to the local file system,

if your local ****current**** path is /root/user/

so in dfs it will be /user/root/user/dir

follow has some operators of dfs:

```
**delete**/
```

```
FileSystem fs = FileSystem.get(this.getConf());
```

```
fs.delete(new Path(args[1]), true);
```

(2011-7-19) mapred & mapreduce

import org.apache.hadoop.mapreduce is new version;

import org.apache.hadoop.mapred is lower

if you write with new methods, use first ones.

(2011-7-19) how to look over faults of hadoop programmes?

use mrunit & junit

1- if you use >= 0.2.0 hadoop jar, then

```
import org.apache.hadoop.mrunit.mapreduce.MapDriver;
```

```
not
```

```
import org.apache.hadoop.mrunit.MapDriver;
```

2-put all useful jar packets in and then meunit, junit

3-unit test is just replacement of hadoop input/output system

(2011-7-11) Run a hadoop job

- 1-put all jar files in the /root/Lucia
and the input files and results
- 2-copy the target jar file in the /root/mr/hadoop
and input files to hadoop/files
- 3-create a directory file which contains all the input and output files
- 4-start sshd service
- 5-start hadoop routine- bin/start-all.sh
input passwords 123456
- 6-now you can examine the dfs
all user files are in the /user
bin/hadoop dfs -ls /user
- 7-put input files from local fs to dfs
bin/hadoop dfs -put /files/input input
- 8-run the job
bin/hadoop jar sample.jar MainClass /input /output
- 9-get files from dfs
bin/hadoop dfs -get /output files/output
- 10-delete the files in /user on dfs
- 11-stop hadoop process
bin/stop-all.sh
- 12-stop sshd service
service sshd stop
- 13-copy files/output to Lucia and delete it

you'd better delete the logs in logs/userlogs if they arent useful.

(2011-7-15)

if you implement key and value using self-definding classes, you can treat key as one property of value.

(2011-8-28) about drwxr-xr-x

第一位表示文件类型。d 是目录文件，l 是链接文件，-是普通文件，p 是管道

第 2-4 位表示这个文件的属主拥有的权限，r 是读，w 是写，x 是执行。

第 5-7 位表示和这个文件属主所在同一个组的用户所具有的权限。

第 8-10 位表示其他用户所具有的权限。

如：

```
drwxr-xr-x  user1  group1      filename
```

表示 filename 是个目录，user1 拥有读写执行的权限，和 user1 所在同一个 group1 组里的用户拥有只读和执行权限，剩下其他用户拥有只读和执行权限。

(2011-8-28)how to make directory:

```
bin/hadoop  dfs  -mkdir  /user/hadoop
```

(2011-8-29)

if you input 'cd home' , it will not work.

(2011-8-29) ssh 的级联

1-log in master-iwshadoop0, then use following command to log in other slaves:

```
****ssh [hostname]: ssh iwshadoop1
```

use 'exit' to log out

```
**** vi /etc/hostname to see the local host name
```

```
**** vi /etc/hosts to see the hosts that the localhost can connect with
```

The operations as showed above indicate that every node can connect to the others.

```
**** Esc + : + q! close vim
```

results given as follows:

iwshadoop1: ok
iwshadoop2: error
iwshadoop3: ok
iwshadoop4: ok
iwshadoop5: ok
iwshadoop6: error
iwshadoop7: ok
iwshadoop8: error
ssh 的级联，相当于登入一台主机后再登另一台主机

2-then, we check the conf files
core-site.xml shows which one is the namenode;
hdfs-site.xml ???
mapred-site.xml: which one is the jobtracker
hadoop-env.sh : about command of hadoop

3-finally, check the master-slave relationships of all nodes

(2011-8-30)

Today, i will do three things,

1,first, learn how to use vi editor;
**mkdir [dir name] used to make directory,
**touch [file name] used to make empty file that could be editated
**you should first press [esc] to enter command mode then ..
**1dd to delete lines
x to delete content

2,secondly, make sure if files can be translated by SSH.

** use scp
scp /home/iwshadoop/ttt iwshadoop@10.3.6.51:/home/iwshadoop/ttt

if there has always existed the same name file, the process will simply cover the old one.

3,modify the conf files to check if three nodes of all can work well.

everything is ok!

(2011-9-1)

****use cp -i source .. dir to copy files to one place in Linux, where -i is used to ask users if to overwrite existed files.**

(2011-9-3)

**** rm [file]**

****rm -r [dir]**

**** cp -r [dir] [dir] copy directory**

**** bin/hadoop dfs -put [file/dir] [dir] which can relay files or directory form local fs to hdfs.**

here if you input "bin/hadoop dfs -put files/input /user/input", hdfs will contain all files under files/input in /user/input, instead make a new /user/input named directory to contain files/input directory.

**** bin/hadoop dfs -rmr /user/* used for all files in directory /user deletaration.**

****you can use "pscp -r" to translate directories.**

(2011-9-7) the procedure to write mapreduce program:

- 1- write main mappers and reducers;
- 2- use mrunit to test them right;
- 3- write input and output parts;
- 4- use local system to test them right;
- 5- use distributed system to test them right.

(2011-9-7)

The works you are doing now are not like small homeworks, you should test if one choiced strategy has a high efficiency before officially starting.

you should have a prototype framework first and then test it.

(2011-9-7) How to write and read multi-files?

How to write and read multi-files?

SequenceFileInputformat cannot read files recursively, but it can read multiple files under one directory.

if you want, you can add many directory paths.

(2011-9-9/11)

reducer 开始读入数据时就已经开始了进度了。

****rm -r * 删除目录，这里是删除所有目录

(2011-9-12)

It seems that running a mapreduce program will cost a lot of space for logs.

**look over the directory's space -- du -h [dir]

all slaves' logs will be grouped in the master.

(2011-9-14)

A reduce task contains records having one more kind of keys, in reduce end point, program first sort these records according to these keys, then calls reduce function for each key.

So there may exist a reduce task containing large scale of intimate records which would make this reduce task run slowly. This will cause hadoop create an attempt task.

At last experiment:

We found jobtracker's logs have many records of attempt tasks, we check if so do tasktracker.

one task will be at least attempted on time, so its ID would be
atempt_201109141022_0002_m_0001_0

>>>>>The most important is how to decrease the immediate records?

(2011-8-27)

=====

HOW TO LOGIN?

10.3.6.50

login as: iwshadoop

password: 123456

cd home/iwshadoop/bin/hadoop-0.20.2

if you make mistake of spelling, for example "iswhadoop", the access will be denied.

=====

HOW TO DELETE FILES?

rm -r

rm

to delete files or directories.

=====

HOW TO USE PUTTY TRANSLATE FILES?

you cannot translate any files to /home, but to sub-dirs is ok, i donnot know why.

here is the commands:

pscp D:\test\t.txt iwshadoop@10.3.6.50:/home/iwshadoop

pscp iwshadoop@10.3.6.50:/home/iwshadoop/t.txt D:\

if you want to copy directories, add -r as option.

=====

Vi 编辑器的基本使用方法

\$ vi myfile 先按一下「ESC」键转到「命令行模式（command mode）」按一下「:」冒号键进入「Last line mode」，例如： :w filename （输入「w filename」将文章以指定的文件名 filename 保存） :wq (输入「wq」，存盘并退出 vi) : q! (输入 q!， 不存盘强制退出 vi)

按「i」切换进入插入模式「insert mode」，按“i”进入插入模式后是从光标当前位置开始输入文件；

(2011-9-18) InputFormat works on InputFormat, InputSplit, RecordReader

First, InputFormat split files to as many slaves as the number of setted mappers, then RecordReader processes each slave and makes them into many key-value pairs. One split to one RecordReader.

(2011-9-18) OutputFormat works on OutputFormat, RecordWriter, OutputCommitter

First, OutputFormat checks the correctness of output paths, then create writers for all records, commit if necessary.

(2011-9-18)

a job can run singlely or be called by other job.

(2011-9-20/21/22/23)

1- check if hdfs can read files with * pattern.

NO!

2- if you did not increase one counter, it will return 0 in jobClient.

3- if you sate an object having a name same as Class name with first letter as lowercase will change selvely with the changing of Class name, but on the other case, this will not happen.

4- how to combine two arries?

```
int[] a3=new int[a1.length+a2.length];
System.arraycopy(a1,0,a3,0,a1.length);
System.arraycopy(a2,0,a3,a1.length,a2.length);
*****
```

5- how to read and write Objects and read all files under one directory.

6- java.io.NotSerializableException

org.apache.hadoop.mapreduce.lib.output.TextOutputFormat can not be Serializable.

全部 implements java.io.Serializable 试试。

7- 在编译包的时候一定不要带错编译，否则应用该报也会有错，因为打包时配置信息会带错。!

已试验成功!

(2011-9-25/26/27/28) How does MRUnit work?

when it calls a reducer, it will just create one Reducer object used to process all parts of results of mappers, so The gloable variables you state in this redducer will be shared with all data patch.

(2011-10-1/6) about hadoop file system

if you write a class extending Configured and implementing Tool while using it as a common class namely you did not call

"run" function, when you get FileSystem through "FileSystem fs = FileSystem.get(this.getConf());", it will throw

"NullPointerException", because you just treat it as a common class.

2- set outputpath of SequenceFileOutputFormat is the directory of results of jobs.

(2011-11-2)

1- **Hash 函数

2- **中间过程压缩

3- **怎么只做 mapper?

了解 mr，就要鼓起勇气去了解它的配置和 shulffe

这几天主要进行编码和学习，进行本机测验就行。
后面再来考虑配置的事情，并且文件共享等等。

The performance of mapreduce is based on twofolds: program itself ; configuration of running jobs.

key and value can not use interface or abstract class.

【2011-11-2】**override the checkOutputSpecs

I just override the checkOutputSpecs of OutputFormat as

```
/******  
@Override  
    public void checkOutputSpecs(JobContext context)  
        throws IOException {  
  
    }  
/******  
then if the putput dir already exists, it will not throw exception.
```

(2011-11-2)input & output

input & output

两部分都包含两部分：Format 和 RecordReader/Writer

Format: 检查路径是否正确，可用或已存在，以及文件的格式，序列或纯文本

RecordReader/Writer: 不关注文件格式，只关注记录本身。

对于 input 来说，不显示确定文件格式，如果是纯文本，那记录键值一定是 Text 格式，否则是具体的类；

对于 output 来说，必须显示确定格式，是文本的，还是序列的。

(2011-11-5)Stream of data

对于一个 mapper 或者 reducer 来说，它只是跑一个进程，而数据像流一样流过这两个函数。

(2011-11-5)reuse

为了消除短周期现象，尽量减少在 map 函数和 reduce 函数中申请新对象。

对象可以重用，甚至是变量，因为 context.write()后，变量或者对象就可以被回收，信息已经写出去了。

(2011-11-8/9)some tips

- 1- 网速确实和网线与更新有关。
- 2- 副本数量设小一点是不是好一点？
在 hdfs-site.xml

3- 当启动集群的时候，为什么 `datanode` 无法活动？

1) 有新的节点加入，或者 `namenode` 重装了，

这时候需要重新格式化，`bin/hadoop namenode -format`

在格式化之前，删除所有节点上的 `/tmp` 文件夹，

它是用来存放老的文件系统目录的

2) 刚启动时，往往不行，还没连上 `datanode`，所以要等一会儿

或一会儿就连上了。可以用 `bin/hadoop dfsadmin -report` 查看

以上是大多数情况的原因。

4- 在我的实现中，`mapper` 的读入路径不存在会抛异常，但是程序还是照常进行。先对输入路径做一个检查。

(2011-11-11/12/18/21)

`mapper & reducer` 的配置不会传给 `jobClient`，只能通过 `counter` 来

(2011-11-11/12/18/21)

如果使用全局变量减少短周期的话

`map` 和 `reduce` 函数

要注意对成员变量清空。

(2011-10-8)

if you want to move a files into one folder, you should put:

```
**pscp /sa/fdfw/kk /home/node/tool/
```

bute not

```
**pscp /sa/fdfw/kk /home/node/tool/
```

(2011-10-13/16)

****猜想：**上传数据的时候，namenode 先将一部分数据放在本机上，再联系其他 datanode，联系上了，根据散列值将一部分数据放到其他节点上。

****确实是网线的原因造成网络总是中断。**

---How to resolve the heap problem in hadoop?

****in conf dir: hadoop-env.sh add "export HADOOP_HEAPSIZE=2000"**

---I want to add task Failed Time

in mapred-site.xml

The easiest way will be to set this configuration parameter:

```
<property>
  <name>mapred.task.timeout</name>
  <value>1800000</value> <!-- 30 minutes -->
</property>
in mapred-site.xml
```

---GC outofhead

这类错误是由于 hadoop-default.xml (mapred-site.xml)中的内存设置太小所致，即如下这个配置：

```
<property>
<name>mapred.child.java.opts</name>
<value>-Xmx200m</value>
<description>Java opts for the task tracker child processes.
The following symbol, if present, will be interpolated: @taskid@ is replaced
by current TaskID. Any other occurrences of '@' will go unchanged.
For example, to enable verbose gc logging to a file named for the taskid in
/tmp and to set the heap maximum to be a gigabyte, pass a 'value' of:
-Xmx1024m -verbose:gc -Xloggc:/tmp/@taskid@.gc
```

The configuration variable mapred.child.ulimit can be used to control the maximum virtual memory of the child processes.

```
</description>
</property>
```

默认只有 200M 的内存，可以在启动 hadoop 任务的时候，增大配置以覆盖掉这个选项。即用如下选项：

mapred.child.java.opts=-Xmx2048m

--if you input "bin/hadoop namenode -format" will cuase the unformat of datanode only if you delete the dirs under tmp or input "bin/hadoop datanode -format"

(2011-12-14) 如何将 hadoop 控制台的 log 记录输出到文件中。

在 conf/log4j.properties 里
更改

log4j.appender.console=org.apache.log4j.ConsoleAppender

log4j.appender.console.target=System.err

为

log4j.appender.console=org.apache.log4j.FileAppender

log4j.appender.console.file={%filepath%}

log 的配置信息是在作业启动时读的，所以不需要重启 hadoop 进程。

(2011-12-14)

hadoop 只能对.gz 的压缩文件进行处理，不能对.tar.gz 的文件进行处理，或者说会存在错误。

(2011-12-14)Compress

可以通过修改 hadoop-site.xml 的配置选项，可以在程序中用 JobConf 的类接口设置，或者在提交作业时

用-D 选项来设置该选项）中配置一个选项：

```
<property>
  <name>mapred.compress.map.output</name>
  <value>true</value>
</property>
```

这样，作业就会将产生的中间结果写入 slave local 的时候，对结果进行压缩，reduce 读取时也能够根

据 gz 的后缀知道该中间结果是压缩文件，于是采用相应的读取格式来读取。

(2011-10-2)安装 hadoop-0.20.2

1- 进入 bios

del

2- 安装 Ubuntu

先分区 swap, /

2*- 使用权限操作

在所有命令前加上 sudo

获取管理员权限:

sudo passwd root

如果忘记密码, 按照提示设置密码

su 进入管理员权限, 当然要输入密码。

exit 退出管理员

2** 关闭防火墙

sudo ufw disable

3- 下载安装并开启 ssh 服务

\$ sudo apt-get install openssh-server

然后确认 sshserver 是否启动了:

\$ ps -e |grep ssh

如果只有 ssh-agent 那 ssh-server 还没有启动, 需要执行命令\$ /etc/init.d/ssh start, 如果看到 sshd 那说明 ssh-server 已经启动了。

这一步完成之后就可以利用 ssh 远程登录了

4- 集群之间设置无密码登陆 ssh(本机示范)

主机名是 node0

在 node0 下, 即 cd /home/node0

生成密钥对:

ssh-keygen -d

>>>>此时会生成.ssh 目录, 这个目录不可见, 在目录里面有一对密钥对

将 id_dsa.pub 考到 authorized_keys 中

cat .ssh/id_dsa.pub >> .ssh/authorized_keys

>>>>此时就可以无密码登陆

测试本机:

ssh localhost

/////多机; 将.ssh/id_dsa.pub 复制到所有节点的.ssh/authorized_keys 中, 没有的新建.ssh 文件

测试 ssh node5(这里 node5 是机名，其实是 hosts 中对应 ip 的名字)
如果不需要则成功！

5- 安装 java

copy 到机器的 usr

```
(sudo)cp tools/jdk-6u20-i586.bin /usr
```

并在每台机器的 root 用户下面安装.

```
$ cd /usr
```

```
$ chmod +x jdk-6u20-linux-i586.bin 给安装文件增加执行权限.
```

```
$ ./jdk-6u11-linux-i586.bin,按提示按几个空格健后,输入 yes 后开始安装 jdk6.
```

安装好后,将目录名修改为 jdk6.

设置 JDK 的环境变量，考虑到 JDK 可能会有其他系统用户也会用到，建议将环境变量直接设置在/etc/profile 中具体内容(如果没有则直接在 profile 文件中添加):

用\$ source /etc/profile 使用 java 环境生效.

用 java -version 检验是否成功。

5*- 获取文件编辑权限

```
sudo gedit '/etc/profile'
```

如果第二次 namenode 重新生成密钥，datanode 就连不上了，要把.ssh/known_hosts 删掉。。

Java Notes

(2011-7-23) a==b ? "dd" : "h"

if you want to print something like this:

```
String str = a==b ? "dd" : "h" + "\n" + "ok";
```

it will be wrong unless you rewrite it as:

```
String str = (a==b ? "dd" : "h") + "\n" + "ok";
```

(2011-7-23) iterator & container

if you use iterator, you can just use it only once, it will not work anymore,

and it is bad idea to get more than one iterators,

if you use container, it works in nested iteration.

(2011-7-28) iterator

Every iterator you get from container can be used only once, but you can get a new one at every time, and this new one iterator is also from start.

(2011-7-28)logger and system.out

logger and system.out will not act as the order you make.

(2011-7-28)continue not return

it is a mistake problem i make,

in elp0.0.12, which acts like that, when i nestly iterate the values, it seems like it doesnot check all the cases.

because i use "return" in nested iteration, if i want to check all the mistakes, i should use "continue" instead of

"return"

(2011-7-28) How to get Long random

as follows:

```
*****
```

```
long i = 10;
Random r = new Random();
System.out.println(Math.abs(r.nextLong() % i));
*****
```

(2011-7-28)log4j

log4j 的配置文件放到与 src 同级的根目录下，采用 `PropertyConfigurator.configure("log4j-config");`函数激活配置文件。

(2011-10-24) Java Integer

Java Integer int 的可以取到的最大值为 2 的 31 次方减 1，即为 2147483647，为 21 亿 4 千万多

(2011-9-12)

I have found a good way to read files big than M order of scalability which can not be opened by common editors.

****use FileInputStream to read the file, iterate it and print every char type variable using standard output.**

****you should not be supposed to use a string to store this input for the sluggish processing.**

(2011-9-14) The scalability of long:

```
System.out.println(Long.MAX_VALUE);//最大数:9223372036854775807
System.out.println(Long.MIN_VALUE);//最小数:-9223372036854775808
that shows long is enough for ID.
```

(2011-9-14)

```
*****
```

i found that if you get bytes from a string, it will treat every char as one byte.

e.g. `byte[] bytes = "iamTom"` the length of bytes will be 6.

a short string occupies less space than long, so i use string to translate data.

(2011-9-20/21/22/23)

if you sate an object having a name same as Class name with first letter as lowercase will change selvely with the changing of Class name, but on the other case, this will not happen.

(2011-9-20/21/22/23) how to combine two arries?

```
*****
int[]  a3=new  int[a1.length+a2.length];
System.arraycopy(a1,0,a3,0,a1.length);
System.arraycopy(a2,0,a3,a1.length,a2.length);
*****
```

(2011-9-24)

****java : java.io.FileNotFoundException:** D:\Project\elp4mr\0.2.5 实验\tmp\subClassOf\iter-0
(拒绝访问。)

if you look out a file just with its parent path, this exception will be throwed,
you should make sure that file name is added! like:
D:\Project\elp4mr\0.2.5 实验\tmp\subClassOf\iter-0\1.txt

** 关于写出对象

打开对象写出流，不关的话可以一直写出去，但是假如关的话，再重新新建对象，会重头开始写，覆盖掉原来所有的内容！要注意！

(2011-9-25/26/27/28)About java debugging

you should allocate a identifier to a new object, otherwise, to do debugging will throw "can not find sorce files"

exception.

`new T().test();` should be `T t = new T(); t.test();`

***使用单步执行过滤器，可以不用进入子函数。

(2011-10-1/6) case penetrating

there is a problem in switch sentences, if you donot add break, whatever case penetrating will happen.

(2010-11-24) if you add two same Integer objects in HashSet

1- if you add two same Integer objects in HashSet, they will be reduced to single.

(2010-11-24) one example about log4j

2- one example about log4j

```
public static void main(String[] args) {
    Logger logger = Logger.getLogger(TestLogger.class.getName());
    logger.setLevel(Level.ALL);
    try {
        FileHandler fileHandler = new FileHandler("D:\\Project\\elp4mr\\0.2.6 实验
\\testlogger.txt");
        fileHandler.setLevel(Level.ALL);
        fileHandler.setFormatter(new SimpleFormatter());
        logger.addHandler(fileHandler);
    } catch (SecurityException e) {
        e.printStackTrace();
    } catch (IOException e) {
        e.printStackTrace();
    }
    logger.fine("良好");
}
```

(2011-10-26) How to readline

```
try {
    FileReader fileReader = new FileReader(file);
    BufferedReader bf=new BufferedReader(fileReader);
    String axiom;
    try {
```

```

        axiom = bf.readLine();
        while(axiom != null){
            addNewAxiom(axiom);
            axiom = bf.readLine();
        }
    } catch (IOException e) {
        e.printStackTrace();
    }
} catch (FileNotFoundException e) {
    e.printStackTrace();
}
}

```

(2011-10-26)java heap space

-Xms512m -Xmx512m

(2011-10-30)

- 1- You can not use wildcard strategy.
- 2- You can use interface as the key-value of mapper or reducer.
- 3- 不同字段，串行序列的值也不同。

(2011-11-1)

Try to use byte array as information cell.

Long's MAX_VALUE = 9223372036854775807

(2011-11-1)How to de-encode Long from-to byte[]

```

/*
public static long decodeLong(byte[] value, int start) {
    ByteBuffer buf = ByteBuffer.wrap(value, start, value.length - start);
    return buf.getLong();
}

public static void encodeLong(byte[] value, int start, long number) {
    ByteBuffer buf = ByteBuffer.wrap(value, start, value.length - start);

```

```

        buf.putLong(number);
    }
}
*/

```

putLong(number) will use 8 bytes.

3- Long has 8 bytes, while Integer has 4 bytes.

```

4- byte[] bKey = key.getBytes();
    oKey.set(bKey, 0, 9);

```

(2011-11-5)使用 System.arraycopy

```

int[] src={1,3,5,6,7,8};
int[] dest = new int[6];

```

```

System.arraycopy(src, 0, dest, 0, 6);

```

System 提供了一个静态方法 arraycopy(),我们可以使用它来实现数组之间的复制.
其函数原型是:

```

public static void arraycopy(Object src, int srcPos, Object dest, int destPos, int length)

```

src:源数组; srcPos:源数组要复制的起始位置;
dest:目的数组; destPos:目的数组放置的起始位置;
length:复制的长度.

注意: src and dest 都必须同类型或者可以进行转换类型的数组.

有趣的是这个函数可以实现自己到自己复制,

比如: int[] fun ={0,1,2,3,4,5,6};

```

System.arraycopy(fun,0,fun,3,3);

```

(2011-11-11/12/18/21)最近用到了 Set 的 toArray 方法

```

/*****
最近用到了 Set 的 toArray 方法

```

首先我的 Set 种保存的是 String 类型的数据,我想把这个 Set 转换为 String 的数组,于是就用到 Set 的 toArray()方法:

view plainprint?

```
1. Set<String> set = new HashSet<String>();
2. ....
3. String[] strs = (String[]) set.toArray();
```

```
Set<String> set = new HashSet<String>(); ..... String[] strs = (String[]) set.toArray();
```

程序编译的时候很正常，当运行的时候却报了 `ClassCastException`，于是就纳闷了：

`String` 是 `Object` 的子类，而 `set.toArray()` 返回的是一个 `Object` 数组，为什么就不能转换呢？

百思不得其解，于是搜了写资料，才终于弄明白了缘由。

第一点：`String` 是一种类型，`String[]`是另一种类型，这是不同的概念。

`Object` 可以强转为 `String`（只要可以）不代表 `Object[]`类型可以强转为 `String[]`类型。

第二点（此段话摘抄自 `steedhorse` 在一篇帖子中的回复）：对于 `Set` 而言，它只知道它内部保存的是 `Object`，所以默认情况下，`toArray` 只能是返回一个由这些 `Object` 构成的 `Object` 数组出来。但程序的作者或许更清楚其内部元素的更具体的类型，因此，`HashSet` 类提供了 `toArray` 的另一个重载版本，允许用户指定一种比 `Object[]`更具体的数组类型，方法是传递一个用户想要的数组类型的一个数组实例进去，多长都无所谓（因此我们常常使用一个 `0` 长度的，毕竟把类型带进去就 `OK` 了），于是，`toArray` 内部就会按照你想要的这种类型，给构造一个数组出来。这样构造出来的数组，当然是很安全地被调用者转换回那个实际的类型。

`String` 和 `String[]`不是一种类型，这可算是常识了，可是真到用的时候就犯了想当然的错误。

还真是搞笑....

```
/******
```

(2011-11-19/24/26/28)正则表达式

. can match " " and \t, but not \n and \r

- 1 `[.]*` is not allowed;
- 2 `.` 是一下子将后面的都读进来再看（←这个解释是错误的）。

关于正则表达式的使用规则：

- a- 首先将要泛化的模式的所有可能实例拿出来；
 - b- 以串行的顺序抽取模式即可，不必从整体到局部。
-

3 []* 是可以的

4 关于组的计算是从左起数左括号，组号等于括号序号，括号范围是内容，0 号是整体；

(2011-11-29)

1- 取消过程解码

2- 取消冗长包名

(2011-12-3/4)

1- eclipse 自动导入包 ctrl+shift+o

2- java % 运算对负数处理后是负数。

如 $-1 \% 7 = -1$

正常的应该是 6

(2011-12-15) 如何在 ubuntu 下运行 jar

如何在 ubuntu 下运行 jar

配置 java 相应环境变量：

```
export JAVA_HOME=/usr/lib/jvm/java-6-sun
```

```
export PATH=$PATH:/usr/lib/jvm/java-6-sun/bin
```

```
export CLASSPATH=./:/usr/lib/jvm/java-6-sun/lib/tools.jar:/usr/lib/jvm/java-6-
```

```
sun/lib/dt.jar:/usr/lib/jvm/java-6-sun/lib/rt.jar:/usr/lib/jvm/java-6-sun/lib
```

运行

```
java -jar XXX.jar [参数]
```