OOM cases in MapReduce jobs in StackOverflow.com

1 (Unsolved)

Q: Java Heap space error in Hadoop

User: I am currently reading file of 50 MB in reducer setup step successfully, But when file is larger than that approx(500MB) it gives me "out of memory error".

Expert:

asked 14 hours ago by Sanjay Bhosale

2

Q: Hadoop Error: Java heap space

So, after seeing the a percent or so of running the job I get an **error** that says, "**Error**: Java **heap** space" and then something along the lines of, "Application container killed" I am literally ... running an empty map and reduce job. However, the job does take in an input that is, roughly, about 100 gigs. For whatever reason, I run out of **heap** space. Although the job does nothing. I am using default ...

hadoop

asked may 7 by Xibz

3

Q: Java heap space error while running hadoop

Every time I try to run a **hadoop** jar , I'm getting Java **heap** space **error**. I have installed **hadoop**-1.2.1 in my VMware. This is my conf/mapred-site.xml: I have also added the following to **hadoop** ... -env.sh: Even after, each time I run the job it shows Java **heap** space exception and the job fails. Could some one help me out. Whenever I run my mapreduce my Linux becomes very slow. Too slow that it takes long even to open a browser. ...

java hadoop heap-memory

asked sep 6 by NeethuPL

4

Q: Hadoop java mapper -copyFromLocal heap size error

As part of my Java mapper I have a command executes some code on the local node and copies a local output file to the **hadoop** fs. Unfortunately I'm getting the following output: **Error** occurred ... during initialization of VM Could not reserve enough space for object **heap** I've tried adjusting mapred.map.child.java.opts to -Xmx512M, but unfortunately no luck. When I ssh into the node, I can ...

java jvm hadoop heap mapreduce

asked sep 23 '11 by Joris

5

Q: Heap error when using custom RecordReader with large file

Guide' but I get a **heap error** when trying to read in to the BytesWritable. I believe this is because the byte array is of size 85713669, but I'm not sure how to overcome this issue. Here is the code: } ... I've written a custom file reader to not split my input files as they are large gzipped files and I want my first mapper job to simply gunzip them. I followed the example in '**Hadoop** The Definitive ...

java hadoop heap

asked feb 11 '13 by Shane

6

Q: Out of heap error when creating Index in Apache Hive

of RAM, 8 cores and 8 data disks. We use **Hadoop** v 2.4.1 and Hive v 0.13.1 . Before we got this far we had problems with running out of **heap** in the hive “console” but after increasing the max size we … instead run into this new problem. We have tried various **heap** sizes (and garbage collectors/settings) but it seems to not make any difference (or at least not make the job complete without **errors** …

indexing out-of-memory hive

asked aug 26 by Magnus Eriksson

6

Q: Error: Java heap space

In Ubuntu, when I am running the **hadoop** example : In log, I am getting the **error** as : INFO mapred.JobClient: Task Id : attempt_201303251213_0012_m_000000_2, Status : FAILED **Error**: Java …**heap** space 13/03/25 15:03:43 INFO mapred.JobClient: Task Id :attempt_201303251213_0012_m_000001_2, Status : FAILED **Error**: Java **heap** space13/03/25 15:04:28 INFO mapred.JobClient: Job Failed …

hadoop

asked mar 25 '13 by Senthil Porunan

6

Q: CDH 4.1: Error running child : java.lang.OutOfMemoryError: Java heap space

I've trying to overcome sudden problem. Befor that problem I've used old VM. I've downloaded the new one VM and still can't make my job run. I get Java **heap** space **error**. I've alredy read this one … post: out of Memory **Error** in **Hadoop** Here is my configs from /etc/**hadoop**/conf: sudo vi **hadoop**-env.sh Here is my mapred-site.xml Nothing helps :(Here is the console output: Here is my log: What can I try next? Please help. Thank you. …

hadoop mapreduce cloudera yarn

asked dec 2 '12 by Sergey

6

Q: hadoop mapper over consumption of memory(heap)

and with the mapper code I dont expect my **heap** memory to ever reach 256M, yet it fails with java **heap**space **error**. I will be thankful if you can give some insight into why the mapper is consuming so much memory. EDIT: … I wrote a simple hash join program in **hadoop** map reduce. The idea is the following: A small table is distributed to every mapper using DistributedCache provided by **hadoop** framework. The large table …

hadoop mapreduce hashmap mapper

asked mar 9 '13 by mc_87

6

Q: increase jvm heap space while runnig from hadoop unix

I am running a java class test.java from **hadoop** command : I am using a stringBuilder, and its size is going out of memory : I know in java we can run a java program by providing a **heap** space … size : How can I do this while running with **hadoop**, if I run it like : it throws exception : I would like to change the **heap** size permanently. …

java unix exception hadoop

asked aug 31 '13 by naveen kumar

6

Q: hadoop windows (work ok) linux java heap space

) (using cygwin) on a 4 GB RAM machine, the application works fine, but when I run it on linux- ubuntu on a 2 GB RAM machine, it process some rows but then it throws a "Java **heap** space" **error**, or sometimes ... the thread is killed. For the linux: I already tried to change the **hadoop** export HEAP_SIZE and also the Xmx and Xms parameters on the app and it made some difference but not too much, the **error** ...

java linux heap hadoop space

asked nov 1 '10 by KoRnE

6

Q: PIG using HCatLoader, Java heap space error

- Connected to metastore. 2014-04-11 01:12:37,785 [main] **ERROR** org.apache.pig.tools.grunt.Grunt -**ERROR** 2998: Unhandled internal **error**. Java **heap** space Details at logfile: /home/aphadoop/pig_1397158893348.log Does anybody encountered the same **error** and solved, Please let me know. ... I am using hive-0.12.0,pig-0.12.0,mysql-5.6 and **hadoop**-1.2.1 in pseudo distribution mode. I configured PIG_CLASSPATH etc.. details according to the link, https://cwiki.apache.org/confluence/display ...

hadoop apache-pig hcatalog

asked apr 12 by Dorababu G

6

Q: Error running child : java.lang.OutOfMemoryError: Java heap space

I have read a lot on the internet, but found no solution for my problem. I use **Hadoop** 2.6.0. The main goal for the MapReduce is to run through a SequenceFile and do some analysis on the key/value ... pairs. Here the output from STDOUT My configurations are nearly default, nothing related to the Java **heap** size. I've also tried this, which does not made a difference. The main programm ...

java hadoop

asked 14 hours ago by Christian D.

6

Q: Hadoop: Can you silently discard a failed map task?

I am processing large amounts of data using **hadoop** MapReduce. The problem is that, ocassionaly, a corrupt file causes Map task to throw a java **heap** space **error** or something similar. It would be nice ... , if possible, to just discard whatever that map task was doing, kill it, and move on with the job, never mind the lost data. I don't want the whole M/R job to fail because of that. Is this possible in **hadoop** and how? ...

java hadoop mapreduce

asked jan 9 '14 by miljanm

6

Q: Querstion regarding hadoop-env.sh

I am Facing **Error**: Java **heap** space and **Error**: GC overhead limit exceeded So i started looking into**hadoop**-env.sh. so thats what i understand so far, Please correct me if i am wrong ... of 9GB But tasktracker is invoked with 7GB memory, so this wil conflict. as max memory for tasktracker and child JVMS invoked by tasktracker is 7GB, but they are consuming 9G. So the **heap** space **error** occured, is my calculation correct? ...

java hadoop hadoop-streaming hadoop-partitioning hadoop2

asked jun 27 by user2950086

6

Q: java.lang.OutOfMemoryError on running Hadoop job

tried the solution of assigning more **heap** as mentioned in this thread: out of Memory **Error** in **Hadoop**WordCountMapper code: ... to the WordCount example provided with **Hadoop**. I've 4 classes in all to carry out the processing: StanfordLemmatizer [contains goodies for lemmatizing from Stanford's coreNLP package v3.3.0], WordCount ...

java hadoop stanford-nlp

asked nov 27 '13 by Aditya

6

Q: how to change mapper memory requirement in hadoop?

that lowering mapred.max.split.size was a good idea: more mappers with lower memory requirements. BUT, I got the "java.lang.OutOfMemoryError: Java **heap** space" **error** again, again and again. It seems that, i did not understand how **hadoop** works. Any suggestions? ... In a map-reduce job, i got the **error**"java.lang.OutOfMemoryError: Java **heap** space". Since I get this **error** in a mapper function; I thought that when I lower the input size to the mapper I will have ...

hadoop mapreduce

asked sep 11 '13 by ndemir

6

Q: Hadoop conf to determine num map tasks

a Java **Heap** Space **error**. I've tried setting many different conf properties in my **Hadoop** cluster to make the job split into more tasks but nothing seems to have any effect. I have tried setting ... I have a job, like all my **Hadoop** jobs, it seems to have a total of 2 map tasks when running from what I can see in the **Hadoop**interface. However, this means it is loading so much data that I get ...

hadoop configuration mapreduce hdfs

asked jul 23 '13 by Katie

6

Q: Mahout on Elastic MapReduce: Java Heap Space

I'm running Mahout 0.6 from the command line on an Amazon Elastic MapReduce cluster trying to canopy-cluster ~1500 short documents, and the jobs keep failing with a "**Error**: Java **heap** space" message ... . Based on previous questions here and elsewhere, I've cranked up every memory knob I can find: conf/**hadoop**-env.sh: setting all the **heap** spaces there up to 1.5GB on small instances and even 4GB ...

hadoop heap mahout emr

asked apr 29 '12 by David M.

6

Q: Hadoop Streaming Memory Usage

I'm wondering where the memory is used in the following job: **Hadoop** Mapper/Reducer **Heap** Size: Streaming API: Mapper: Reducer: Input File is a 350MByte file containg a single line full ... getting always out of memory **errors**. Is this a possible Bug in **Hadoop** (e.g. unaccary copies) or do I just don't understand some required memory intensive steps? Would be really thankful for any further hints. ...

java memory hadoop mapreduce

asked jul 31 '13 by mt_

6

A: Java Heap Space Error in running mahout item similarity job on Amazon EMR

/TaskConfiguration_H2.html). Before making further changes you will need to determine what process is actually suffering from the **heap error** and then tune accordingly. If it does turn out ... to be container/task jvm **heap** then it can be adjusted with configure-**hadoop**(http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/emr-plan-bootstrap.html#PredefinedbootstrapActions_ConfigureHadoop). Also drop the memory-intensive bootstrap action, it is only for AMI 1.x. ...

answered jan 6 by ChristopherB

6

Q: Writing a Hadoop Reducer which writes to a Stream

I have a **Hadoop** reducer which throws **heap** space **errors** while trying to produce very long output records. Is there a way to write a Reducer to use Streams for output, so that I can run through the data for the record without marshalling the whole record into memory? ...

hadoop streaming reducers

asked sep 10 by mcintyre321

6

Q: Getting java heap space error while running a mapreduce code for large dataset

-500MB. But for datasets above 1GB I am getting an **error** like this: I think my program is consuming too much memory and need to be optimized. I even tried to solve this by increasing my java **heap** space ... I am a beginner of MapReduce programming and have coded the following Java program for running in a**Hadoop** cluster comprising 1 NameNode and 3 DatanNodes : The program is run on a dataset like ...

java hadoop mapreduce

asked apr 13 by Monami Sen

6

Q: Reducer's Heap out of memory

on occasion I also get **errors** about bash failing to get memory for what I assume is the spill operation. Would this be the **Hadoop** node running out of memory? If so would just turning down the **heap** ... So I have a few Pig scripts that keep dying in there reduce phase of the job with the **errors** that the Java **heap**keeps running out of space. To this date my only solution has been to increase Reducer ...

hadoop mapreduce apache-pig piglatin

asked jan 2 '12 by NerdyNick

6

Q: Memory problems with Java in the context of Hadoop

I want to compute a multiway join in **Hadoop** framework. When the records of each relation get bigger from a threshold and beyond I face two memory problems, 1) **Error**: GC overhead limit exceeded, 2 ...) **Error**: Java **heap** space. The threshold is the 1.000.000 / relation for a chain join and a star join. In the join computation I use some hash tables i.e. These **errors** occur when I hash the input ...

java hadoop garbage-collection heap

asked sep 22 '13 by vpap

6

Q: Mahout on EMR Error: Java heap space

I've ran a clustering job on EMR. The dataset is huge. Everything worked well untill: So. The basic question is how to fix that?

java hadoop amazon-web-services mahout emr

asked sep 29 '12 by denys

6

Q: A join operation using Hadoop MapReduce

a HashMap and then take a cross product. (eg. Join of two datasets in Mapreduce/**Hadoop**) This solution is very good and works for majority of the cases but in my case my issue is rather different. I am ... dealing with a data which has got billions of records and taking a cross product of two sets is impossible because in many cases the hashmap will end up having few million objects. So I encounter a **Heap** ...

hadoop mapreduce elastic-map-reduce

asked may 19 '13 by Eastern Monk

6

Q: Building Inverted Index exceed the Java Heap Size

the **error**. Which means that as the data increase number of items/product_ids that I index for words like New or old etc get bigger which cause the **Heap** Size to overflow. So, the question is how can avoid java**heap** size overflow and accomplish this task. ... of data from large system). The building of inverted index get executed as a map reduce job on **Hadoop**. Inverted index is build with the help of scala. Structure of the inverted index as follows ...

scala hadoop avro scalding

asked jul 31 '13 by Null-Hypothesis

6

Q: Mahout Canopy Clustering, K-means Clustering : Java Heap Space - out of memory

centers by tweaking canopy parameters (T1, T2), it works very well. More then a certain number of canopy centers, the jobs keep failing with a "**Error**: Java **heap** space" message at 67% of reduce phase. K ... can find: **hadoop**-env.sh: setting all the **heap** spaces there up to 16GB on namenode and even 8GB on datanodes. mapred-site.xml: adding mapred.{map, reduce}.child.java.opts properties, and setting ...

hadoop heap mahout

asked jun 12 '13 by Nebulach

6

Q: How to handle unsplittable 500 MB+ input files in hadoop?

out of memory **errors**, even with a java **heap** size of 6 GB. Now I wonder how I could split the data so that it better matches hadoop's 64 MB block size. I cannot simply split the big files ... I am writing a **hadoop**MapReduce job that is running over all source code files of a complete Debian mirror (≈ 40 GB). Since the Debian mirror data is on a separate machine and not in the **hadoop** ...

hadoop mapreduce

asked may 25 by Michael

6

Q: Hadoop UniqValueCount Map and Aggregate Reducer for Large Dataset (1 billion records)

, but keep getting "Out of Memory" and Java **heap** size **errors** on **Hadoop** - at the same time, I am able to run this fairly easily on a single box using a Python Set (hashtable, if you will.) I am using a fairly ... I have a data set that has

approximately 1 billion data points. There are about 46 million unique data points I want to extract from this. I want to use **Hadoop** to extract the unique values ...

hadoop mapreduce hadoop-streaming elastic-map-reduce

asked jan 18 '13 by Suman

6

Q: Hive Issue - java.lang.OutOfMemoryError: Java heap space

:597) at org.apache.hadoop.util.RunJar.main(RunJar.java:208) FAILED: Execution **Error**, return code -101 from org.apache.hadoop.hive.ql.exec.DDLTask My Environment variables as defined in **hadoop**-env.sh are Definition of hive-env.sh Where could the problem be? Please advise. ... **heap** space at org.apache.thrift.protocol.TBinaryProtocol.readStringBody(TBinaryProtocol.java:353) at org.apache.thrift.protocol.TBinaryProtocol.readMessageBegin(TBinaryProtocol.java:215 ...

hadoop hive

asked apr 22 by user3528338

6

A: Hadoop example job fails in Standalone mode with: “Unable to load native-hadoop library”

issue is the you get. Check you input, increase the **heap** size, if necessary. You might also have a look at this related question: out of Memory **Error** in **Hadoop** ... The warning tells you that the compression codec is not (properly) installed for **Hadoop**. To install the compression, have a look at: http://code.google.com/p/**hadoop**-snappy/ However, a more serious ...

answered dec 1 '12 by Lorand Bendig

6

Q: Out of memory error in Mapreduce shuffle phase

I am getting strange **errors** while running a wordcount-like mapreduce program. I have a **hadoop** cluster with 20 slaves, each having 4 GB RAM. I configured my map tasks to have a **heap** of 300MB and my ... was googling and found this link but I don't really know what to make of it: **hadoop** common link I don't understand why **hadoop** would experience any problems in copying and merging if it is able ...

hadoop mapreduce

asked oct 10 '13 by DDW

6

A: hive collect_set crashes query

This is probably the memory problem, since aggregates data in the memory. Try increasing **heap** size and enabling concurrent GC (via setting **Hadoop** to e.g). This answer has more information about "GC overhead limit" **error**. ...

answered jan 14 '14 by Nigel Tufnel

6

A: What is the relation between 'mapreduce.map.memory.mb' and 'mapred.map.child.java.opts' in A...

via mapred.map.child.java.opts (or mapreduce.map.java.opts in **Hadoop** 2+). If the mapper process runs out of **heap** memory, the mapper throws a java out of memory exceptions: **Error** ... physical memory used; 970.1 MB of 1.0 GB virtual memory used. Killing container. **Hadoop** mapper is a java process and each Java process has its own **heap** memory maximum allocation settings configured ...

answered sep 20 by user1234883

6

Q: MapReduce jobs in hive-0.8.1-cdh4.0.1 Failed.

Queries in hive-0.8.1-cdh4.0.1 that invoke the Reducer results in Task Failed. The queries having MAPJOIn is working fine but JOIN gives **error**. eg: The log file shows that its due to Java **heap** space problem. ...

hive

asked oct 15 '12 by Jickson T George

6

A: out of Memory Error in Hadoop

After trying so many combinations, finally I concluded the same **error** on my environment (Ubuntu 12.04,**Hadoop** 1.0.4) is due to two issues. Same as Zach Gamer mentioned above. don't forget ... to execute "ssh localhost" first. Believe or not! No ssh would throw an **error** message on Java **heap** space as well. ...

answered nov 16 '12 by etlolap

6

Q: OutOfMemory Error when running the wikipedia bayes example on mahout

i ran mahout wikipedia example with the 7 gig wiki backup.. , but when testing the classifier, i am getting the a OutOfMemory **Error** i have pasted the output below, i set the mahout **heap** size and java **heap** size to 2500m ...

java hadoop mahout

asked apr 9 '12 by Nauman Bashir

6

A: Load snappy-compressed files into Elastic MapReduce

decompress out files. One symptom of this problem was a **heap** space **error**: When I switched to a much larger instance and cranked up the mapred.child.java.opts setting, I got a new **error**: Hadoop's snappy ... The answer is, "it can't be done." At least, not for the specific case of applying **hadoop** streaming to snappy-compressed files originating outside of **hadoop**. I (thoroughly!) explored two main ...

answered apr 4 '13 by Abe

6

A: How to insert data into Parquet table in Hive

What are the **error** messages that you get on hive server side? I had a similar problem. In the hive server log I saw some **heap** memory problems. I could solve the problem on my **hadoop** installation using higher values in mapred-site.xml ...

answered apr 26 by woopi

6

A: pig join gets OutOfMemoryError in reducer when mapred.job.shuffle.input.buffer.percent=0.70

Generally speaking, mapred.job.shuffle.input.buffer.percent=0.70 will not trigger OutOfMemory **error**, because this configuration ensures at most 70% of reducer's **heap** is used to store shuffled data ... % of**heap** in shuffle phase, which may cause OutOfMemory **error**. But in general, Pig does not have combine() in Join operator. 2) JVM manages the memory itself and divides its **heap** into Eden, S0, S1 and old ...

answered aug 14 '13 by Lijie Xu

6

A: Hive: SQL request to split a table into N tables of approximately the same size?

First of all, I would dig into why you're getting **heap** size **errors**. This usually indicates a misconfigured cluster. In theory, Hive/**Hadoop** should be able to do almost everything by streaming to/from ...

answered may 17 by Joe K

6

A: Loading gz file with pig and storing it into HBase with COMPRESSION => 'GZ'

at the file extension to see what type of compression codec it was used to compress and calls the relevant decompressor. **Hadoop** first checks all the codecs installed and will report an **error** if it cant find ... the codec required for your file. So, .And this is the reason of **heap** size **error** in some mapreduce job as mapred child cant get enough space to keep the uncompressed data. . Also, there are only 3 ...

answered may 8 by Chandra kant

6

Q: How to run large Mahout fuzzy kmeans clustering without running out of memory?

I am running Mahout 0.7 fuzzy k-means clustering on Amazon's EMR (AMI 2.3.1) and I am running out of memory. My overall question: how do I get it working most easily? Here is an invocation: Mor ...

hadoop cluster-analysis mahout k-means

asked apr 19 '13 by dfrankow

6

Q: OutofMemoryError when reading a local file via DistributedCache

/standalone mode). However, when I execute it in the **Hadoop** cluster (fully distributed mode), I get an "**OutOfMemoryError**: Java heap space", with the same 40 MB file. I don't understand why this happens, as the file isn't that large. This is the code: Any help would be appreciated, thanks in advance. ... understand why the program works locally but it doesn't distributedly. Thanks for all the answers. I'm using **Hadoop**1.0.3. The cluster is composed of three machines, all of them running Ubuntu Linux 12.04 ...

hadoop mapreduce out-of-memory

asked nov 19 '12 by iporto

6

Q: how to determine the size of “mapred.child.java.opts” and HADOOP_CLIENT_OPTS in mahout canopy

Is size is about 50M. mapred-site.xml: **hadoop**-env.sh: When running mahout canopy,it always throws**OutOfMemoryError**: Question is,how to determine the size of "mapred.child.java.opts" and HADOOP_CLIENT_OPTS? ...

mahout

asked dec 6 by user3156087

6

Q: How to use “typedbytes” or “rawbytes” in Hadoop Streaming?

the binary file to HDFS with "**hadoop** fs -copyFromLocal". When I try to use it as input to a map-reduce job, it fails with an **OutOfMemoryError** on the line where it tries to make a byte array ... I have a problem that would be solved by **Hadoop** Streaming in "typedbytes" or "rawbytes" mode, which allow one to analyze binary data in a language other than Java. (Without this, Streaming ...

hadoop binary streaming

asked mar 2 '13 by Jim Pivarski

6

Q: Mahout - Exception: Java Heap space

I'm trying to convert some texts to mahout sequence files using: But all I get is a **OutOfMemoryError**, as here: I am using Mahout 0.9, **Hadoop** 1.2.1 and OpenJDK Java7u25 defining MAHOUT_HEAPSIZE ...

hadoop mahout

asked apr 7 by user3422072

6

Q: HBase: How to handle large query results

I have a table with almost 3 million records in it. I want to be allow a user to query large sections of it, or even the whole table. I'm new to hbase/**hadoop** (and fairly inexperienced in db ... function. I could just pass the Scanner, but I want to close it. My problem is when the ArrayList gets filled to about 2 million records, I get an **OutOfMemoryError**: I'm pretty sure my approach ...

java hadoop out-of-memory hbase

asked jul 21 by Bryany

6

A: Hadoop Pipes: how to pass large data records to map/reduce tasks

Hadoop is not designed for records about 100MB in size. You will get **OutOfMemoryError** and uneven splits because some records are 1MB and some are 100MB. By Ahmdal's Law your parallelism will suffer ... greatly, reducing throughput. I see two options. You can use **Hadoop** streaming to map your large files into your C++ executable as-is. Since this will send your data via stdin it will naturally ...

answered oct 26 '10 by Spike Gronim

6

Q: OOM exception in Hadoop Reduce child

I am getting **OOM** exception (Java heap space) for reduce child. In the reducer, I am appending all the values to a StringBuilder which would be the output of the reducer process. The number of values ... removed it later on. Some sample sizes of iterator are as follows: 238695, 1, 13, 673, 1, 1 etc. These are not very large values. Why do I keep getting the **OOM** exception? Any help would be valuable to me. Stack trace ...

hadoop mapreduce out-of-memory

asked oct 11 '12 by Raghava

6

Q: setting hadoop job configuration programmatically

I am getting **OOM** exception (Java heap space) for reduce child. I read in the documentation that increasing the value of to -Xmx512M or more would help. Since I am not the admin, I cannot change ... that value in mapred-site.xml. I would like to set that value only for my job through the java program. I tried setting it using class as follows, but that didn't work. The version of **Hadoop** is 1.0.3 What is the proper way of setting the configuration values programmatically? ...

configuration hadoop mapreduce

asked oct 10 '12 by Raghava

6

A: -XX:OnOutOfMemoryError="kill -9 %p" Problem

Running as a **hadoop** option I run in to the same issues. This was the answer: Here is stdout on **OOM**: I also tried: It started, but on **OOM** it But STDERR has: sh: kill -9 1164: command not found These Wouldn't even start: ...

answered may 23 '13 by Jeremy

6

Q: Is it possible to intervene if a task fails?

I have a mapreduce job running on many urls and parsing them. I need way to handle a scenario in which one parsing task crashes on a fatal error like **OOM** error. In the normal **hadoop** behaivour a task ...

hadoop

asked may 8 '12 by user1251654

6

Q: Limit CPU / Stack for Java method call?

I am using an NLP library (Stanford NER) that throws **OOM** errors for rare input documents. I plan to eventually isolate these documents and figure out what about them causes the errors ... , but this is hard to do (I'm running in **Hadoop**, so I just know the error occurs 17% through split 379/500 or something like that). As an interim solution, I'd like to be able to apply a CPU and memory limit ...

java nlp stanford-nlp

asked jul 4 '09 by Kevin Peterson

6

Q: out of Memory Error in Hadoop

suggest a solution so that i can try **out** the example. The entire Exception is listed below. I am new to**Hadoop** I might have done something dumb . Any suggestion will be highly appreciated. ... I tried installing**Hadoop** following this http://hadoop.apache.org/common/docs/stable/single_node_setup.html document. When I tried executing this I am getting the following Exception Please ...

java hadoop

asked dec 11 '11 by Anuj

6

Q: Hadoop YARN Map Task running out of physical and virtual memory

I have the following method that I run from my map task in a multithreaded execution , however this works fine in a standalone mod e, but when I runt this in **Hadoop** YARN it runs **out** of the physical ...

java hadoop selenium-webdriver yarn ghostdriver

asked jan 7 '14 by user1965449

6

Q: Reducer's Heap out of memory

on occasion I also get errors about bash failing to get memory for what I assume is the spill operation. Would this be the **Hadoop** node running **out** of memory? If so would just turning down the heap ... So I have a few Pig scripts that keep dying in there reduce phase of the job with the errors that the Java heap keeps running **out** of space. To this date my only solution has been to increase Reducer ...

hadoop mapreduce apache-pig piglatin

asked jan 2 '12 by NerdyNick

6

Q: Dropping Hive table throws Out of memory error

analytics rename to analytics_backup.It was hanging in the terminal for 30-45 mins and then throws the**out**-of-memory error. Is there anyone have noticed this kind of issue and any solution to overcome this. I am using the CDH3 **Hadoop**/Hive version. Thanks in advance. ...

out-of-memory hive

asked may 19 by Siva

6

Q: Why the identity mapper can get out of memory?

In an reduce-only **Hadoop** job input files are handled by the identity mapper and sent to the reducers without modification. In some job of mine I got very surprised to see the job failing in the map ... phase with "**Out** of memory error" and "GC overhead limit exceeded". In my understanding, a memory leak on the identity mapper is **out** of the question. What can be the cause of such error? ...

java hadoop out-of-memory amazon-emr

asked sep 6 '12 by sortega

6

Q: Too slow or out of memory problems in Machine Learning/Data Mining [closed]

resorting to heavy-weight **hadoop** like systems] What I meant to learn from the community: (subjectively) your real life problems/algorithms that are not TOO large to be called big data, but still big ... difficulties with because they are too slow or need excessively large memory? As a hobby research project we built an **out**-of-core programming model to handle data larger than system memory and it natively ...

parallel-processing machine-learning analytics data-mining large-data

asked mar 14 '13 by mvarshney

6

Q: hadoop mapper over consumption of memory(heap)

I wrote a simple hash join program in **hadoop** map reduce. The idea is the following: A small table is distributed to every mapper using DistributedCache provided by **hadoop** framework. The large table ...) on the hashmap, and if the key exists in the hash map it is written **out**. There is no need of a reducer at this point of time. This is the code which we use: While testing this code, our small table ...

hadoop mapreduce hashmap mapper

asked mar 9 '13 by mc_87

6

Q: What does “Usage threshold is not supported” mean with Hadoop PIG?

I'm running **Hadoop** PIG, and all tasks fails because of the following exception: So, what does it mean? I tried to google for it, but to no avail. Am I **out** of memory? EDIT Running and , java ...

java hadoop apache-pig

asked sep 3 '13 by warl0ck

6

Q: Java Heap space error in Hadoop

I have setup **hadoop** cluster with single node as OS : CentOS Linux release 7.0.1406 **Hadoop** : 2.6.0 Java : 1.7 My problem is with **hadoop** heap memory. I am currently reading file of 50 MB in reducer ... setup step successfully, But when file is larger than that approx(500MB) it gives me "**out** of memory error". When i browse through http:/ip_address:50070/dfshealth.html#tab-overview it gives me below ...

java linux hadoop memory

asked 15h ago by Sanjay Bhosale

6

Q: Will reducer out of java heap space

I am implementing a program using **Hadoop**. My question is how to deal with java **out** of space problem, I added some property configuration into xml file, but it didn't work. Increasing number ... as key, and column vector as value.Is there any way I can get **out** of this dilemma? PS:I was first thinking that reducer will take column one by one, and that won't cause **out** of memory issue ...

hadoop

asked apr 20 '13 by shirley

6

Q: increase jvm heap space while runnig from hadoop unix

I am running a java class test.java from **hadoop** command : I am using a stringBuilder, and its size is going**out** of memory : I know in java we can run a java program by providing a heap space ... size : How can I do this while running with **hadoop**, if I run it like : it throws exception : I would like to change the heap size permanently. ...

java unix exception hadoop

asked aug 31 '13 by naveen kumar

6

Q: Use wget with Hadoop?

on my local acccount on the remote machine), but no luck. Does such a way even exist? Any other suggestions to get this working? I've already tried using Yahoo! VM which comes pre-configured with **Hadoop**, but it's too slow and plus runs **out** of memory since the dataset is large. ... I have a dataset (~31GB, zipped file with extension .gz) which is present on a web location, and I want to run my **Hadoop** program on it. The program is a slight modification from the original ...

java hadoop mapreduce wget

asked nov 28 '13 by Aditya

6

Q: java.lang.OutOfMemoryError on running Hadoop job

to the WordCount example provided with **Hadoop**. I've 4 classes in all to carry **out** the processing: StanfordLemmatizer [contains goodies for lemmatizing from Stanford's coreNLP package v3.3.0], WordCount ... tried the solution of assigning more heap as mentioned in this thread: **out** of Memory Error in**Hadoop** WordCountMapper code: ...

java hadoop stanford-nlp

asked nov 27 '13 by Aditya

6

Q: Out of memory due to hash maps used in map-side aggregation

MY Hive Query is throwing this exception. I tried this on 8 EMR core instances with 1 large master instance on 8Gb of data. First i tried with external table (location of data is path of s3). After ...

hadoop hive amazon-emr hiveql

asked may 22 '13 by Naresh

6

Q: Hadoop searching words from one file in another file

to write to another output file I tried a few examples in **hadoop**. I have two questions Two files are approximately 200MB each. Checking every word in another file might cause **out** of memory ... I want to build a **hadoop** application which can read words from one file and search in another file. If the word exists - it has to write to one output file If the word doesn't exist - it has ...

hadoop mapreduce

asked jan 24 '10 by Boolean

6

Q: Out of heap error when creating Index in Apache Hive

of RAM, 8 cores and 8 data disks. We use **Hadoop** v 2.4.1 and Hive v 0.13.1 . Before we got this far we had problems with running **out** of heap in the hive “console” but after increasing the max size we ... start to see jobs failing and after this point most of the remaining jobs actually fail. We have a 10 machine **Hadoop**cluster (2 pure admin and 8 slave machines). The slave machines each have 72 GB ...

indexing out-of-memory hive

asked aug 26 by Magnus Eriksson

6

Q: Hadoop UniqValueCount Map and Aggregate Reducer for Large Dataset (1 billion records)

, but keep getting "**Out** of Memory" and Java heap size errors on **Hadoop** - at the same time, I am able to run this fairly easily on a single box using a Python Set (hashtable, if you will.) I am using a fairly ... I have a data set that has approximately 1 billion data points. There are about 46 million unique data points I want to extract from this. I want to use **Hadoop** to extract the unique values ...

hadoop mapreduce hadoop-streaming elastic-map-reduce

asked jan 18 '13 by Suman

6

Q: CDH 4.1: Error running child : java.lang.OutOfMemoryError: Java heap space

post: **out** of Memory Error in **Hadoop** Here is my configs from /etc/**hadoop**/conf: sudo vi **hadoop**-env.sh Here is my mapred-site.xml Nothing helps :(Here is the console output: Here is my log: What can I try next? Please help. Thank you. ...

hadoop mapreduce cloudera yarn

asked dec 2 '12 by Sergey

6

A: Hadoop example job fails in Standalone mode with: “Unable to load native-hadoop library”

issue is the you get. Check you input, increase the heap size, if necessary. You might also have a look at this related question: **out** of Memory Error in **Hadoop** ... The warning tells you that the compression codec is not (properly) installed for **Hadoop**. To install the compression, have a look at: http://code.google.com/p/**hadoop**-snappy/ However, a more serious ...

answered dec 1 '12 by Lorand Bendig

6

Q: Shuffle, merger and fetcher errors when processing large files in hadoop

I am running a word-count like mapreduce job processing 200 files of 1Gb each. I am running the job on a**hadoop** cluster comprising 4 datanodes (2cpu each) with 8Gb of memory and about 200G of space ... output I still get errors in the reducer phase. I use 4 reducers. Thus I have tried various configurations of the**hadoop** cluster: The standard configuration of the cluster was: This configuration ...

hadoop configuration mapreduce out-of-memory shuffle

asked may 29 by CSDS

6

A: running an elementary mapreduce job with java on hadoop

. In this case you should be able to tell **hadoop** to allocate you the memory. Try the following. the option -Xmx1G says allow up 1 Gigabyte. This other stackoverflow question is also very similar. **out** of Memory Error in **Hadoop** ...

answered may 1 '13 by greedybuddha

6

A: out of Memory Error in Hadoop

I installed **hadoop** 1.0.4 from the binary tar and had the **out** of memory problem. I tried Tudor's, Zach Garner's, Nishant Nagwani's and Andris Birkmanis's solutions but none of them worked for me ... . Editing the bin/**hadoop** to ignore $HADOOP_CLIENT_OPTS worked for me: I'm assuming that there is a better way to do this but I could not find it. ...

answered nov 6 '12 by Brian C.

6

Q: Apparent memory-leak in hadoop

I have an apparent memory leak in a **hadoop** program I'm running. Specifically I get the message: ERROR GC overhead limit exceeded followed later by the exception I'm running on what should ... be very small data sets in an initial trial, so I shouldn't be hitting any memory limit. More to the point I don't want to change the **hadoop** configuration; if the program can't run with the current ...

java memory-leaks hadoop

asked nov 30 '12 by dsollen

6

Q: Hadoop: Heap space and gc problems

algorithm works fine for small datasets, but for a medium dataset has heap space problems. My algorithm reaches a certain tree level and then it goes **out** of heap space, or has gc overhead problems ... . At that point, i made some calculations and i saw that every task doesnt need more than 100MB memory. So for 8 tasks, i am using about 800MB of memory. I don't know what is going on. I even updated my **hadoop** ...

java garbage-collection hadoop heap multicore

asked mar 14 '12 by jojoba

6

A: Hadoop sorting for massive data

. That being said, **Hadoop** design strength is in distributed sorting (the magic that happens between the mapper and reducer) so if running **out** of memory is your concern, you want to organize your data ... First, understand that **Hadoop** is designed for batch processing (think 18-wheeler not Maserati) so if this search has a constrained time limit to your users, **Hadoop** is not the right tool for the job ...

answered jun 13 '13 by Engineiro

6

A: Hadoop JobClient: Error Reading task output

the files are "divided" into blocks the system runs **out** of memory pretty quickly. So to solve this you have to "fill" the blocks and arrange your new files so that they are spread nicely into blocks. **Hadoop** ... I had a similar problem and was

able to find a solution. The problem lies on how **hadoop** deals with smaller files. In my case, I had about 150 text files that added up to 10MB. Because of how ...

answered mar 15 by Enrique

6

Q: Computing simple moving average using Map Reduce in MongoDB

, why is it that **Hadoop** Reduce phase doesn't crash **out** of memory, since it has to deal with at least several TBs of mapped data. ... I stumbled upon this article: http://blog.cloudera.com/blog/2011/04/simple-moving-average-secondary-sort-and-mapreduce-part-3/ which mentions how to calculate moving average using **Hadoop** ...

mongodb hadoop mapreduce

asked may 16 '13 by P.Prasad

6

A: How many Mapreduce Jobs can be run simultaneously

of running some **hadoop** code, which requires some amount of memory, so eventually you would run **out** of memory on your machine. You might also have to configure job queues cleverly in order to run a ton at the same time. Now, what is possible is a very different question than what is a good idea... ...

answered oct 30 '13 by Joe K

6

Q: Cassandra setInputSplitSize is not working properly

I am using **Hadoop** + Cassandra. I use setInputSplitSize(1000) to not overload mappers (and receive **out** of heap memory) as default it is 64K. All together I have only 2M lines to process. Actually ... %. When I check the log, I found 40K-64K rows processed. It is not crashing or giving **out** of memory, but these 2-3 tasks begin in the middle of processing and continue for 2-3 hours after all other have ...

java hadoop mapreduce cassandra

asked aug 11 '11 by Anton

6

A: How to run large Mahout fuzzy kmeans clustering without running out of memory?

mapper by default. When you subtract **out** all the JVM overhead, room for splits and combining, etc, you probably don't have a whole lot left. You set **Hadoop** params in a bootstrap action. Choose ... Yes you're running **out** of memory. As far as I know, that "memory intensive workload" bootstrap action is long since deprecated, so may do nothing. See the note on that page. A should use 384MB per ...

answered apr 20 '13 by Sean Owen

6

A: Pig: Hadoop jobs Fail

Check your logs, increase the verbosity level if needed, but probably you're facing and **Out** of Mem error. Check this answer on how to change Pig logging. To change the memory in **Hadoop** change ...

answered dec 17 by Paulo Fidalgo

6

Q: How to export a large table (100M+ rows) to a text file?

fields, etc.) and store it int a big text file, for later processing with **Hadoop**. So far, I tried two things: Using Python, I browse the table by chunks (typically 10'000 records at a time) using ... the full table with this. Using the command-line

tool, I tried to output the result of my query in form to a text file directly. Because of the size, it ran **out** of memory and crashed. I am currently ...

python mysql database hadoop export

asked jan 18 '13 by Wookai

6

A: Hadoop Pipes: how to pass large data records to map/reduce tasks

Hadoop is not designed for records about 100MB in size. You will get OutOfMemoryError and uneven splits because some records are 1MB and some are 100MB. By Ahmdal's Law your parallelism will suffer ... greatly, reducing throughput. I see two options. You can use **Hadoop** streaming to map your large files into your C++ executable as-is. Since this will send your data via stdin it will naturally ...

answered oct 26 '10 by Spike Gronim

6

A: Running Hadoop: insufficient memory for the Java Runtime Environment to continue

This indicates you have run **out** of virtual memory, try increasing the swap space, or decreasing the heap to leave the rest of your program mroe virtual memory. a 32-bit program is limited to ~3 GB ... On Windows a 32-bit program is limited to around 1.5 GB of virtual memory. As **hadoop** is a big data solution it is typically run on much bigger machines. e.g. 256 GB to 1 TB is not unusual. Given 32 GB is pretty cheap these days I would consider getting at least this much, or a lot more memory. ...

answered jan 11 by Peter Lawrey

6

A: Hadoop: Is it possible to have in memory structures in map function and aggregate them?

way. Pulling tricks like collecting data in memory in a is a minor and sometimes necessary sin, so, nothing really wrong with it. It does mean you really need to know the semantics that **Hadoop** guarantees, test well, and think about running **out** of memory if not careful. ...

answered mar 1 '12 by Sean Owen

6

Q: How to set heap size for EMR Master

I have a job which I am trigger from in EMR. The master triggers the mapper. Once it is done, it loads a heavweight operation in memory and then evenutualy will dump **out**. Right now, the job which ... runs on the cluster fails after a few minutes because it runs **out** of heap space. By default it sets about 1000m on its master Tried the exact action below, but that did not work . The program is still ...

elastic-map-reduce emr

asked aug 6 '13 by user2655578

6

Q: How smart is the Java JVM about GC'ing during a reduce() operation on a long Scala list or S...

OK, let me see if I can explain. I have some code that wraps a Java iterator (from **Hadoop**, as it happens) in a Scala Stream, so that it potentially can be read more than once, by client code that I ... the iterator will be extremely large, so that storing all the items in it will lead to **out**-of-memory errors. However, in general, the situations where the client code needs the multiple-iteration ...

java scala stream garbage-collection jvm

asked sep 24 '12 by Urban Vagabond

6

Q: Detailed dataflow in hadoop's mapreduce?

I am struggling a bit to understand the dataflow in mapreduce. Recently a very demanding job crashed when my disks ran **out** of memory in the reduce phase. I find it difficult to estimate how much disk ... with permutation groups of words. Since identical words need to be joined the reduce function requires a temporary hash map which is always <= 3GB. Since I have 12GB of RAM and my **hadoop** daemons require 1GB ...

java memory hadoop mapreduce

asked oct 21 '13 by DDW

6

Q: Mahout on Elastic MapReduce: Java Heap Space

. Based on previous questions here and elsewhere, I've cranked up every memory knob I can find: conf/**hadoop**-env.sh: setting all the heap spaces there up to 1.5GB on small instances and even 4GB ...

6

Q: hive job stuck at map=100%, reduce 0%

I'm running hive-0.12.0 on **hadoop**-2.2.0. After submitting the query: I get the following errors in the logs: And then the last line repeats every second or so ad infinitum. If I look ... at container logs I see: I've searched for the Exit code 143, but most the stuff **out** there refers to memory issue and I have memory set pretty large (following the advice of Container is running beyond ...

hadoop hive

asked jul 7 by harschware

6

Q: Limit CPU / Stack for Java method call?

I am using an NLP library (Stanford NER) that throws OOM errors for rare input documents. I plan to eventually isolate these documents and figure **out** what about them causes the errors ... , but this is hard to do (I'm running in **Hadoop**, so I just know the error occurs 17% through split 379/500 or something like that). As an interim solution, I'd like to be able to apply a CPU and memory limit ...

java nlp stanford-nlp

asked jul 4 '09 by Kevin Peterson