

1

answer

Q: [hadoop fs -ls out of memory error](#)

I have 300000+ files on a HDFS data directory. When I do a **hadoop** fs -ls and I am getting an **out** of memory error saying GC Limit has exceeded. The cluster nodes have 256 GB of RAM each. How do I fix it? ...

hadoop

hdfs

asked jul 15 by [Sambit Tripathy](#)

11

answers

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](#)

2

answers

Q: [Java Out of Memory exception in Ubuntu when using Flume/Hadoop](#)

I'm getting an **out** of memory exception due to lack of Java heap space when I try and download tweets using Flume and pipe them into **Hadoop**. I have set the heap space currently to 4GB in the mapred ... - site.xml of **Hadoop**, like so: I am hoping to download tweets continually for two days but can't get past 45 minutes without errors. Since I do have the disk space to hold all of this, I am assuming ...

java

hadoop

heap

flume

asked jul 29 '13 by [Andrew Martin](#)

0

answers

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](#)

3

answers

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](#)

1

answer

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](#)

1

answer

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](#)

2

answers

Q: [MongoDB map-reduce slow and runs out of memory](#)

not yet need the overhead of **Hadoop**. For testing purposes I insert 50 million rows of the type With an index on user_id on an EC2 Large Instance. Its a Single instance mongodb (not sharded ...). takes less than a second. However a mapreduce where I wanted to count the number of user entries took all night and returned with an **out** of memory error, either I must be doing something stupid or mongo ...

mongodb

mapreduce

asked aug 9 '11 by [lordOfChaos](#)

1

answer

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](#)

1

answer

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](#)

1

answer

Q: Hadoop YARN: Failed to launch container

I am getting an **out** of memory error on datanodes in my **Hadoop** 2.3.0 cluster when running Hive queries. What settings should I be looking at to prevent the nodemanager from going down? ...

hadoop

asked apr 29 by [poliu2s](#)

0

answers

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](#)

1

answer

Q: Hadoop Balancer fails with - IOException: Couldn't set up IO streams (LeaseRenewer Warning)

I am stumbling across this error while running the **Hadoop** Balancer via Namenode. Anytips on cracking this. The process is also blocking the current user and giving an **Out** of Memory error on issuing any other command. ...

hadoop

asked may 9 by [Yash Sharma](#)

1

answer

Q: Hadoop 2.2.0 Streaming Memory Limitation

therefore it is only possible to start one and only one task per node **out** of the box each **hadoop** instance starts about 7 **hadoop** containers with default **hadoop** settings. each **hadoop** task forks ... the parent memory footprint and since the memory used by our parent was more than half of total memory, we run **out** of memory again. if we lower the map memory, **hadoop** will allocate 2 containers per node ...

bash

memory

hadoop

hadoop-streaming

yarn

asked feb 21 '14 by [Patrick](#)

0

answers

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 1 hour ago by [Sanjay Bhosale](#)

1

answer

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](#)

1

answer

Q: increase jvm heap space while running 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](#)

0

answers

Q: Settings for a Hadoop cluster with out swap partition

I have a smallish **hadoop** cluster running v2.3. Just MapReduce, no HDFS. none of my nodes have a swap partition. what should be my vmem:pmem value so I dont run in to **out** of memory issues when ...

hadoop

mapreduce

yarn

asked apr 16 by [powerrox](#)

1

answer

Q: Use wget with Hadoop?

on my local account 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](#)

1

answer

Q: Hadoop Basic Examples WordCount

I am getting this error with a mostly **out** of the box configuration from version 0.20.203.0 Where should I look for a potential issue. Most of the configuration is **out** of the box. I was able ... to visit the local websites for hdfs, task manager. I am guessing the error is related to a permissions issue on cygwin and windows. Also, googling the problem, they say there might be some kind of **out** ...

hadoop

asked aug 13 '11 by [Berlin Brown](#)

2

answers

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](#)

1
answer

Q: hadoop - tasks keep running after job finished

I running an **hadoop** cluster with 20 nodes. I run a heavy job which uses a lot of memory. for some reason it seems that some tasks get "**out** of touch" with the job. why does it seem so? I have ...

hadoop

asked jul 11 '12 by [user1251654](#)

2
answers

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](#)

1
answer

Q: Hadoop Streaming Memory Usage

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. ... 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 ...

java memory hadoop mapreduce

asked jul 31 '13 by [mt_](#)

0
answers

Q: Hadoop, HAR files and globbing doesn't work

: It was trivial to move these to hdfs and to build a har archive; I used the following command to make the archive Listing the contents of the har (bin/**hadoop** fs -lsr har:///user/**hadoop** ... /epc_test.har) and everything looks as I'd expect. When we come to run the **hadoop** job with this command, trying to wildcard the archive: it fails with the following exception Running the job with the non ...

hadoop har

asked feb 13 '14 by [DWAB](#)

3
answers

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](#)

0
answers

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](#)

1
answer

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](#)

2
answers

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](#)

1
answer

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

I am writing a **hadoop** MapReduce job that is running over all source code files of a complete Debian mirror (\approx 40 GB). Since the Debian mirror data is on a separate machine and not in the **hadoop** ... then be reduced to a single entry. The next MapReduce job will then operate on debian packages as keys and all their files as values. However, I noticed that **hadoop** works very poorly with output values that can ...

hadoop mapreduce

asked may 25 by [Michael](#)

0
answers

Q: Installing Hadoop from .deb installer

I found an officially **hadoop** 1.2.1 .deb installer, and I'm trying to install it on Ubuntu 12.04 (I previously installed the JDK). After the installation, I found these scripts: **hadoop** **hadoop** ... -daemon.sh **hadoop**-setup-applications.sh **hadoop**-setup-hdfs.sh **hadoop**-validate-setup.sh **hadoop**-create-user.sh **hadoop**-daemons.sh **hadoop**-setup-conf.sh **hadoop**-setup-single-node.sh I tried to execute **hadoop** ...

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

3

answers

Q: How to collect Hadoop Cluster Size/Number of Cores Information

I am running my **hadoop** jobs on a cluster consisting of multiple machines whose sizes are not known (main memory, number of cores, size etc.. per machine). Without using any OS specific library (*.so ... files I mean), is there any class or tools for **hadoop** in itself or some additional libraries where I could collect information like while the **Hadoop** MR jobs are being executed: Total Number of cores ...

hadoop mapreduce cluster-computing

asked jul 17 '12 by Bob

1

answer

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

1

answer

Q: Hadoop start-dfs.sh fail

I am new to **Hadoop**, I config cluster follow the instruction After config, I start HDFS daemons with I check log file to make sure is run, but I see only text as below: Hope anyone can help? ...

hadoop ulimit

asked aug 10 by Minh Ha

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

1

answer

Q: hadoop build from source error on hadoop-mapreduce-client-jobclient

I tried to build **hadoop**-2.3.0 from source. I caught some errors but I cannot resolve the problem below.. Is there someone who can help me? Followings are my environment Since I want to show you ...

maven hadoop

asked jun 10 by Joffrey

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 So if settings are correct, Java-based **Hadoop** tasks should never get killed by **Hadoop** so you should never see the "Killing container" error like above. If you experience Java **out of** memory errors, you have to increase both memory settings. ...

answered sep 20 by user1234883

1

answer

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

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.

A: Why to code java to write Mapreduce program where as same can be achieved by hive and pig or...

It depends on the amount of data. Basically **Hadoop** (Hive + Pig) also use Java. With **Hadoop**, you can use MapReduce with hundreds of terabyte of data. If you want to do that inside your Java app ... , you've got to manage yourself to not run **out of** memory and the computation is not distributed automatically. ...

answered nov 10 by Benjamin M

A: MongoDB vs. Cassandra

/reduce implementation; Cassandra provides native **Hadoop** support, including for Hive (a SQL data warehouse built on **Hadoop** map/reduce) and Pig (a **Hadoop**-specific analysis language that many think ... over how your replication works, including support for multiple data centers. More concerned about simple setup, maintenance and code Both are trivial to set up, with reasonable **out-of-the-box** ...

answered may 24 '10 by [Michael](#)

1

answer

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](#)

0

answers

Q: Hadoop: Run MapReduce jobs on specific group of servers

Currently, we are using **Hadoop** in 3 separate projects on production, each cluster consists of different type of servers: Cluster 1: 35TB storage, medium memory and medium CPU Cluster 2: 200TB ... storage, high memory and medium CPU Cluster 3: 60TB storage, high memory, high CPU and SSD harddisks Now we want to make 1 cluster **out** of the 3 groups. So all data would be distributed among members ...

[hadoop](#) [hadoop2](#)

asked jun 15 by [user3655729](#)

5

answers

Q: running Hadoop software on office computers (when they are idle)

Is there a project which helps setup a **Hadoop** cluster on office desktops, when they are idle? I'd like to experiment with **Hadoop**/MR/hbase but don't have acces to 5-10 computers. The computers ... for this to work I need a fairly light weight monitor running on each machine. When the computer has been idle for X hours, it will join the cluster. If the user logs on, it has to drop **out** of the cluster and return all CPU/memory back. Does something like this exist? ...

[hadoop](#) [mapreduce](#) [cluster-computing](#) [hbase](#)

asked apr 14 '12 by [Shahbaz](#)

1

answer

Q: Does hadoop reducer have a timeout for inputs?

, but the job errored **out** with a message indicating as such. The version of **hadoop** I'm working with is the version you get with Microsoft HDInsight. It appears to be a **hadoop** 1.1.0 snapshot for windows ... I have a **hadoop** job that I'm trying to run on a 8-node Windows HDP cluster. The job has over 137000 inputs and it processes between 84% and 92% of the map tasks in about 3.5 hours. Then the reducer ...

[java](#) [hadoop](#) [hortonworks-data-platform](#)

asked dec 28 '12 by [Jay R.](#)

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](#)

A: cloudera hadoop mapreduce job GC overhead limit exceeded error

The Mahout jobs are very memory intensive. I don't know whether the mappers or reducers are the culprits, but, either way you will have to tell **Hadoop** to give them more RAM. "GC Overhead Limit ... Exceeded" is just a way of saying "**out** of memory" -- means the JVM gave up trying to reclaim the last 0.01% of available RAM. How you set this is indeed a little complex, because there are several ...

answered oct 18 '13 by [Sean Owen](#)

1

answer

Q: Why do I have multiple instances of Java running after installing HDnsight?

I've installed HDInsight on a desktop computer to learn to work with Hive. When I restarted my computer and logged into my profile everything was moving slow. I've figured **out** that **Hadoop** is running multiple instances of Java.exe. Is there a way to limit the amount of memory that **Hadoop** uses for this? ...

[hadoop](#) [hive](#) [hdinsight](#) [azul-zulu](#)

asked dec 3 by [Tremvelope_Filter](#)

0

answers

Q: Hadoop CompositeInputFormat block matrix-vector multiplication

I am trying to implement a block matrix-vector multiplication algorithm with **Hadoop** according to the schematics from <http://i.stanford.edu/~ullman/mmds/ch5.pdf> page 162. My matrix is going ... join in **Hadoop** concerning splits, ordering etc. and in terms of splitting I would take care of that by having one file for each block named accordingly. To me it looks like I need to modify the code ...

[java](#) [hadoop](#) [matrix-multiplication](#) [pagerank](#)

asked sep 4 '12 by [sigurd](#)