0
answers

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

---

0
answers

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

1
answer

Q: Sqoop incremental import to S3 Wrong FS error

When using the flag in the , the job will fail. Here is the full command: The exact same command without the --check-column` work without any **error** and upload to the correct s3 directory …

hadoop   cloudera   sqoop   asked feb 22 '13 by ender

1
answer

Q: CDH5.2: MR, Unable to initialize any output collector

-mapred-1.7.6-cdh5.2.0-hadoop2.jar, **hadoop**-common-2.5.0-cdh5.2.0.jar, **hadoop**-mapreduce-client-core-2.5.0-cdh5.2.0.jar I've run the following job I get the following **error**, is it a Java **heap** problem, any comments ? Thank you in advance …

java   hadoop   mapreduce   yarn   cloudera-cdh   asked nov 14 by guzu92

0
answers

Q: Error when trying to run algorithm in Spark

based on cdh4 (http://www.apache.org/dyn/closer.cgi/spark/spark-1.2.0/spark-1.2.0-bin-cdh4.tgz) For the moment however I haven't been able to successfully run it as I keep getting a java **heap error** … . The input file itself is only 136kb, so I don't think memory is the case. Below is an the command that I ran and the **error** message that I get. Command: Verbose output: Stderr from one of executor (executorID 0. Both of the executor had the same stderr output) …

apache-spark   asked jan 15 by Hafidz Zulkifli

1
answer

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

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

---

3
answers

Q: spark java.lang.OutOfMemoryError: Java heap space

my cluster: 1 master, 11 slaves, each node has 6G memory. my setting: spark.executor.memory=4g, Dspark.akka.frameSize=512 here is my problem: First,I read some data(2.19G) from hdfs to …

out-of-memory   apache-spark   asked jan 15 '14 by hequn8128

1
answer

Q: hadoop connection refused on port 9000

I want to setup a **hadoop**-cluster in pseudo-distributed mode for development. Trying to start the **hadoop** cluster fails due to refused connection on port 9000. These are my configs (pretty standard … files: **hadoop**-pschmidt-datanode-zaphod.log **hadoop**-pschmidt-namenode-zaphod.log After reformatting the hdfs following output is displayed: Using : Using returns nothing, assuming …

java   ubuntu   hadoop   hdfs   asked aug 19 '13 by dyster fyr

---

0
answers

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

Q: OutOfMemoryError while Logistic regression in SparkR

1 answer

269369 rows. The **error** message I got: Dimension of data (sample) : I have also tried hosting the same csv file over local file system as well as on HDFS. I think it will need more **Hadoop** … datanotes to store large datasets?. If it is then how can I set up Spark **Hadoop** cluster to get rid of this. (or I am doing something wrong) Tip: I think increasing Java and Spark **heap** space will help me …

r | apache-spark

asked oct 6 by Hanry

A: libssh: Memory lost in simple implementation

It happened as I suspected. This could be the bug in . I ran my program in and valgrind showed no leaks. Here is the snippet of the output: So, of course, this is not a bug in .

answered jan 8 by Rohit

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

---

0 answers

Q: NegativeArraySizeException when creating a SequenceFile with large (>1GB) BytesWritable valu…

I have tried different ways to create a large **Hadoop** SequenceFile with simply one short(<100bytes) key but one large (>1GB) value (BytesWriteable). The following sample works for out-of-box: https … ://svn.apache.org/repos/asf/**hadoop**/common/trunk/**hadoop**-mapreduce-project/**hadoop**-mapreduce-client/**hadoop**-mapreduce-client-jobclient/src/test/java/org/apache/**hadoop**/mapred/BigMapOutput.java which writes …

hadoop | out-of-memory | heap | large-files | sequencefile

asked jun 9 by user815613

---

1 answer

Q: Hadoop 2.4 cannot launch job on aws s3n

I deployed **Hadoop** 2.4 on AWS EC2 using S3 native file system as a replacement of HDFS. I tried several example apps, all gave me the following stack tracing msgs (an older thread on Jul 24 hang … (core-site.xml): https://wiki.apache.org/**hadoop**/AmazonS3 core-site.xml: I tried **Hadoop** v1 as well and turns out Hadoop1 on s3n file system works. But it seems it doesn't work for **Hadoop** v2. Please help. Thanks in advance. …

java | hadoop | amazon-s3

asked aug 12 by tddg

---

2 answers

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

---

0 answers

Q: FileNotFoundException while running Hadoop MR Job

and it works fine for the Sample WordCount program. I am new to **Hadoop**. I googled but could not find anything substantial. I am using **hadoop**-0.20.2-cdh3u6 on a single node setup. …

hadoop | mapreduce

asked aug 28 '13 by Nayan

---

2 answers

Q: What is wrong with this Java for HDInsight Hadoop?

I am trying to find out why the below Java does not work when I try to run it on **hadoop**. The **error** is: c:**hadoop**-training\tutorial02-jobtracker>**hadoop** jar PageStat.jar PageStat jobtra cker … /input/visit_5000000.txt jobtracker/output 13/07/29 11:24:50 INFO input.FileInputFormat: Total input paths to process : 1 log4j:ERROR Failed to rename [c:**Hadoop****hadoop**-1.1.0-SNAPSHOT\logs …

java | hadoop | hdinsight

asked jul 30 '13 by user2244947

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

---

2 answers

Q: Hadoop: strange NullPointer Exception running in local mode

see in the mapper tests "processRecordsFromSuspiciousFile") but it doesn't return **errors**. I also checked through logs without any success. Is it something related to wrong or missing local-mode parameters (number threads, **heap**-memory, etc)? or something wrong in the code? … I just started to study **Hadoop** and I have extracted an example from a book. So I have created a MapReducer to run locally that extracts temperatures from NCDC free data files. This is a sample …

java | hadoop

asked sep 22 '13 by Randomize

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

2
answers

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

1
answer

Q: hadoop yarn single node performance tuning

(bytes) snapshot=13738057728 Virtual memory (bytes) snapshot=66425016320 Total committed **heap** usage (bytes)=10799808512 Shuffle **Errors** BAD_ID=0 … I have **hadoop** 2.5.2 single mode installation on my Ubuntu VM, which is: 4-core, 3GHz per core; 4G memory. This VM is not for production, only for demo and learning. Then, I wrote a vey simple map …

performance   hadoop   mapreduce   yarn

asked dec 16 by Tiger.Xing

2
answers

Q: setting JAVA_HOME in linux

i set java_home: but when i have this **error** in running **hadoop**: EDIT I set on start-all.sh but not workeed EDIT2 I edit file and set java folder but not working yet …

java   linux   hadoop

asked dec 10 '13 by 2 8

0
answers

Q: Can I use the ResettableIterator interface of org.apache.commons.collections in my mapreduce…

and **heap** space **error** is provided here : "Getting java **heap** space **error** while running a mapreduce code for large dataset" … into a list and then traverse the list the second time. But since I am dealing with big data the list is becoming too long and java **heap** space problem is being reported. Then I came across the Resettable …

java   hadoop   mapreduce

asked apr 21 by Monami Sen

A: After starting hadoop, i cannot start mahout ..!

. The **error** itself says the problem: you requested a **heap** that's too large. So, I'd go see what **heap** you requested in **Hadoop** config and check it. This is the kind of info you should post in a question. … There's no OpenJDK 0.6 -- you mean 1.6. Yes Java 6 is supported, as you see in the project documentation. This does not seem to have anything to do with Mahout as it's an **error** from the JVM itself …

answered apr 13 '13 by Sean Owen

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

3
answers

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

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

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

0
answers

Q: Can't run giraph code over Cloudera CDH

I am trying to run the shortest path code in Giraph which is installed over CDH (**Hadoop** is installed on a cluster consisting of four nodes). I tried it running like: The following was the output …

hadoop   cloudera-cdh   giraph

asked sep 25 by arpp

0
answers

Q: Launching jobs in a for loop

launching 50 jobs, one by one causes difficulties to **hadoop**? EDIT: Just checked the job logs, this is the first **error** occurring: And after that we see a bunch of: EDIT2: Some ideas? The **heap** … cleared then?) EDIT3: @Donald: I haven't resized the memory for the **hadoop** daemons, so they should have a **heap** of

1GB each. The maptasks have 800 MB of **heap** from which 450 MB is used for io.sort …

asked nov 12 '13 by DDW

0
answers

Q: Mahout error on links-simple-sorted.txt

errors: **Errors** stay the same after rerun under 2GB of **heap**. If anyone has seen it before, please let me know. Thanks! … Following the "Mahout in Action" example of Chapter 6. links-simple-sorted.txt After format conversion ran the command of on a single node with **heap** of 1GB: After about 5 hours got the following …

mahout

asked feb 18 '14 by user3322621

1
answer

Q: Cause of error java.lang.IndexOutOfBoundsException: Index: 0, Size: 0

I wrote **hadoop** the program from the following Map phase As a result of execution I received the following messages I also will give Map a phase of the program which was executed successfully … directly from a database) Help to eliminate the **error**. The **error** is in a line because probably is equal to zero. But why as a result of reading from a database (the table specified in request is completed with 6 records) wasn't filled? …

java   linux   hadoop   arraylist   indexoutofboundsexception

asked apr 27 '13 by user2306966

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

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

1
answer

Q: Cloudera Manager Embedded PostgreSQL Hive Metastore Server OutOfMemoryError Issue

I looked at the logs for the hive metastore and discovered java.lang.OutOfMemoryErrors: At this point, the hive metastore gets shutdown and restarted: Now, to fix this, I've changed the max **heap** … that the actual metastore 'database' is being run as part of my cloudera manager and I'm wondering if that PostgreSQL process is running out of memory. I looked for ways to increase the java **heap** size …

hive   cloudera   hue   beeswax

asked may 7 '13 by Vijay Ratnagiri

A: After starting hadoop, i cannot start mahout ..!

It's exactly what is says in the **error** message Could not reserve enough space for object **heap** check your your **hadoop** config files: and for any properties where you have allocated memory …

answered apr 15 '13 by Ion Cojocaru

A: Hadoop on Amazon Cloud

Have you tried Amazon Elastic MapReduce? This is a simple API that brings up **Hadoop** clusters of a specified size on demand. That's easier then to create own cluster manually. But once the jobflow … permanently (e.g. if you are running HBase on top of **Hadoop**) you may actually need own cluster on EC2. In this case you may find Cloudera's distribution of **Hadoop** for Amazon EC2 useful. Altering **Hadoop** …

answered nov 8 '11 by Alexey Tigarev

1
answer

Q: How to work with Mahout?

I have tried to work with mahout using the link: http://girlincomputerscience.blogspot.in/2010/11/apache-mahout.html When i execute the command it shows: What may be the possible reasons for the **errors**? …

apache   hadoop   cloud   mahout   recommendation-engine

asked jul 17 by user2799441

1
answer

Q: Unable to start Hbase Thrift server

the terminal was still hanged. The page did not open but the terminal resumed showing the **error** message # # java.lang.OutOfMemoryError: Java **heap** space # -XX:OnOutOfMemoryError="kill -9 %p" # Executing /bin/sh -c "kill -9 18526"... Killed Please suggest me what to do. Regards … I have install **Hadoop** 1.2.1 and Hbase 0.94.8 in psudo-distributed mode in my ubuntu system running Ubuntu 12.04. Everything is working fine and all the deamons are running. In order to access hbase …

php   ubuntu   hadoop   hbase   thrift

asked feb 19 '14 by xavii

1
answer

Q: Display the imported table in HDFS using sqoop

i have set up a single node **hadoop** cluster and configured it to work with apache Hive, now when i imported a mysql table using the following command (using sqoop) sqoop import --connect … run the import command again it gives me the **error** that dir already exists please help me find the solution EDIT: i havent created any tables prior to the import, do i have to create a table …

hadoop   fedora   hive   sqoop

asked aug 4 '12 by John x

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

1
answer

Q: Determination of execution of mapper

on which mapper is executed. Here I give a code But despite it on one of the subordinate nodes the /export/**hadoop**-1.0.1/bin/readwaysfromdatabase.txt file wasn't created. Whether follows from here, what …

java   linux   hadoop   mapreduce   mapper

asked apr 22 '13 by user2306966

A: how to test Hadoop mapreduce

into CPU and **heap** dumps of **Hadoop** job. It also provides an in-depth graphical view of the MapReduce phases – Map, Reduce, Sort, Shuffle, Setup, and Cleanup. HDFS data validation Provides a simple … and easy-to-use utility to find discrepancies and **errors** in HDFS data. User may check for data violations under various categories like data type, null values, or regular expression values. Cluster …

answered aug 24 by Mash

1
answer

Q: Sqoop Oracle Hive Error

I am having a problem with the sqoop finding my table's upper bound value. It has to be 203 but its coming null as a result of which the query (in bolded) is wrong.

oracle   hadoop   hive   sqoop

asked apr 15 '13 by Arjun Harish

2
answers

Q: RHadoop reduce job failed

I am following RHadoop tutorial, https://github.com/RevolutionAnalytics/rmr2/blob/master/docs/tutorial.md and running the second example, but I am getting **errors** which I can't resolve. The code … is as the following: The map job is successful, but reduce job failed, part of the **error** message is as the following: Job failed as tasks failed. failedMaps:0 failedReduces:1 could somebody help? I couldn't proceed further from here. Thanks. …

hadoop   rhadoop

asked jul 24 by user3714023

1
answer

Q: Reason and java.lang.NullPointerException bug fixing

I wrote the program from the following phase And as a result of execution on a cluster I received the following messages The **error** contains in a line In what cause of **error**? What it generated (owing to what it arose)? Help to remove it. …

java   linux   unix   hadoop   nullpointerexception

asked may 1 '13 by user2328488

1
answer

Q: Apache Spark - MLlib - K-Means Input format

else. I use Spark 1.0.0 and my input textile (400MB) looks like this: I am not sure if I can use the first two, because in the MLlib example file there only use floats. So I also tried the last two: My **error** code in both cases is here: What am I missing? …

scala   k-means   apache-spark   mllib

asked aug 7 by user3400996

1
answer

Q: Memory issues when running Spark job on relatively large input

following **errors** in the logs: and also: How do I go about debugging such an issue? EDIT: I Found the root cause of the problem. It is this piece of code: Particularly allocating a buffer of size …

hadoop   apache-spark   google-hadoop

asked oct 22 by Yaniv Donenfeld

---

15   30   **50**   per page