



SUPPORT QUESTIONS

Find answers, ask questions, and share your expertise



Advanced Search

[Home](#) > [Support](#)

[Support](#)

[Questions](#)

> [Difference between mr and Tez?](#)

Options

[All Forum Topics](#) [Previous](#) [Next](#)

Difference between mr and Tez? ✓ Solved

[Go to solution](#)

Labels:

[Apache Hadoop](#)[Apache Tez](#)

balavignesh_nag

Guru

Created 02-14-2017 09:25 AM



What are the differences between mr mode and tez.

Why tez functions faster than mr?

why few queries fail in tez but gets executed successfully in mr mode?

I know the question I ask have multiple answers but I just wanted to know all the possible scenarios. Thanks

Reply



36,110 VIEWS



1 KUDO

0 AGREE

TAGS (6)

[hadoop](#) [Hadoop Core](#) [mr-history](#) [MR2](#) [tez](#) [Upgrade to HDP 2.5.3 : ConcurrentModificationException When Executing Insert Overwrite : Hive](#)

1 ACCEPTED SOLUTION

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Accept Cookies

Re: Difference between mr and Tez?



mqureshi Super Guru

Created 02-14-2017 03:41 PM



@Bala Vignesh N V

Tez is a DAG (Directed acyclic graph) architecture. A typical Map reduce job has following steps:

1. Read data from file --> one disk access
2. Run mappers
3. Write map output --> second disk access
4. Run shuffle and sort --> read map output, third disk access
5. write shuffle and sort --> write sorted data for reducers --> fourth disk access
6. Run reducers which reads sorted data --> fifth disk output
7. Write reducers output --> sixth disk access

Tez works very similar to Spark (Tez was created by Hortonworks well before Spark):

1. Execute the plan but no need to read data from disk.
2. Once ready to do some calculations (similar to actions in spark), get the data from disk and perform all steps and produce output.

Only one read and one write.

Notice the efficiency introduced by not going to disk multiple times. Intermediate results are stored in memory (not written to disks). On top of that there is vectorization (process batch of rows instead of one row at a time). All this adds to efficiencies in query time.

Now to answer your question on why Tez queries fail but executed in MR. This should not happen. Possible bugs or sometimes people working with Hive have used MapReduce for a while and know how to make things work but not as familiar with Tez. I think, Tez queries should not fail any more than Map Reduce.

I highly recommend skimming quickly over following slides, specially starting from slide 7.

http://www.slideshare.net/Hadoop_Summit/w-235phall1pandey

View solution in original post

Reply



10,973 VIEWS



8 KUDOS

0 AGREE

7 REPLIES

Cloudera uses cookies to provide and improve our site's services.

Re: Difference between mr and Tez?

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.



dkozlowski Guru

Accept Cookies

Created 02-14-2017 02:57 PM



@Bala Vignesh N V

Please, have a look at the below links. Hopefully, there will give you some more details:

- <http://hortonworks.com/blog/evaluating-hive-with-tez-as-a-fast-query-engine/>
- <http://hortonworks.com/apache/mapreduce/>
- <https://community.hortonworks.com/questions/65772/when-tezmr-is-better-for-query-execution.html>

Reply

 10,972 VIEWS  0 KUDOS

Re: Difference between mr and Tez?



snukavarapu  Cloudera Employee

Created 02-14-2017 03:31 PM



Hi @Bala Vignesh N V

Here are some links that you can go through to understand differences between MR and Tez.

Difference between MR and Tez & Why Tez is faster?

http://hortonworks.com/apache/tez/#section_2

<https://cwiki.apache.org/confluence/display/Hive/Hive+on+Tez>

http://www.slideshare.net/Hadoop_Summit/w-235phall1pandey (See slides 9 to 16)

why job fails in tez but runs in MR?

There could be several reasons for the jobs failure. It could be because of OOM exceptions when memory is not tuned correctly (refer:

[https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-](https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.4/bk_installing_manually_book/content/determi....)

[2.3.4/bk_installing_manually_book/content/determi....](https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.4/bk_installing_manually_book/content/determi....) I have also seen cases where job fails in MR but runs fine when Tez execution engine is used.

Reply

 10,973 VIEWS  0 KUDOS

Re: Difference between mr and Tez?

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Created 02-14-2017 03:41 PM



Accept Cookies

@Bala Vignesh N V

Tez is a DAG (Directed acyclic graph) architecture. A typical Map reduce job has following steps:

1. Read data from file --> one disk access
2. Run mappers
3. Write map output --> second disk access
4. Run shuffle and sort --> read map output, third disk access
5. write shuffle and sort --> write sorted data for reducers --> fourth disk access
6. Run reducers which reads sorted data --> fifth disk output
7. Write reducers output --> sixth disk access

Tez works very similar to Spark (Tez was created by Hortonworks well before Spark):

1. Execute the plan but no need to read data from disk.
2. Once ready to do some calculations (similar to actions in spark), get the data from disk and perform all steps and produce output.

Only one read and one write.

Notice the efficiency introduced by not going to disk multiple times. Intermediate results are stored in memory (not written to disks). On top of that there is vectorization (process batch of rows instead of one row at a time). All this adds to efficiencies in query time.

Now to answer your question on why Tez queries fail but executed in MR. This should not happen. Possible bugs or sometimes people working with Hive have used MapReduce for a while and know how to make things work but not as familiar with Tez. I think, Tez queries should not fail any more than Map Reduce.

I highly recommend skimming quickly over following slides, specially starting from slide 7.

http://www.slideshare.net/Hadoop_Summit/w-235phall1pandey

View solution in original post

Reply



10,974 VIEWS



8 KUDOS

0 AGREE

Re: Difference between mr and Tez?



balavignesh_nag Guru

Created 02-15-2017 08:18 AM



spot on @mqureshi. Thanks it helps !!

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Reply



10,973 VIEWS



0 KUDOS

Accept Cookies

Re: Difference between mr and Tez?



mqureshi Super Guru

Created 02-16-2017 02:55 PM



@Bala Vignesh N V

Glad, it was helpful. If you think its a complete answer you were looking for, please accept the answer.

Reply

 10,973 VIEWS  0 KUDOS

Re: Difference between mr and Tez?



BharatTiwari9 New Contributor

Created 07-28-2020 11:31 PM



"I highly recommend skimming quickly over following slides, specially starting from slide 7.
http://www.slideshare.net/Hadoop_Summit/w-235phall1pandey"

This slide is not there at the path

Reply

 3,937 VIEWS  0 KUDOS



Hive + Tez: A Performance Deep Dive

slideshare.net/Hadoop_Summit/w-235phall1pandey

Published on 1. Hive+Tez: A Performance deep dive Jitendra Pandey Gopal Vijayaraghavan 2. © Hortonworks Inc. 2014. Stinger Project (announced February 2013) Batch AND Interactive SQL-IN-Hadoop Stinger Initiative A broad, community-based effort to drive the next generation of HIVE Hive 0.13, April

Re: Difference between mr and Tez?



Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.



Accept Cookies

If the table is partitioned and there are delta files (from updates, for eg.), I think mr works but not tez. You may have to run compaction to convert the delta files into base files and then tez will work.

Reply

 10,973 VIEWS  0 KUDOS

Explore the Community



Already a User?

Don't have an account?

Sign In

Register

ANNOUNCEMENTS

Product Announcements

[ANNOUNCE] CDP Public Cloud Data Hub is now available on Google Cloud

Support Announcements

Validations - Cloudera Support's Predictive Alerting

Support Announcements

Cloudera's CDP Upgrade Advisor

Product Announcements

[ANNOUNCE] CDS 3.1 Powered by Apache Spark 3.1.1 is now GA!

Product Announcements

[ANNOUNCE] CDP Private Cloud Base 7.1.6 Released















[View More Announcements](#)

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in [Cloudera's Privacy and Data Policies](#).

Accept Cookies

RECOMMENDATIONS

-  ✓ Hi..whats the difference between "Start Tez sessio...
-  Hive query (tez/mr) hangs due to empty string oper... 
-  ✓ Difference between installing from Apache vs Horto... 
-  Hadoop Query Issue - Different Result on TEZ and M...
-  Jumbune MR and Cluster Analyzer 
-  ✓ What is the difference between Lineage and Impact ... 
-  ✓ What is the difference between HDP Repo and Apache...
-  ✓ Difference between Kafka 0.9.0.2(hortonworks) and ...
-  Image differences between Vmware and Virtualbox qu...
-  ✓ HBase Replication between two clusters for differe...

My Cloudera Home

Community

Documentation

Knowledge Base

Partners

Support

PRODUCTS

Cloudera Data Platform

Cloudera uses cookies to provide and improve our site's services.

Data Hub

By using this site, you consent to use of cookies as outlined in Cloudera's Privacy and Data Policies.

Data Warehouse

Machine Learning

Accept Cookies

CDP Private Cloud

Data Flow

Data Science Workbench

Enterprise Data Hub

SOLUTIONS

Customer Analytics

IoT / Connected Products

Security, Risk, & Compliance

Industry Solutions

Solutions Gallery

DOWNLOADS

Cloudera Data Flow

CDP Private Cloud

Cloudera Director

MORE

About

News & Blogs

Investor Relations

Events

Customers

Careers

Security

Community Guidelines

Community FAQs

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in [Cloudera's Privacy and Data Policies](#).

[Contact Us](#)

[Accept Cookies](#)

United States: +1 888 789 1488

Outside the US: +1 650 362 0488

© 2021 Cloudera, Inc. All rights reserved.

[Terms & Conditions](#) | [Privacy Policy and Data Policy](#) | [Unsubscribe / Do Not Sell My Personal Information](#) |

[Supported Browsers Policy](#)

Apache Hadoop and associated open source project names are trademarks of the Apache Software Foundation. For a complete list of trademarks, [click here](#).

Cloudera uses cookies to provide and improve our site's services.

By using this site, you consent to use of cookies as outlined in [Cloudera's Privacy and Data Policies](#).

Accept Cookies