**IDRP-SCM**

**Optimization Details**

SCPO-DAY

**SCPO-DAY->**

**DP-00-IE-AUTH-KMT-SDAY-> DP-00-IE-DETER\_AUTHO\_KMART\_MSTR**

Date:- 15/12/2018

Developed By :- Heena Salim Shaikh

**JOB DETAILS**

* I have worked on the following scripts:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr No** | **Job name** | **Average Run Time** | **Status** |
| 1 | DP-00-IE-MASTERDATA\_GOLDDEP | 0:48:28 | Did 5 minutes performance improvement but no direct impact of this script on execution time as it is not in critical path hence I have not put the changes in the production |
| 2 | DP-00-IE-DETER\_AUTHO\_KMART\_MSTR | 1:28:17 | 10 minutes performance improvement |

**DP-00-IE-DETER\_AUTHO\_KMART\_MSTR**

This job have following four scripts those are in critical path :-

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Scripts Names** | **Execution Time** |
| 1 | perform\_item\_eligibility\_smith\_\_idrp\_space\_planning\_authorized\_vendor\_package\_stores.sh | 37 |
| 2 | perform\_item\_eligibility\_work\_\_idrp\_vp\_cancarry\_stores.sh | 10 |
| 3 | perform\_item\_eligibility\_work\_\_idrp\_post\_drop\_ship\_proc.sh | 13 |
| 4 | perform\_item\_eligibility\_work\_\_idrp\_vpstores\_after\_order\_dotcom\_edits.sh | 7 |

perform\_item\_eligibility\_smith\_\_idrp\_space\_planning\_authorized\_vendor\_package\_stores.sh takes 37 minutes to execute hence I have worked on this long running script .

**Shell Name** perform\_item\_eligibility\_smith\_\_idrp\_space\_planning\_authorized\_vendor\_package\_stores.sh

**Pig Name** /appl/hdidrp/pig/scripts/item\_eligibility/smith\_\_idrp\_space\_planning\_authorized\_vendor\_package\_stores/perform\_item\_eligibility\_smith\_\_idrp\_space\_planning\_authorized\_vendor\_package\_stores.pig

**Input Files**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Input Files** | **File Count** |
| 1 | /smith/space\_planning/archive/prev\_\*/SG.GDTU100.DTCITEM | 1217173842 |
| 2 | /smith/idrp/eligible\_loc | 165323 |
| 3 | /smith/idrp/shc\_item\_combined | 8239880 |
| 4 | /smith/idrp/vend\_pack\_combined | 11146269 |
| 5 | /work/idrp/items\_vend\_packs\_can\_carry | 1987953 |
| 6 | /smith/space\_planning/current/sbt\_ord\_pt | 462 |
| 7 | /smith/space\_planning/current/sbt\_vend\_packs | 223371 |
| 8 | /smith/space\_planning/current/sbt\_launch\_str | 300729 |
| 9 | /gold/geographic/model\_store | 36722353 |
| 10 | /smith/idrp/ksn\_attribute/current | 3055858 |

**Output Files**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Output Files** | **File Count** |
| 1 | /smith/idrp/space\_planning\_vendor\_package\_stores/current | 466242290 |

**Testing efforts**

**I have carried out the 150 test cases and 20 test cases have given the positive result, but when I combined the all those changes ,I got 6 following test cases those are giving the improvement.**

**Please find the following for the same.**

**TECHNICAL CODE CHANGES**

**1: Adding the compression technique to the files**

I have added the following new code in the script :-

New Code:-

**SET io.compression.codec.lzo.class com.hadoop.compression.lzo.LzoCodec**

**SET pig.tmpfilecompression true**

**SET pig.tmpfilecompression.codec lzo**

Details:-

The pig stores the temporary output of the aggregate functions in the form of the files. By using this compression technique we can compress the temporary output files.

Advantages of the LZO compression:-

Saving the hard disk space

Speedy transfer or copy of the data

**2: Setting Mappers in the script**

I have added the following new code in the script :-

New Code:-

**set mapred.min.split.size 524288;**

Details:-

More number of the mapper more speed.

**3: Changed join rev\_dtc\_file11**

I have modified the older join rev\_dtc\_file11 in following way:-

Older Code :-

**rev\_dtc\_file11 = JOIN rev\_dtc\_file BY REV\_ITEM, hier\_item\_data\_required BY H\_ITEM\_ID using 'skewed';**

Newer Code:-

**rev\_dtc\_file11 = JOIN hier\_item\_data\_required BY H\_ITEM\_ID, rev\_dtc\_file BY REV\_ITEM;**

Details:-

I have compared the data of the two files, and kept larger file last. Generally, we have to adjust the files according to speed.

Also, I have removed the skewed keyword as data was not skewed.

**4: Changed the input\_join**

I have modified the older join rev\_dtc\_file11 in following way:-

Older Code:-

**input\_join = JOIN input\_req\_gen BY ITEM\_ID, sbt\_ord\_join1 BY sbt\_vend\_join1::filtered\_vend\_packs::shc\_item\_id;**

Newer Code:-

**input\_join = JOIN input\_req\_gen BY ITEM\_ID, sbt\_ord\_join1 BY sbt\_vend\_join1::filtered\_vend\_packs::shc\_item\_id using 'replicated';**

Details:-

I have added the replicated keyword to the join.

When we have one table small and one table large then a small table can be fit into the memory hence we can access the data faster.

Replicated join increases the speed and performance of the script.

**5: Adding the filtered\_vend\_packs after loading the corresponding file only**

I have added the filter immediately after loading the file , earlier it was filtering the data at the last.

Moved filter after:-

**smith\_\_idrp\_vend\_pack\_combined\_data =**

**LOAD '$SMITH\_\_IDRP\_VEND\_PACK\_COMBINED\_LOCATION'**

**USING PigStorage('$FIELD\_DELIMITER\_CONTROL\_A')**

**AS ($SMITH\_\_IDRP\_VEND\_PACK\_COMBINED\_SCHEMA);**

**filtered\_vend\_packs = FILTER smith\_\_idrp\_vend\_pack\_combined\_data by (TRIM(ksn\_purchase\_status\_cd) != 'U') ;**

Details:-

If we use the filter after loading the file , it increases the speed and performance of the job.

**6: Adding parallel to test\_data\_1234**

I have added the PARALLEL keyword here to increase the number of the reducers.

Newer Code:-

**test\_data\_1234 = UNION t\_data\_1, t\_data\_2, t\_data\_3 PARALLEL 500;**

Details:-

Reducers do join, union , aggregation and combination operations .More number of the reducers 🡪 More number of data will process 🡪 Speedy Operations

**final\_test1.sh without changes final DONE**

**the\_final\_critical\_script2.pig final with brackets**

**the\_final\_critical\_script3.pig final with setting brackets DONE**

**the\_final\_critical\_script4.pig combined with setting bractes and parallel together**

**the\_original.pig**

**THANK YOU ☺!!**