

Procedure for the migration of the DC06 data

The procedure used to migrate the data is described in this document (under development..)

Status of the migration started on Tue Apr 15 2008

On Wed Apr 16, 10 AM:

Total number of productions migrated so far: 165
Number of productions left: 349
Average rate = 25.68 rows/s
Weighted Average rate = 64.06 rows/s

On Wed Apr 16, 15:15:

Total number of productions migrated so far: 195
Number of productions left: 319
Average rate = 28.46 rows/s
Weighted Average rate = 64.56 rows/s

Wed Apr 16 15:15: The migration script crashed due to the expired AFS tokens (not possible to write the information on the logfile!).

Migration script restarted on Apr 16 16:26.

Migration crashed at 3 AM on Apr 17 (AMGA could not deal with a very big query).
Restarted at Apr 17 11:24 AM.

On Thu Apr 17 11:33:23 2008

Migration status:
Total number of productions migrated so far: 325
Number of productions left: 189
Average rate = 25.35 rows/s
Weighted Average rate = 64.67 rows/s

Migration script crashed at 18:39 while writing to logfile (formatting error, my fault!). Restarted the day after (Friday).

On Fri Apr 18 10:15:55 2008

Total number of productions migrated so far: 408
Number of productions left: 106
Average rate = 23.24 rows/s
Weighted Average rate = 62.51 rows/s

Friday 18 Apr 2008

migration finished for most of the productions (only 19 productions left to migrate, they need a special procedure due to the very big number of jobs).

After migration checks:

some productions give the following error: in the inputFiles table some FileId correspond to more than one

JobId. This means that the file has been used as input for more than one job. This problem can be due to the fact that we are copying data from the integration DB which is an exact copy of the production DB created on Feb19. Some productions were found to have this problem with duplicated input files, and were reprocessed afterward. Then, Gianluca provided a list of 34 productions which have undergone modifications since 19Feb, they have been removed from the new schema and copied again from the production DB.

Monday 21 Apr 2008

Total number of productions migrated so far: 495
Number of productions left: 19
Average rate = 23.14 rows/s
Weighted Average rate = 61.88 rows/s

but again, some productions (20) show the problem about the same file used in input for more than 1 job. The list of productions is: 00001536 00002000 00002001 00002030 00002034 00002036 00002037 00002038 00002039 00002051 00002054 00002058 00002059 00002061 00002069 00002072 00002077 00002078 00002079 00002082 00009280.
Need of further investigation...

Tuesday 29 Apr 2008

It is confirmed that these productions have a problem. Some jobs and files of these production have to be skipped during the migration, because they are jobs produced processing an input file which had already been processed by another job. And of course also the output file of such jobs have to be excluded from the migration. I am currently implementing a fix in the migration suite in order to take account of these problem.

Wednesday 30 Apr 2008 - New Migration Tool for Very Big Productions

Migration of the 19 very big productions (very big meaning with tables with more than 55000 rows): a new tool has been developed, totally based on Oracle, since the tool based on the AMGA api could not be used, as it crashes for big queries.

The oracle migration script has been started on Wed Apr 30 10:10:27 2008

Monday 5 May 2008

The migration script failed as the new schema account ran out of space!
Restarted on Monday. Still running...

Migration status at Wed May 7 15:42:01 2008
Total number of productions migrated so far: 502
Number of productions left: 12
Average rate = 22.74 rows/s
Weighted Average rate = 62.41 rows/s

Monday 2 June 2008 - Migration of files with null production number

See below for more details about these files.

The migration has started on Monday 2nd June at 5 pm. and finished on Wed 4th June at 8:40 am. Migration tool: migrateNullProdFiles.test4() in the directory /home/elisal/bookkeeping/migration/bk/. Logfiles stored in : /home/elisal/bookkeeping/migration/bk/nullProd

Performance of the migration:

So far: files read: 124 and Jobs inserted = 3212302 (and 3952050 files)
average rate 67.0 rows/s -> estimated total time: 16.4 h
weighted average rate 27.0 rows/s -> estimated total time: 40.6 h

Wed 4 June 2008: Productions with multiple input files

Finally it has been decided to migrate these production as they are in the old schema, and to fix the problem a posteriori. The productions are:

['00001536', '00002000', '00002001', '00002030', '00002034', '00002036', '00002037', '00002038', '00002039', '00002051', '00002058', '00002059', '00002061', '00002069', '00002072', '00002077', '00002078']

migration finished at 3:30 pm.

Problems found during the migration

Production 1734

there are 2 jobs which do not have any entry in the files table. Action: these jobs have been skipped (they are useless).

Production 1762

There are 30 jobs with no entry in the files table (see prod 1734).

There are 4 files reported in the *files* table, but they do not have any entry in the *fileParams* table. These are the fileIDs and type: 31030782 (GAUSS output file), 31030794 (GAUSS histogram file), 31030799 (GAUSS output file), 31030807 (Brunel logfile).

None of these files is reported in the *inputFiles* table.

I think that these files can be safely skipped in the migration.

Production 1894

The job with JobId = 11833483 exists in the *jobParams* table, but it doesn't in the *jobs* table. This job has some output files (fileId = 33328384,33328385,33328386) hence it cannot be dropped. Then, the attributes of the *jobs* table have been retrieved from other events of the same production and the entry has been added by hand in the new schema.

In the old schema nothing has been altered.

Files with null production number

In the old schema some files have a null production number. For this reason, they have not been copied to the new schema (the migration is done production by production). The problem is that these files appear in the new schema in the *inputFiles* table, because they have been taken in input by jobs that have been migrated to the new schema. Then, these fileIDs appear in the *inputFiles* table, but they don't in the *files* table, and this is an inconsistency!

This problem has to be solved...

We fixed the problem in this way: select all the files with this problem (null production number, configuration DC06 and gotReplica='yes'). On the basis of the file name, we have reconstructed the job name and the production number. The information has been stored, production by production, on ascii files (one for each productions). After that, an ad hoc migration script has been developed to copy these files from the old schema to the new schema. Migration done on Tuesday 3 June 08.

-- ElisaLanciotti - 16 Apr 2008

This topic: LHCb > MigrationPage

Topic revision: r17 - 2008-06-04 - ElisaLanciotti



Copyright &© 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
Ideas, requests, problems regarding TWiki? Send feedback