CXT – AWS Data Seeding prod checklist

Date : Sat Feb 1, 2020 08:00

DBA : Oscar Pagan, John Semencar

Mgr : Bob Gleason

Db Server : xcxtdbm2p

Db Name : CXTHEPRD

Project # : P30942A

\*\*\* ~~Ship oldest claims first (ie; Year1 = oldest claims, Year3 = most current claims)~~

~~\*\*\* Date for Year1 (oldest claims) : ???~~

\*\*\* the week before Feb 1, we were told not to divide the data into 3 separate years.

\*\*\* we will perform the extract in one 3 year collection.

Total estimate time for all 3 years = 36 to 45 hours.

Notes : Before the PreWork :

* perform the encryption setup
* perform the AWS client, network, S3 bucket setup

The PreWork is one-time only ( apx 45 mins ).

There are 4 tasks ~~for each of the 3 years~~ :

* + the first task completes ~~in apx 7.5 hrs.~~ ( x3 = apx 21 hrs)
  + the remaining tasks complete ~~in apx 2.5 hrs each.~~ (x3 = apx 6 hrs each)

1. proc - edit the date range for ~~each~~ 3years and run the proc
2. exp - ~~replace export directory for each year and~~ run the export
3. encrypt - create the “filelist” and encrypt the exp files
4. ship - transfer the encrypted exp files to AWS S3 bucket

Scripts : /orahome/u01/aetna/scripts/stage/DBA/CHC\_Cloud/PROD

\*\*\* before beginning Prework, the prod db must be cloned and exclude file must exist.

Prework :

* + upd\_tab\_BATCH\_primary.sql
  + disable archive log mode !!
  + add\_undo.sql
  + c~~r\_tabs\_indexes-2.sql~~ cr\_driver\_table.sql
  + cr\_dpdba.sql
  + mkdir\_oraexport\_u01.ksh
  + c~~r\_dp\_dir\_y1.sql~~ cr\_dp\_dir\_3years.sql
  + c~~r\_dp\_dir\_y2.sql~~
  + c~~r\_dp\_dir\_y3.sql~~
  + load\_exclude.ksh
  + cr\_exclude\_index.sql
  + gather\_exclude\_index\_stats.sql
  + disable\_auto-stats.sql

The 4 Tasks :

1. c~~r\_proc\_y1.sql~~

c~~r\_proc\_y2.sql~~

~~cr\_proc\_y3.sql~~

~~run\_proc\_y1.ksh~~

~~run\_proc\_y2.ksh~~

~~run\_proc\_y3.ksh~~

cr\_proc\_3years.sql

run\_proc\_3years.ksh

\*\*\* sql used for date range :

WHERE CV.IC\_PURGE\_DATE >= TO\_TIMESTAMP('02/01/2017','mm/dd/yyyy')

AND CV.IC\_PURGE\_DATE < TO\_TIMESTAMP('02/02/2020','mm/dd/yyyy')

1. e~~xp\_dataSeed\_y1.ksh~~

~~exp\_dataSeed\_y2.ksh~~

~~exp\_dataSeed\_y3.ksh~~

cr\_driver\_indexes.sql

gather\_CCES\_CLAIM\_VERSION\_stats.sql

exp\_dataSeed\_3years.ksh

1. e~~ncrypt\_files\_y1.ksh~~

~~encrypt\_files\_y2.ksh~~

~~encrypt\_files\_y3.ksh~~

~~encrypt\_files\_3years.ksh~~

encryption was moving slowly, so we parallelized it by :

* created separate scripts, each with its own filelist
* manually created separate filelists with a range of files

x1\_encrypt\_files\_3years.ksh

x2\_encrypt\_files\_3years.ksh

x3\_encrypt\_files\_3years.ksh

x4\_encrypt\_files\_3years.ksh

x5\_encrypt\_files\_3years.ksh

1. s~~hip\_AWS\_y1.ksh~~

~~ship\_AWS\_y2.ksh~~

~~ship\_AWS\_y3.ksh~~

ship\_AWS\_3years.ksh

Post impl notes :

clone db began at 08:00 on impl day. Done by 11:00.

during that time the exclude file was ftp’d.

preWork began at noon.

The table size for the proc was larger than expected. Estimated 3x7gb = 21gb. Actual = 30gb.

More data means larger exp size. Estimated = 750gb. Actual = 1.1 tb.

The exclude file had apx the same mount of rows as last summer’s testing :

* Aug 2019 = 31,665,502 ods\_full\_exclude\_file.txt
* Feb 2020 = 31,601,445 ODS\_EXCLUDE\_D020120.txt

Export and encryption sizes : xcxtdbm2p:/oraexport/u01/CXTHEPRD> du -ms \*

* 1036281 3years (1.08 tb)
* 252959 gpg\_files ( 259 gb)

!!! We were able to speed up almost every task using a few different methods:

* bumped up parallelism from 8 to 16 where applicable.
* created index on exclude table.
* ran stats on exclude table and its indexes.
* did not create driver table’s () indexes until after the proc completed.
* ran stats on the driver’s indexes before the export.

Total impl time from 08:00 Sat until 18:00 Sun.

The end . . .