K10plus upgrading to CBS9/Postgres

Nils Achtergarde
Migration to CBS 9.0

Migration will be done during 09/21/2023 - 09/24/2023

• Extraction from Sybase ASE and conversion to PostgreSQL takes some amount of time
• Sybase CBS will be set to readonly and will be available for some time
• PostgreSQL 15 will be used.
Preparations for the migration

• Some cleanup/error corrections had to be done before (e.g. erasing some non-printable characters in titles)

• Old GBV-only CBS was erased last week

• A separate CBS-instance was build up (already running with CBS 9.0) for supplying articles to discovery systems

• db_s_title_log was cleaned up, which freed up 150GB in the database contained in 1.3 billion lines

• CBSlib-programs had to be adapted, recompiled and tested

• ILL-scripts still have to be adapted
Advantages of PostgreSQL

• Opensource

• Widely used, even in environments with enormous requirements regarding performance, e.g. german stock-exchange

• No license costs

• Standard pattern matching perl-like and e.g. for date-conversion

• Native support in multiple programming languages

• Many Extensions e.g. for handling geolocations
Compression in PostgreSQL

- Compression of the title-tables was adapted
- Based on zstd
- Different levels of compression can be defined, relation between cpu-usage and saving of space have to be considered
- With level 2 we have saved half the space in the database
- Reduces network-traffic for database-replication
- Only adapted in K10plus-CBS for now
Partitioning of tables – Part 1

• Partitions are defined per table based on a column, e.g. for the table dbs_titles_global the column fno. This leads to separate tables for each fileset, e.g. dbs_titles_global_21, dbs_titles_global_12, ....

• With partitions deletions can be done in “notime”

• Speedup of selections
Partitioning of tables – Part2

• Already was possible with Sybase, but with high license-costs
• Commands still can be used transparently, e.g. update, select, insert
• Only adapted in K10plus-CBS for now
• Partitions on log-tables based on year/month are considered
Entries in dbs_title_log

Datenreihe 3
Syncronization

• Synchronization is mostly done by asynchronous database-replication
• Push-/pull-mechanism is mostly replaced
• The sync has to be monitored
• Was available with Sybase, but cost-intensive and not easy to implement
• Copying of filesets with bcpttl works again for PostgreSQL, but only for K10plus. Even works for copying filesets from Sybase to PostgreSQL
Replacement of Bibserv

- Bibserv is replaced by CJM in CBS 9.0
- CJM works fine and looks fine for CSDBC-Exports
- Was intensely tested by BSZ/OCLC, since really necessarily needed by BSZ for updates to local library systems
- “Bibserv”-like server will be still be used, replicated each weekend
- Jenkins is used for general jobs to be run
Database-backup

• Database-backups were done with Logdumps every 10 minutes
• In PostgreSQL database-backups will be done based on the WAL-files, also every 10 minutes
• Backups will be run on a mirror-system
Structure of CBS-environment

BSZ

VZG

Nils Achtergarde | CBS-Partner-Meeting | Hamburg, 5th September 2023
Last but not least

- The table title_overflow no more is use, HOORAY!!!
- Maintenance for PostgreSQL has to be implemented
- Template-system really helps to build up new instances and keep these consistent