File system usage during v43 upgrade

Summary

Because of the implications for storage requirements for the v43 upgrade, the upgrade will not proceed unless the filesystem containing the database has sufficient space available. The expected amount required during the upgrade is 1.5 times the existing database size, plus 10GB margin, is available.

Background

The V43 upgrade procedure, related to the database, is as follows:

1. Make an incremental backup of the PostgreSQL 9.6 database with an associated restore point.
2. Migrate the v42 database from PostgreSQL 9.6 to 12.
3. Delete the PostgreSQL 9.6 database.
4. Make a full backup of the PostgreSQL 12 database with an associated restore point.

Resilient uses a PostgreSQL database. As part of the upgrade to Resilient v43, the PostgreSQL database is upgraded from v9.6 to v12.8. This necessitates a data migration procedure which has an effect on the storage requirements for Resilient.

One of the first steps in the Resilient upgrade procedure is to backup the database so that it would be possible to roll the upgrade back to the the point before the upgrade was performed. The quicker the backup step, the quicker the upgrade can be. In v42 we chose to use `pgbackrest` to facilitate making a quicker backup.

`pgbackrest` supports making either a full backup or incremental backups. A full backup is a compressed image of all the data in the database. Incremental backups only contain the data that was changed since earlier backups, and are quicker to perform. Since v42, `pgbackrest` has been performing a full backup at 00:01am every Sunday, and an incremental backup at the same time every other day. Consequently, in making a backup as part of the upgrade process, only an incremental backup is necessary. Using `pgbackrest` imposes an additional storage cost. To mitigate this cost, only one week's worth of backups is retained.

PostgreSQL and pgbackrest storage locations

The Resilient PostgreSQL DB is stored under /crypt/database/<PostgreSQL version>

The pgbackrest backups are stored in /crypt/pgbackrest_repo/backup/<pgbackrest stanza> which is by default `/crypt/pgbackrest_repo/backup/ibm-security-soar/`

Details about the pgbackrest backups can be retrieved using

```
sudo -u postgres pgbackrest --stanza=<stanze name> info
```

e.g.

```
 $ sudo -u postgres pgbackrest --stanza=ibm-security-soar info

stanza: ibm-security-soar
    status: ok
    cipher: none

    db (current)
```

```
wal archive min/max (9.6): 000000010000000200000073/000000010000000200000078

full backup: 20210928-161354F
    timestamp start/stop: 2021-09-28 16:13:54 / 2021-09-28 16:14:45
    wal start/stop: 000000010000000200000073 / 000000010000000200000073
    database size: 17.5GB, database backup size: 17.5GB
    repo1: backup set size: 2.6GB, backup size: 2.6GB

incr backup: 20210928-161354F_20210928-173012I
    timestamp start/stop: 2021-09-28 17:30:12 / 2021-09-28 17:30:28
    wal start/stop: 000000010000000200000075 / 000000010000000200000075
    database size: 17.5GB, database backup size: 4.8GB
    repo1: backup set size: 2.7GB, backup size: 666.8MB
    backup reference list: 20210928-161354F
```

The approximate size on disk of the backups are given by the "backup size" field. Please consult the pgbackrest documentation for more details. Full backups are annotated with "full backup". Incremental backups are annotated "incr backup".

To uncover how many bytes of diskspace is used in either case use

```
$ sudo du -bs <directory path>
```

A more human-readable result is possible using

```
$ sudo du -hs <directory path>
```

e.g.

```
$ sudo du -hs /crypt/pgbackrest_repo/backup/
3.4G  /crypt/pgbackrest_repo/backup/
```

The v43 upgrade database storage implications

1 Make an incremental backup of the PostgreSQL 9.6 database with an associated restore point.

The size of this backup is dependent on the data in the database that has changed. An indication of the likely storage requirements can be gauged from the size of the incremental backups taken over the last week. Use

```
$ sudo -u postgres pgbackrest --stanza=ibm-security-soar info
```

to retrieve this information. For planning purposes consider the maximum value of the previous week's incremental backups.

2 Migrate the v42 database from PostgreSQL 9.6 to 12.

This duplicates the existing 9.6 database, with some changes, resulting in an extra filesystem footprint normally equal to the 9.6 database. The size of the existing database can be determined by using

```
$ sudo du -hs /crypt/database
```

3 Delete the PostgreSQL 9.6 database.

This reduces the filesystem impact by the corresponding amount, roughly equal to the value retrieved in step 2.

4 Make a full backup of the PostgreSQL 12 database with an associated restore point.

Normally a backup of a PostgreSQL 12 database is the same size as a backup of a corresponding PostgreSQL 9.6 database. Use

```
$ sudo -u postgres pgbackrest --stanza=ibm-security-soar info
```

to retrieve this information. For planning purposes consider the full backup size from the previous week's backups.

After the upgrade has ended, and Resilient 43 is running, Resilient will perform certain v42 to v43 data migrations tasks. The extent of their impact on backup size will depend on the data in the Resilient database, and cannot be calculated in advance.

The high-water mark for filesystem usage during the v42 to v43 upgrade is just after the migration from PostgreSQL v9.6 to v12, incorporating the extra v12 database and the v9.6 backups. Collectively these might amount to 1.5 times the size of the existing v42 database, using a conservative estimate for the resulting files.

The upgrade .run file will initially check the free space on the filesystem, and will not proceed unless the requisite amount is available, plus a 10GB potential growth factor.

Filesystem locations to potentially reclaim storage space

It may be that space used by obsolete files could be reclaimed in order to meet the v43 install requirements. As well as customary locations, such as `/tmp`, there are SOAR-specific directories that could be examined.

System backups are stored in `/crypt/backups/`. Any such backups are likely to be individually large. SOAR log files are stored under `/usr/share/co3/logs` and certain directories under `/var/logs/`:

- `/var/logs/resilient-scripting`
- `/var/logs/resilient-email`
- `/var/logs/resilient-messaging`
- `/var/logs/resilient-app-manager`
- `/var/logs/elasticsearch`
- `/var/logs/pgbackrest`

Files under a `daily` subdirectory are compressed historical log files, marked for the day they were created. Though these files are compressed, for a long running active system they could account for a considerable amount of storage.

Running `resPackageLogs` before deleting a large number of log files may grant confidence that log files will exist for diagnosing unnoticed existing problems before the upgrade.