

# Mémoire technique

## Utilitaire d'archivage

Justin Bossard

Antoine Feuillette

Octobre 2024

## Table des matières

<b>1</b>	<b>Présentation du projet</b>	<b>2</b>
<b>2</b>	<b>Utilisation des variables</b>	<b>2</b>
2.1	Variables du fichier de configuration . . . . .	2
2.1.1	Configuration générale . . . . .	2
2.1.2	Serveur d'archivage . . . . .	2
2.1.3	Envoi de mails . . . . .	2
2.2	Variable utilisée dans le script . . . . .	3
<b>3</b>	<b>Organisation</b>	<b>3</b>
3.1	Script ( <code>archive.sh</code> ) . . . . .	4
3.1.1	Fonction d'écriture de logs . . . . .	4
3.1.2	Fonction d'envoi de mails . . . . .	4
3.1.3	Fonction de combinaison . . . . .	5
3.2	Fichier de configuration ( <code>archive.conf</code> ) . . . . .	6
3.3	Fichier de journaux ( <code>archive.log</code> ) . . . . .	6
3.4	Fichier de somme de contrôle ( <code>.prevChecksum</code> ) . . . . .	6
<b>4</b>	<b>Principe de fonctionnement</b>	<b>7</b>
<b>5</b>	<b>Justification des choix techniques</b>	<b>7</b>
5.1	Choix du langage de script . . . . .	7
5.2	Test pour des changements dans le dump SQL . . . . .	7
5.3	Envoi de mails . . . . .	8
5.4	Serveur Web . . . . .	8
5.5	Serveur SSH . . . . .	8
<b>Annexe</b>		<b>9</b>
5.6	Fichier de configuration ( <code>archive.conf</code> ) . . . . .	9
5.7	Script ( <code>archive.sh</code> ) . . . . .	10

# 1 Présentation du projet

Ce projet consiste à programmer un utilitaire d'archivage, qui télécharge une archive au [format ZIP](#) depuis un serveur Web. Cette archive contient un dump SQL, qui doit être à la racine et l'unique fichier SQL.

Cet utilitaire est prévu pour s'exécuter quotidiennement et déterminer si le dump SQL a subi des changements par rapport à la veille. Si c'est le cas, il faut archiver le dump sur un serveur distant via le protocole SFTP, dans une archive TGZ, au format `AAAADDMM.tgz`.

Il faut également établir un suivi par mail, et un historique des opérations dans un fichier de journalisation.

## 2 Utilisation des variables

### 2.1 Variables du fichier de configuration

Le fichier de configuration `archive.conf` comprend les variables suivantes :

#### 2.1.1 Configuration générale

**emplacementLog** Définit où les logs sont enregistrés. Attention, l'utilisateur qui exécute le script doit avoir les droits d'écriture dans le dossier parent.

Par défaut sur `./archive.log`.

**logStdout** En cas d'échec, redirige le motif à la sortie standard (0) ou pas (1).

Par défaut sur 0.

**archiveURL** Définit l'emplacement de l'archive via une URL, accessible depuis l'ordinateur client.

#### 2.1.2 Serveur d'archivage

**adresseArchivage** Adresse IP du serveur d'archivage, qui doit être accessible via SSH. Si le port 22 est utilisé, pas besoin de le préciser.

**usernameSSH** Nom d'utilisateur à utiliser sur le serveur d'archivage

**pathSSH** Chemin sur lequel enregistrer les archives sur le serveur. Le chemin doit déjà exister et être accessible en lecture et écriture pour l'utilisateur renseigné à **usernameSSH**.

**dureeConservation** Durée à partir de laquelle les anciennes archives seront supprimées, en jours.

Par défaut sur 30.

#### 2.1.3 Envoi de mails

**envoyerMail** Dans quel cas envoyer un mail, jamais (0), en cas d'échec de l'exécution (1) ou toujours (2).

Par défaut sur 1.

**mailDestinataires**=(dest1@mail.org dest2@mail.org) Destinataires du mail, séparés par des espaces. Si cette liste est vide, et peu importe la valeur de **envoyerMail**, le programme quittera sans envoyer de mail et sans erreur.

**objSucces** Objet du mail à envoyer en cas de succès.

Par défaut sur Archivage du \$(date +%d %B %Y') réussi.

**objEchec** Objet du mail à envoyer en cas d'échec.

Par défaut sur Archivage du \$(date +%d %B %Y') échoué.

**joindreLog** Dans quelle situation joindre le fichier de logs complet, jamais (0), en cas d'échec (1) ou toujours (2).

Attention, il s'agit du fichier de log entier, le motif d'échec, si c'est le cas, est toujours indiqué dans le corps du message.

Par défaut sur 1

**muttrcUtilisateur** Utiliser le ~/.muttrc de l'utilisateur (0) ou non (1).

Par défaut sur 1.

**2.1.3.1 Serveur SMTP** Ces options n'auront aucune incidence si **muttrcUtilisateur**=0.

**serveurHote** Serveur SMTP qui gère l'envoi de mails.

**port** Port sur lequel contacter le serveur. En général, on a :

- 25 : sans chiffrement
- 465 : chiffrement implicite (TLS/SSL)
- 587 : chiffrement explicite (STARTTLS)

**mailEnvoyeur** Mail envoyeur des informations de l'utilitaire, enregistré sur le serveur renseigné dans **serveurHote**.

**motDePasse** Mot de passe associé au mail pour s'identifier sur **serveurHote**.

## 2.2 Variable utilisée dans le script

Toutes les variables du fichier de configuration sont utilisées dans le script. On ne définit qu'une variable dans le script, comme tel :

```
currentChecksum=$(sha256sum ./*.sql | cut -d ' ' -f1)
```

Cela permet de stocker la somme de contrôle du dump SQL en cours de traitement pour pouvoir la comparer avec la somme de contrôle sauvegardée.

## 3 Organisation

On a l'organisation de fichier suivante :

```
.
|- archive.sh          # Script
|- archive.conf        # Fichier de configuration
|- archive.log         # Logs du script, possibilité de modifier l'emplacement
|- .prevChecksum      # Somme de contrôle du précédent fichier
```

### 3.1 Script (archive.sh)

Pour simplifier la lecture et l'écriture du programme, nous avons écrit une fonction pour écrire les logs, une fonction pour envoyer un mail, et une fonction qui combine les deux.

#### 3.1.1 Fonction d'écriture de logs

```
function ecrireLog() {
    if [[ $1 -eq 0 ]]; then
        echo "[ $(date +%T - %d %b %Y') ] : Succès, checksum=$2" >>
        ↪ "$emplacementLog"
    else
        [[ $logStdout -eq 0 ]] && echo "$2"
        echo "[ $(date +%T - %d %b %Y') ] : Échec, $2" >> "$emplacementLog"
    fi
}
```

Cette fonction admet deux paramètres, le premier définit le succès ou l'échec de l'opération, le second dépend du premier :

- en cas de succès, il s'agit de la somme de contrôle du fichier SQL sauvegardé
- en cas d'échec, il s'agit du motif de l'échec

On ajoute ensuite à la fin du fichier de log la date et l'heure d'écriture, si l'opération est un succès ou non, suivi de la checksum dans le premier cas et du motif d'échec dans le second.

#### 3.1.2 Fonction d'envoi de mails

```
function envoyerMail() {
    if [[ ${#mailDestinataires[*]} -ne 0 && (($1 -eq 1 && $envoyerMail -eq 1)
    ↪ || $envoyerMail -eq 2) ]]; then

        if [[ $muttrcUtilisateur -eq 0 ]]; then
            echo "$([[ $1 -eq 0 ]] && echo L'opération de ce jour est un
            ↪ succès || echo $2)" | \
            mutt -x \
                -s "$([[ $1 -eq 0 ]] && echo $objSucces || echo
                ↪ $objEchec)" \
                $([[ $joindreLog -eq 2 || ($1 -eq 1 && $joindreLog -eq 1)
                ↪ ]] && echo "-a $emplacementLog --") \
                "$(echo ${mailDestinataires[*]})"

        else
            echo "$([[ $1 -eq 0 ]] && echo L'opération d'archivage de ce
            ↪ jour est un succès. || echo $2)" | \
            mutt -nx \
                -e "set from = \"$mailEnvoyeur\" \" \
                -e "set smtp_pass = \"$motDePasse\" \" \
                -e "set smtp_url =
                ↪ \"$smtps://$mailEnvoyeur@$serveurHote:$port\" \" \
                -e "set send_charset = \"utf-8\" \" \
                -s "$([[ $1 -eq 0 ]] && echo $objSucces || echo
                ↪ $objEchec)" \
```

```

        $([[ $joindreLog -eq 2 || ($1 -eq 1 && $joindreLog -eq 1)
        ↪ ]] && echo "-a $emplacementLog --") \
        "$(echo ${mailDestinataires[*]})"
    fi
fi

[[ $? -ne 0 ]] && ecrireLog 1 "erreur lors de l'envoi du mail."
}

```

Cette fonction admet également deux paramètres, le premier définit le succès ou l'échec de l'opération, le second, n'est utilisé qu'en cas d'échec, et contient alors le motif de celui-ci.

Le booléen utiliser par le premier **if** permet d'envoyer un mail si la liste de destinataires du fichier de configuration n'est pas vide *et* que on est en situation d'envoyer un mail (échec et `envoyerMail=1`, ou `envoyerMail=2`).

On vérifie ensuite si l'utilisateur veut utiliser sa configuration de **mutt** ou non. Si ce n'est pas le cas, on précise que **mutt** ne doit pas utiliser le fichier de configuration de l'utilisateur (option `-n`), et on passe les paramètres nécessaires à l'envoi d'un mail en argument (option `-e "set option = \"valeur\""`) :

- **from** permet de définir l'identité de l'expéditeur. On utilise l'adresse mail renseignée dans le fichier de configuration
- **smtp\_pass** le mot de passe qui permet de se connecter au serveur SMTP
- **smtp\_url** Définit le serveur auquel se connecter et le protocole de connexion (**smtps** ici). On renseigne également le port de connexion défini dans la configuration.
- **send\_charset** Définit l'encodage du corps du mail et des pièces jointes. On choisit **utf-8** pour pouvoir prendre en compte l'ensemble des caractères Unicode.

On définit le sujet du mail avec l'option `-s`, qui varie en fonction de l'aboutissement de l'opération (on utilise la syntaxe : `[[ succès ? ]] && oui || non`).

Ensuite, selon les préférences de l'utilisateur, on joint ou non le fichier de log. Pour cela, on vérifie s'il faut joindre systématiquement le log, ou s'il faut le joindre en cas d'échec et que c'en est un. Si la condition est vraie, on renvoie l'option `-a` avec le chemin de log, sinon on ne renvoie rien.

Enfin, on affiche tous les éléments de la liste des destinataires.

### 3.1.3 Fonction de combinaison

```

function combo() {
    if [[ "$1" -eq 0 ]]; then
        ecrireLog 0 "$2"
        envoyerMail 0
    else
        ecrireLog 1 "$2"
        envoyerMail 1 "L'opération d'archivage de ce jour a échoué pour le
        ↪ motif suivant : $2"
    fi
}

```

Cette fonction combine simplement les deux fonctions précédentes, et permet d'unifier les motifs d'arrêts de fonctions. En premier argument, on définit le succès (0) ou l'échec (1) de l'opération. En second, on donne la checksum du fichier concerné en cas de succès, ou le motif de l'arrêt en cas d'erreur.

### 3.2 Fichier de configuration (**archive.conf**)

Le fichier de configuration est organisé en plusieurs sections :

1. Configuration générale
2. **Configuration du serveur d'archivage** Permet de configurer la connexion à un serveur SFTP via SSH
3. **Configuration des mails** Permet de configurer les infos générales des mails : quand en envoyer, objets, destinataires...
  - **Configuration serveur SMTP** Permet de définir comment se connecter au serveur d'envoi SMTP spécialement pour ce script, si nécessaire

### 3.3 Fichier de journaux (**archive.log**)

Le fichier de journalisation a la syntaxe suivante pour une ligne (en utilisant une version édulcorée de la syntaxe bash pour succès ou échec) :

```
[ heure - date ] : [[ Succès, checksum || Échec, cause de l'échec ]]
```

Une nouvelle entrée par ligne, en ordre décroissant d'ancienneté.

### 3.4 Fichier de somme de contrôle (**.prevChecksum**)

Le fichier **.prevChecksum** n'a pas vocation à être consulté par l'utilisateur, d'où le fait qu'il soit caché. Il permet de stocker la somme de contrôle en 256 bits du dump SQL précédent, pour pouvoir déterminer si des modifications sont advenues.

## 4 Principe de fonctionnement



## 5 Justification des choix techniques

### 5.1 Choix du langage de script

Nous avons choisi d'écrire le script en [bash](#), car il s'agit d'un langage de script clair et concis. De plus, il est parfaitement intégré à la plupart des distributions GNU/Linux, en tant que shell par défaut, s'intègre facilement aux distributions \*BSD<sup>1</sup>, et peut s'intégrer à Windows<sup>2</sup>.

### 5.2 Test pour des changements dans le dump SQL

Plusieurs manières sont envisageables pour détecter des changements dans un fichier. La manière la plus sûre consiste à conserver une copie du fichier et à comparer octet par octet avec un autre fichier. Néanmoins, cela pose deux problèmes principaux : tout d'abord, si les fichiers sont grands, cela peut prendre un temps non négligeable, ensuite, il est nécessaire de stocker constamment un fichier inutile autrement.

Nous avons donc décidé de procéder par calcul d'une [somme de contrôle](#) du dump, que l'on

<sup>1</sup><https://docs.freebsd.org/fr/articles/linux-users/#shells>

<sup>2</sup><https://korben.info/installer-shell-bash-linux-windows-10.html>

stocke dans le fichier `.prevChecksum`, et qui permet un suivi abstrait des modifications. Cela a plusieurs avantages : la comparaison prend toujours le même temps, l'espace de stockage nécessaire est négligeable, et cela permet d'avoir un historique en inscrivant la somme de contrôle dans les logs. De cette manière, s'il y a besoin d'utiliser des archives par la suite, on pourra vérifier l'intégrité de la sauvegarde avec sa somme de contrôle, en regardant dans l'historique des logs.

Nous utilisons la fonction `sha256sum`, de GNU coreutils, qui permet de calculer la somme de contrôle sur 256 bits d'un fichier.

Il est amplement suffisant et d'usage d'utiliser 256 bits pour ce genre de situation, pour limiter au maximum le risque de collision on pourrait la calculer sur 512 bits, néanmoins pour des fichiers de taille importante ce calcul prendra plus de temps et de ressources.

### 5.3 Envoi de mails

Pour effectuer l'envoi des mails, on utilise `mutt`, un logiciel libre permettant de se connecter simplement à un serveur SMTP.

Pour tester cette fonctionnalité, et dans la mesure où la plupart des fournisseurs mails ont arrêtés de fournir un accès SMTP à des applications comme `mutt` ([Google](#), [Microsoft](#), [Yahoo](#)<sup>3</sup>...), nous n'avons trouvé que qu'un fournisseur permettant d'accéder *gratuitement* à leurs serveurs par SMTP : [Zoho Mail](#).

Nous avons donc créer une adresse Zoho, et l'envoi de mails et de pièces a été concluant, en fonction de la configuration renseignée (pièce jointe systématique, en cas d'échec...).

### 5.4 Serveur Web

Pour pouvoir télécharger l'archive contenant le dump SQL à tout instant, nous avons mis en place un serveur Web avec [Apache](#), accessible via HTTP (port 80).

Pour cela, il suffit d'installer Apache et de le démarrer automatiquement. Ensuite, on place l'archive d'intérêt dans `/var/www/html` et on y accède avec l'adresse IP (e.g. : `http(s)://200.200.10.02/sql_dump.zip`).

### 5.5 Serveur SSH

Le serveur d'archivage doit être accessible via SSH, nous avons donc installé OpenSSH sur le serveur, il suffit de démarrer le démon SSH automatiquement puis d'accéder à la machine par le port 22 (par défaut).

---

<sup>3</sup>Fonctionne théoriquement, mais nous n'avons pas réussi, et certains posts laissent entendre que cette fonctionnalité est régulièrement désactivée.



# Annexe

## 5.6 Fichier de configuration (archive.conf)

```
1 #####
2 ## CONFIGURATION DE L'UTILITAIRE D'ARCHIVAGE ##
3 #####
4
5 ## Emplacement des logs de l'utilitaire
6 emplacementLog="archive.log"
7
8 ## En cas d'erreur, renvoyer le motif à la sortie standard en plus des logs
9 # 0 : vrai
10 # 1 : faux
11 logStdout="0"
12
13 ## Adresse de l'archive ZIP contenant un fichier SQL
14 archiveURL="http://176.190.35.242:80/sqldump.zip"
15
16
17 #####
18 ## SERVEUR D'ARCHIVAGE ##
19 #####
20
21 ## Adresse du serveur sur lequel on veut archiver les dumps
22 adresseArchivage="176.190.35.242"
23
24 ## Utilisateur qui va recevoir les sauvegardes, sur le serveur d'archivage
25 usernameSSH="john-doe"
26
27 ## Emplacement des sauvegardes, en chemin absolu (possibilité d'utiliser `~`)
28 pathSSH="~/Sauvegardes"
29
30 ## Durée de conservation des archives sur le serveur, en jours
31 dureeConservation="30"
32
33
34 #####
35 ## MAIL ##
36 #####
37
38 ## Dans quel cas envoyer un mail
39 # 0 : Jamais
40 # 1 : En cas d'échec
41 # 2 : Toujours
42 envoyerMail="1"
43
44 ## Adresses mail sur lesquels on veut envoyer un état de la dernière
45 ↪ exécution (succès / échec). Séparer chaque adresse par des espaces.
46 # Pour envoyer des mails internes, on peut utiliser les noms d'utilisateurs
47 ↪ (e.g. : root)
```

```

46 mailDestinataires=(dest1@mail.com)
47
48 ## Objet des mails en cas de succès
49 objSucces="Archivage du $(date +%d %B %Y') réussi"
50 ## Objet des mails en cas d'échec
51 objEchec="Archivage du $(date +%d %B %Y') échoué"
52
53 ## Joindre le log
54 # 0 : Jamais
55 # 1 : En cas d'échec
56 # 2 : Toujours
57 joindreLog="1"
58
59 ## Utiliser le `.muttrc` de l'utilisateur. Si oui, la modifications des
   ↪ options d'après n'aura aucune incidence.
60 # 0 : vrai
61 # 1 : faux
62 muttrcUtilisateur="1"
63
64 #####
65 ## SERVEUR SMTP ##
66 #####
67
68 ## Authentification sur serveur de messagerie, si nécessaire.
69 serveurHote="smtp.mail.org"
70
71 ## Généralement :
72 # 25 (sans chiffrement)
73 # 465 (chiffrement implicite)
74 # 587 (chiffrement explicite)
75 port="465"
76
77 mailEnvoyeur="envoyeur@mail.org"
78 motDePasse="mdp"

```

## 5.7 Script (archive.sh)

```

1  #!/bin/bash
2
3  # Importer le fichier de configuration
4  source "./archive.conf"
5
6  #####
7  ## Fonction d'écriture de log
8  # $1 : succès (0) / échec (1) de l'opération
9  # $2 : si échec, description
10 #      si succès, checksum du fichier concerné
11 function ecrireLog() {
12     if [[ $1 -eq 0 ]]; then

```

```

13         echo "[ $(date +%T - %d %b %Y) ] : Succès, checksum=$2" >>
           ↪ "$emplacementLog"
14     else
15         [[ $logStdout -eq 0 ]] && echo "$2"
16         echo "[ $(date +%T - %d %b %Y) ] : Échec, $2" >> "$emplacementLog"
17     fi
18 }
19
20 ## Fonction d'envoi de mail
21 # $1 : succès (0) / échec (1)
22 # $2 : si échec, corps du message
23 function envoyerMail() {
24     if [[ ${#mailDestinataires[*]} -ne 0 && (($1 -eq 1 && $envoyerMail -eq 1)
           ↪ || $envoyerMail -eq 2) ]]; then
25
26         if [[ $muttrcUtilisateur -eq 0 ]]; then
27             echo "$([[ $1 -eq 0 ]] && echo L'opération de ce jour est un
           ↪ succès || echo $2)" | \
28             mutt -x \
29                 -s "$([[ $1 -eq 0 ]] && echo $objSucces || echo
           ↪ $objEchec)" \
30                 "$([[ $joindreLog -eq 2 || ($1 -eq 1 && $joindreLog -eq 1)
           ↪ ]] && echo "-a $emplacementLog --") \
31                 "$(echo ${mailDestinataires[*]})"
32
33         else
34             echo "$([[ $1 -eq 0 ]] && echo L'opération d'archivage de ce
           ↪ jour est un succès. || echo $2)" | \
35             mutt -nx \
36                 -e "set from = \"$mailEnvoyeur\" \" \" \
37                 -e "set smtp_pass = \"$motDePasse\" \" \" \
38                 -e "set smtp_url =
           ↪ \"$smtps://$mailEnvoyeur@$serveurHote:$port\" \" \" \
39                 -e "set send_charset = \"utf-8\" \" \" \
40                 -s "$([[ $1 -eq 0 ]] && echo $objSucces || echo
           ↪ $objEchec)" \
41                 "$([[ $joindreLog -eq 2 || ($1 -eq 1 && $joindreLog -eq 1)
           ↪ ]] && echo "-a $emplacementLog --") \
42                 "$(echo ${mailDestinataires[*]})"
43         fi
44     fi
45
46     [[ $? -ne 0 ]] && ecrireLog 1 "erreur lors de l'envoi du mail."
47 }
48
49 ## Fonction combo
50 # $1 : succès (0) / échec (1)
51 # $2 : si échec, description
52 #     si succès, checksum du fichier concerné
53 function combo() {

```

```

54     if [[ "$1" -eq 0 ]]; then
55         ecrireLog 0 "$2"
56         envoyerMail 0
57     else
58         ecrireLog 1 "$2"
59         envoyerMail 1 "L'opération d'archivage de ce jour a échoué pour le
        ↳ motif suivant : $2"
60     fi
61 }
62
63 #####
64 # Ménage sur le serveur d'archivage
65 if ! ssh "$usernameSSH@$adresseArchivage" "find $pathSSH -name *.tgz -type f
    ↳ -ctime +$dureeConservation -exec rm '{} ' \; && exit"; then
66     combo 1 "impossible d'accéder au serveur d'archivage renseigné via SSH,
        ↳ vérifier la configuration."
67     exit 1
68 fi
69
70 # Vérifier si l'URL existe
71 if ! wget -q "$archiveURL"; then
72     combo 1 "URL renseignée inaccessible, ou `wget` n'est pas installé."
73     exit 1
74 fi
75
76 # Décompresser l'archive
77 if ! unzip -q ./*.zip; then
78     combo 1 "erreur lors de l'invocation de `unzip`."
79     exit 1
80 fi
81
82 # Supprimer l'archive
83 rm -f ./*.zip
84
85 # Vérifier si il y a un seul fichier SQL
86 if [[ ! $(ls ./*.sql | wc -l) -eq 1 ]]; then
87     combo 1 "structure d'archive à modifier, plusieurs fichiers SQL ou
        ↳ organisation inadéquate."
88     exit 1
89 fi
90
91 # Stocker la somme de contrôle du nouveau dump
92 if ! currentChecksum=$(sha256sum ./*.sql | cut -d ' ' -f1); then
93     combo 1 "fichier SQL non trouvé, ou `sha256sum` inaccessible."
94     exit 1
95 fi
96
97 # Si la sauvegarde de somme de contrôle n'existe pas, la créer
98 touch .prevChecksum
99

```

```

100 # Vérifier les changements
101 if [[ "$(cat .prevChecksum)" == "$currentChecksum" ]]; then
102     rm -f ./*.sql
103     combo 1 "somme de contrôle identique à la dernière enregistrée."
104     exit 1
105 fi
106
107 # Archiver le nouveau dump
108 if ! tar -czf "$(date +%Y%d%m')".tgz ./*.sql; then
109     combo 1 "erreur lors de la compression en tgz, vérifier que `tar` est
110         ↳ installé et accessible."
111     exit 1
112 fi
113
114 # Supprimer le dump SQL
115 rm -f ./*.sql
116
117 # Pousser sur le serveur, avec SSH (SFTP)
118 if ! scp -q ./*.tgz "$usernameSSH@$adresseArchivage:$pathSSH"; then
119     combo 1 "erreur lors de la connexion SSH au serveur d'archivage, vérifier
120         ↳ les informations de connexion."
121     exit 1
122 fi
123
124 # Supprimer l'archive
125 rm -f ./*.tgz
126
127 # Actualiser la somme de contrôle
128 echo "$currentChecksum" > .prevChecksum
129
130 combo 0 "$currentChecksum"
131 exit 0

```