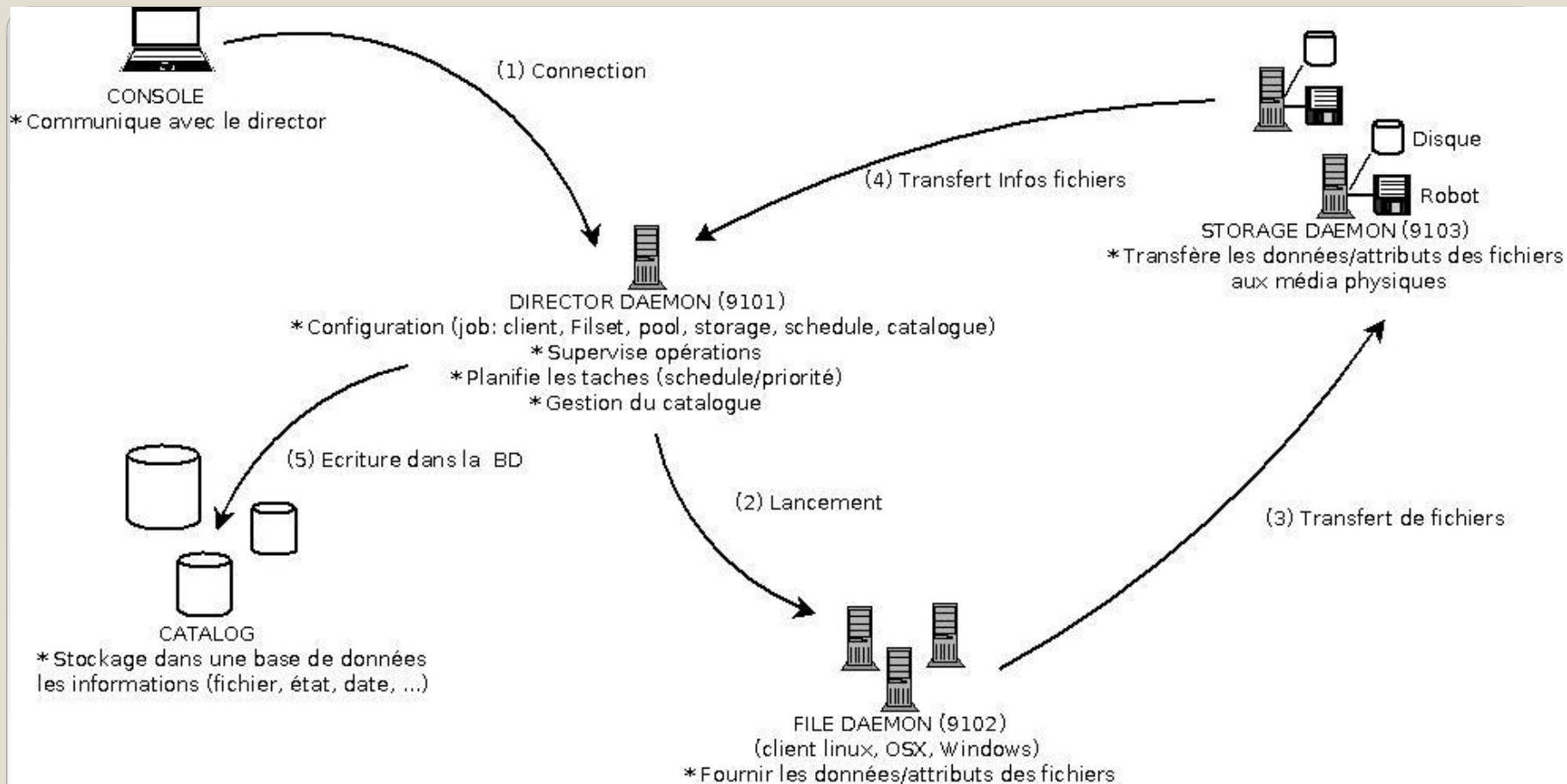


Bacula est un logiciel sous licence GPL,
permettant de **sauvegarder / restaurer / vérifier**
plusieurs ordinateurs sur un réseau hétérogène
(**windows, mac, linux**)
sur **différents supports** (disques, bandes,...)

Il fonctionne en mode client **Client/Serveur**.





Architecture modulaire et distribuée

- **Type Job : Sauvegarde, restauration, vérification**
(cartouche de données est lisible, intrusion à la façon de Tripwire), **migration, copy**
- **Niveaux de sauvegarde : Full, Incr, diff, since, VirtualFull (consolidation)**

```
Job {  
    Name = <nom du job>  
    Type = Backup  
    Level = Incremental  
    JobDefs = DefaultJob  
    Client = <client>-fd  
    FileSet = <nom du fileset>  
    Pool = <nom du pool>  
    Full Backup Pool = <nom du pool>  
    Storage = TS3100  
    Schedule = <nom du scheduleur>  
    RunBeforeJob = <script>  
    RunAfterJob = <script>  
}
```

job

Ghislain ESCORNE

- Adresse et mot de passe du client
- Connexion avec la base de donnée
- Rétention

```
Client {  
  Name = <client>-fd  
  Address = <client>  
  FDPort = 9102  
  Catalog = <catalogue>  
  Password = xxx # password for FileDaemon  
  File Retention = X days  
  Job Retention = Y months  
  AutoPrune = yes  
  Maximum Concurrent Jobs = Z  
}
```

Clients

- Inclure /exclure des fichiers
- Compression logicielle : GZIP, LZO
- Signature des fichiers : SHA1 , MD5

```
FileSet {  
    Name = <nom du fileset>  
    Include {  
        Options {  
            signature = MD5  
        }  
        File = <répertoire à sauvegarder>  
    }  
    Exclude {  
        File = *nosave  
    }  
}
```

Fileset

- **Gestion avancée des pools (fonction du niveau de sauvegarde), multi-volumes, multi-supports**
- **Marquage (label, label barcodes, label automatique suivant le format souhaité) des volumes pour prévenir tout écrasement accidentel.**
- **Data spooling pour écriture sur cartouche asynchrone.**

```
Pool {  
    Name = <nom du pool>  
    Pool Type = Backup  
    Recycle = no  
    AutoPrune = no  
    Volume Retention = X days  
    NextPool = Consolidation  
    Storage = <nom du storage>  
}
```

```
Pool {  
    Name = Catalogue  
    Use Volume Once = yes  
    LabelFormat = "Catalog-"  
    Pool Type = Backup  
    Recycle = yes  
    AutoPrune = yes  
    Volume Retention = 15 days  
}
```

Pool

- Connexion au serveur de stockage
- Quel device utiliser

```
Storage {  
  Name = File  
  Address = <adresse du serveur de stockage>  
  SDPort = 9103  
  Password = « XXX"  
  Device = FileStorage  
  Media Type = File  
  Maximum Concurrent Jobs = 2  
}
```

Storage

- **Schedule** interne pour le lancement automatique des jobs.
- **Exécution simultanée** de plusieurs Jobs.
- Séquencement des Jobs selon une hiérarchie de **priorités**.

```
Schedule {  
  Name = "15"  
  Run = Level=Full on 1 at 20:00  
  Run = Level=Incremental on 15 at 19:30  
  Run = Level=Differential on 30 at 20:00  
}
```

Ordonnancement

Connexion à la base de donnée

```
Catalog {  
  Name = <nom du catalogue>  
  dbname = "<nom de la base" ; dbuser = "<utilisateur>" ; dbpassword = "XXX"  
}
```

Catalogue

- Console d'administration
(commandes interactives,
états, ...)

```
# bconsole
Connecting to Director backup1:9101
1000 OK: backup1-dir Version: 5.2.1 (30
      October 2011)
Enter a period to cancel a command.
*help
Command      Description
=====      =====
label        Label a tape
list         List objects from catalog
mount        Mount storage
query        Query catalog
restore       Restore files
reload       Reload conf file
run          Run a job
status       Report status
use          Use catalog xxx
*
```

Communication avec le directeur

Scripts

- Interroger la bd
- Lancer des commandes

```
run job=f1-db2-heart level=incremental when="2009-01-10 20:00:00" yes
```

```
run job=f2-home level=incremental client=f2-fd pool=Full storage=TS3100  
when="2009-01-10 20:00:00" yes
```

*query

- 1: List up to 20 places where a File is saved regardless of the directory
- 2: List where the most recent copies of a file are saved
- 3: List last 20 Full Backups for a Client
- 4: List all backups for a Client after a specified time

selected

ted JobId

Files

ected Pool

ime

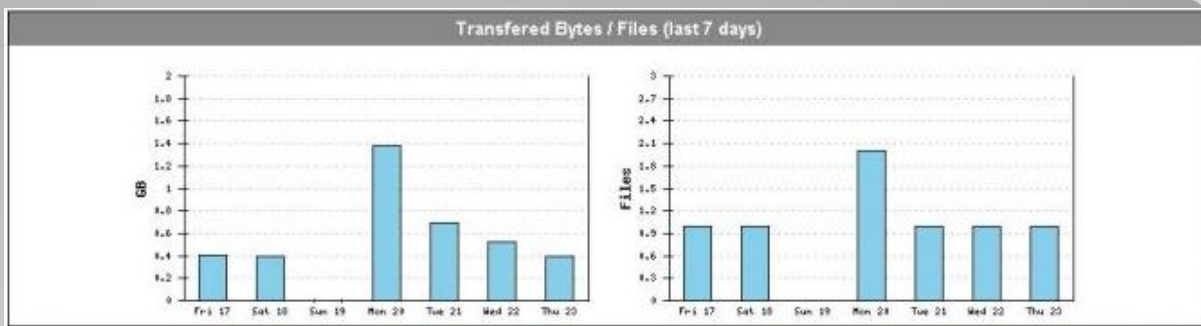
12: List files for a selected pool

...

Choose a query (1-27):

Communication avec le directeur

- Bat
- Bacula-web
- Page web perso



bud-sysop

JobId	Name	Level	JobStatus	Go	End	Duration	Pool	Volume storage
205	bud-sysop	F	T	3124	2011-09-09 17:34:18	1 day 08:59:38	Full-Sauvegardes	72-Full-Sauvegardes TS3100 73-Full-Sauvegardes TS3100
344	bud-sysop	D	T	540	2011-10-22 10:28:18	08:27:22	Diff-Sauvegardes	4-Diff-Sauvegardes TS3100
367	bud-sysop	I	T	159	2011-10-30 22:04:07	01:36:43	Inc-Sauvegardes	11-Inc-Sauvegardes TS3100
383	bud-sysop	F	T	3321	2011-11-10 21:51:54	2 days 02:08:24	Full-Sauvegardes	12-Full-Sauvegardes TS3100 13-Full-Sauvegardes TS3100
434	bud-sysop	D	T	197	2011-11-20 23:13:27	02:07:30	Diff-Sauvegardes	8-Diff-Sauvegardes TS3100
457	bud-sysop	I	T	179	2011-11-30 22:03:50	02:01:33	Inc-Sauvegardes	1-Inc-Sauvegardes TS3100
471	bud-sysop	F	A	3642	2011-12-09 19:19:05	2 days 15:50:10	Full-Sauvegardes	16-Full-Sauvegardes TS3100 21-Full-Sauvegardes TS3100
494	bud-sysop	I	T	177	2011-12-10 21:54:16	01:52:07	Inc-Sauvegardes	1-Inc-Sauvegardes TS3100

Interface graphique

21-Nov 16:12 backup1-sd JobId 10: Despooling elapsed
time = 00:20:52, **Transfer rate = 85.76 M Bytes/second**
5.2.1 (30Oct11):

Build OS: x86_64-unknown-linux-gnu debian 6.0.3
JobId: 10
Job: dugganite-geodynamo.2011-11-
18_21.18.56_17
Backup Level: Full
Client: "backup1-fd" 5.2.1 (30Oct11) x86_64-
unknown-linux-gnu,debian,6.0.3
FileSet: "dugganite-geodynamo" 2011-11-18
21:18:56
Pool: "i80" (From Job resource)
Catalog: "backup1" (From Client resource)
Storage: "i80" (From Pool resource)
Scheduled time: 18-Nov-2011 21:18:48
Start time: 19-Nov-2011 13:21:54
End time: 21-Nov-2011 16:57:06
Elapsed time: **2 days 3 hours 35 mins 12 secs**

Priority: 10
FD Files Written: **496,365**
SD Files Written: 496,365
FD Bytes Written: 2,545,303,284,096 (**2.545 TB**)
SD Bytes Written: 2,545,389,168,377 (2.545 TB)
Rate: **13705.6 KB/s**
Software Compression: None
VSS: no
Encryption: no
Accurate: yes
Volume name(s): **000045|000001**
Volume Session Id: 4
Volume Session Time: 1321629235
Last Volume Bytes: 1,182,668,368,896 (1.182 TB)
Non-fatal FD errors: 0
SD Errors: 0
FD termination status: OK
SD termination status: OK
Termination: **Backup OK**

Compte rendu détaillé par mail

Ghislain ESCORNE

- **Authentification par échange de mots de passe CRAM-MD5 entre chaque composant (*daemon*).**
- **Chiffrement TLS (ssl) entre chaque composant.**
- **Accès par console géré par ACL**

```
Console {  
  Name = restricted-user  
  Password = "UntrustedUser"  
  JobACL = "Restricted Client Save"  
  ClientACL = restricted-client  
  StorageACL = main-storage  
  ScheduleACL = *all*  
  PoolACL = *all*  
  FileSetACL = "Restricted Client's FileSet"  
  CatalogACL = DefaultCatalog  
  CommandACL = run  
}
```

Sécurité

- **bls** : liste des fichiers
- **bextract** : restauration des fichiers
- **bcopy** : copy d'une archive
- **bscan** : Régénérer le catalogue par balayage des volumes de sauvegarde
- **btape** : test du lecteur de bande, vérification d'une cassette, récupération du label
- **dbcheck** : test la cohérence de la base de données
- ..., Restauration d'un système complet (**bare metal**)

Utilitaire / outils autonomes

❖ ISTERRE

- ❑ 255 utilisateurs
- ❑ 200 To de données en production
- ❑ 20 Serveurs Linux + NAS (NDMP en projet avec la version bacula entreprise)
 - 80 To sauvegarde (bacula)
 - 20 To archive (bacula)
- ❑ 500 Postes de travail (BackupPc)

❖ IPAG

- Mails (300 Go)
- NAS (NDMP) en projet

Matériel :

- 2008 : IBM TS3100 (24 Lto4, 1 lecteur, 7000HT)
- 2010 : Baie AOE (3*24 2 To, 45000HT) (rsnapshot)
- 2011 : Quantum scalar i80 (80 Lto5, 2 lecteurs 12000 HT)

Contexte

Nom machine	http://url-tableau-de-bord-general								
	répertoire/ ou [module]	disponibilité	sensibilité	volume	descriptif	snapshots		sauvegardes	
						fréquence	rétenion	fréquence	rétenion
/un/repertoire/ici	J+0	2	100Go	blabla bla	1	15	10	60	
/un/autre/là	J+5	1	1To	bla bla	3	15	15	100	
[un_module_rsyncd]	J+0	3	< 300Go	bla bla bla	N/A	N/A	5	15	
[un_autre_module]	J+10	2	500Go	bla bla bla	1	15	N/A	N/A	

Ce tableau synthétise les sauvegardes souhaitées pour les machines *geodata*. On ne considère pas dans ce tableau l'aspect *archivage* qui portera sur des périodes plus longues.

- répertoire ou module : *fullpath* du répertoire (nb : la sauvegarde ne dépasse pas les frontières d'un *filesystem*) ou *[nom du module] rsyncd* à sauvegarder
- disponibilité : degré souhaité de disponibilité des données suite à une perte (valeur indicative). Choisir parmi :
 - J+0 (le jour même)
 - J+5 (sous une semaine max)
 - J+10 (sous deux semaines max)
- sensibilité : degré de sensibilité des données
 - 1 = peu sensible (les données peuvent être retéléchargées/recalculées/réécrites moyennant un travail faible)
 - 2 = sensible (les données peuvent être retéléchargées/recalculées/réécrites moyennant un travail important)
 - 3 = très sensible (les données sont impossibles à retrouver). Par convention, le niveau 3 implique la création d'une archive séparée
- volume : estimation du volume (préciser l'unité). Possibilité d'utiliser les signes de comparaison (ex: < 500Go).
- descriptif : descriptif succinct des données et de l'impact d'une perte éventuelle
- snapshots : configuration souhaitée du système de *snapshots* (possible uniquement sur certains systèmes de stockage)
 - fréquence : fréquence des snapshots (nombre moyen de jours entre chaque snapshots)
 - rétenion : âge le plus élevé (en nombre de jours) parmi les snapshots disponibles
- sauvegardes : configuration souhaitée du système de sauvegarde
 - fréquence : fréquence des sauvegardes (nombre moyen de jours entre chaque passage du système de sauvegarde, que ce soit pour un *full* ou un *incrément*)
 - rétenion : âge le plus élevé (en nombre de jours) parmi les sauvegardes disponibles

Planification

- Accès à la console => bconsole
- Lancement job => run
- Restitution => restore
- Visualisation de l'état général => status dir
- Visualisation de l'état du storage => status storage
- Visualisation du client => status client
- Visualisation de l'occupation des media => list media
- Interrogation de la base de données => query

Demo

- Gère la gestion avancée des pools et multi volume/support permet d'étendre la sauvegarde facilement.
- Gère le recyclage automatique
- Scripts (lancement de commandes, interrogation de la base de données)
- Compte rendu détaillé par mail (état, débit, quantité, ...)
- Consolidation des données
- Outil de sauvegarde complet

Fonctionnalités intéressantes

- Jobs en attente de libération de ressource (*Max Wait Time*)
- Pas de menu contextuel adapté et pas de possibilité de sortir => Il faut taper « . » Dans la console
- Client est attaché au catalogue
- Prise en charge du robot difficile (configuration à adapter)
- Charger manuellement une cassette quelques fois (*update slots scan storage=TS3100*)
- Démarrage long (dialogue avec le storage, base de données)

Problèmes survenus