

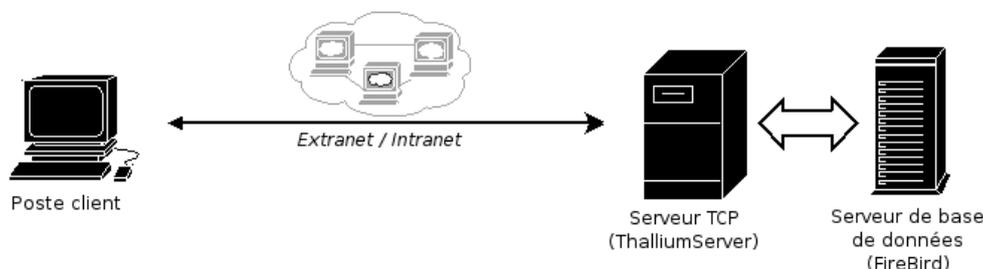
Le stockage de données avec UltraBackup

Ce document est destiné à proposer un aperçu des fonctionnalités de stockages telles qu'elles sont implémentées par UltraBackup 2007 et UltraBackup Netstation. Nous tentons ainsi de répondre à la légitime question « *mais comment sont traités mes données sauvegardées ?...* ». Astase autorise l'analyse et l'exploration des éléments dont il est question dans ce document, mais nous vous rappelons qu'elle se fait aux risques et périls des documents sauvegardés par le logiciel. Vous pouvez librement créer des outils qui exploitent les données présentés dans ce document, et si vous le désirez nous pourrions fournir sur demande via [nos forums](#) des informations complémentaires qui pourraient éventuellement compléter cette présentation.

Virtualisation de l'arborescence

Architecture d'UltraBackup

UltraBackup utilise à la base une architecture trois tiers pour fonctionner. Le « poste client » désigne l'ordinateur sur lequel fonctionne le client de sauvegarde. Le « serveur TCP » et le « Serveur de base de données » constituent les deux éléments du système de sauvegarde. Les versions courantes d'UltraBackup ne séparent plus le système de gestion des données du serveur en lui-même, bien qu'il soit possible en pratique de la réaliser. Les versions non Netstation d'UltraBackup fusionnent sur le même poste les trois entités dans une version plus légère adaptée à un fonctionnement local.



Le serveur TCP est implémenté sous la forme d'une application service Windows (bien qu'on puisse également le faire fonctionner en tant qu'application standard). Il gère la réception des données et s'interface avec la base de données qui est chargée de stocker toutes les informations sur les arborescences de données sauvegardées. Cette base de données permet de détecter les modifications apportées aux fichiers sans avoir besoin d'accéder au périphérique de stockage, et sert également à éviter de recopier la structure des données sauvegardées.

Base de données locale des métadonnées clientes

La base de données évoquée précédemment est au format **FireBird 1.5** et se situe aux emplacements suivants :

Stockage de données avec UltraBackup Netstation/UltraBackup 2007

- [%CSIDL_COMMON_APPDATA]\EmbeddedThalliumBackup\database\backupdb.fdb pour UltraBackup 2007,
- [%CSIDL_COMMON_APPDATA]\ThalliumBackup\database\backupdb.fdb pour UltraBackup Netstation.

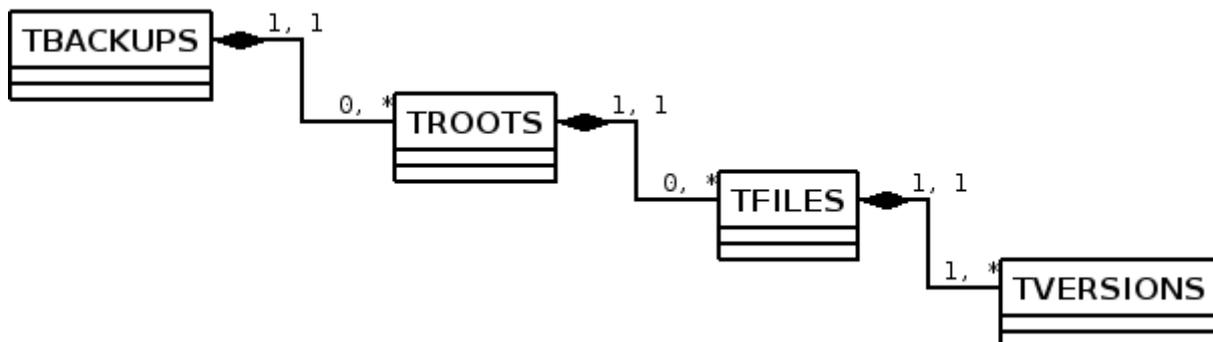
Vous pouvez librement consulter cette base de données avec un logiciel de gestion comme **IBExpert**, distribué par [HK Software](#). Néanmoins, nous déclinons toute responsabilité quand aux modifications manuelles des données effectuées dans la base.

Si vous souhaitez explorer la base de données, vous devez au préalable arrêter le service de sauvegarde. Ceci peut être fait en ligne de commande avec les droits administrateur en utilisant la ligne de commande suivante :

```
Net stop ThalliumServer
```

Racines de sauvegardes, Fichiers, et Versions

La base de données interne à UltraBackup possède, pour un backup donné, des informations qui permettent de décrire les données devant être sauvegardées et déjà sauvegardées. La modélisation adoptée par le logiciel est la suivante :



- Une sauvegarde est constituée de racines de sauvegardes, c'est-à-dire de dossiers ou de fichiers sources qui constituent les éléments de base à sauvegarder (table **TROOTS**).
- Chaque racine de sauvegarde contient l'ensemble fichiers à sauvegarder (table **TFILES**).
- Enfin, à chaque fichier est lié une ou plusieurs versions, c'est-à-dire un ou plusieurs fichiers correspondant à un état dans le temps de l'élément source (table **TVERSIONS**). Si la sauvegarde est configurée en mode simple, à tout fichier correspondra une et une seule version. Dans le cas contraire, le nombre de versions associées à un fichier dépendra du nombre de sauvegardes effectuées sur le document, du nombre de modifications qui lui ont été apportées et de la stratégie de suppression automatique des versions mise en place côté client.

Chaque objet de la base est identifié par un numéro unique codé sur 64 bits. UltraBackup utilise l'identifiant de version pour établir une correspondance entre la version et sa représentation stockée sur le disque.

Transfert des données

Système de transfert par blocs

Les données sont transférées suivant un modèle bloc-à-bloc. Chaque fichier est découpé en morceaux sur lesquels sont appliquées l'algorithme de compression éventuellement choisi puis l'algorithme de chiffrement. Les blocs sont éventuellement signés puis ils sont transmis au service de sauvegarde.

La taille des blocs est négociée durant l'envoi, en utilisant comme valeur par défaut celle spécifiée dans la fenêtre de configuration du gestionnaire de sauvegarde. Augmenter la taille des blocs améliore la vitesse de la sauvegarde mais augmente la quantité de mémoire vive nécessaire durant le processus de copie.

Le stockage des flux reçus conserve le découpage des données. La signature numérique des blocs, si elle est transmise par le client, est stockée dans l'archive afin d'être restituée lors de la restauration sans qu'un nouveau calcul par le serveur soit nécessaire.

Le découpage des flux doit impérativement être conservé, car sinon lors de la restauration il est impossible à UltraBackup d'appliquer dans le sens inverse les mêmes algorithmes sur les mêmes plages de données afin de restituer

Codage des algorithmes utilisés avant le transfert des versions

Chaque bloc de données peut être éventuellement compressé et encrypté. Le processus est effectué côté client avant l'envoi : le serveur stocke les données telles qu'il les reçoit sans effectuer de processus arrière. Lors de la restauration, le client récupère la liste des algorithmes utilisés et les applique de manière inverse afin de retrouver le contenu original du fichier.

Chaque algorithme est identifié par une chaîne de caractères unique qui est stocké dans la base de données et dans l'en tête du fichier stocké.

Les algorithmes suivant peuvent être utilisés pour chiffrer un fichier :

Identifiants :	Algorithme utilisé :
blowfish	Blowfish
3des	Triple DES
twofish	Twofish
serpent	Serpent
Rijndael	Rijndael
Rc6	RC6

Les algorithmes suivant peuvent être utilisés pour compresser un fichier :

Identifiants :	Algorithme utilisé :
deflate	Algorithme deflate, paramètres standard
maxdeflate	Algorithme deflate, compression maximum
fastdeflate	Algorithme deflate, compression la plus rapide

Gestion des fournisseurs de stockage

Pluralités des fournisseurs de stockage

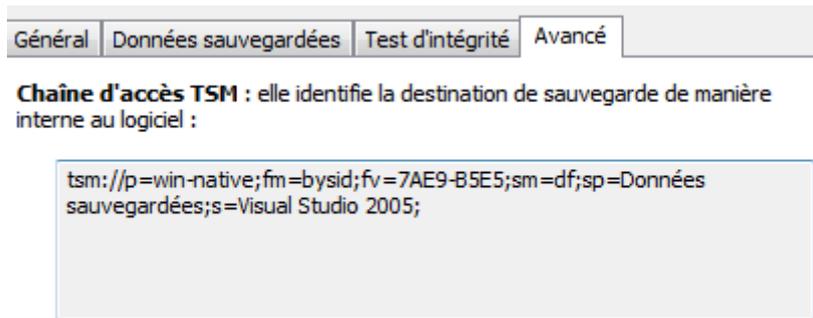
UltraBackup permet d'utiliser des fournisseurs de stockage différents d'une manière modulaire. Les fournisseurs de stockage sont programmés sous la forme de DLL devant implémenter une interface particulière. Le fournisseur de stockage dont il est question dans la suite de ce document est le fournisseur par défaut, dit « natif ». Il est exposé par la librairie **tsmnative.dll**. D'autres fournisseurs de stockage peuvent par exemple gérer les données transmises par le serveur de sauvegarde en les envoyant sur un serveur FTP ou en les écrivant sur des supports n'étant pas pris en charge par défaut par le système d'exploitation.

UltraBackup 2007 et Netstation 1 implémentent un stockage piloté par le serveur ; où le fournisseur doit uniquement fournir un jeu de fonctions permettant de manipuler un flux. Cela signifie que c'est le serveur qui impose son format de stockage et que le rôle du fournisseur de stockage consiste uniquement en le transfert des octets écrits par le serveur sur le média de sauvegarde. Cette approche est amenée à changer avec UltraBackup 5 où le service de sauvegarde n'aura plus accès au descripteur complet de flux, mais plutôt à une interface permettant de lire ou d'écrire des blocs de données. Il sera de la responsabilité du fournisseur de décider comment il ordonnera les blocs reçus et sous quel format il les stockera.

Chargement des fournisseurs de stockage

Chaque sauvegarde est reliée à un compte de sauvegarde, qui spécifie quel fournisseur utiliser. La définition des paramètres de stockage est réalisée via une chaîne préfixée par « tsm:// » stockée dans la base de données. Celle-ci se compose de paires clés/valeur, dont l'interprétation varie suivant le fournisseur. La clé « p » est obligatoire et permet de définir le code protocole lié à la chaîne, c'est-à-dire le fournisseur qui analysera les paramètres de la chaîne. La chaîne protocole correspondant au fournisseur explicité ici est « **win-native** », correspondant au stockage Windows natif.

Dans UltraBackup 2007, elle peut être affichée en utilisant le bouton « Stockage » puis l'onglet « Avancé » du gestionnaire de sauvegardes :



Dans le cas présent, `fm` identifie le type de montage de la ressource (ici réalisée par numéro de série), `fv` le numéro de série du volume de stockage, `sp` le point de stockage (c'est-à-dire dans le contexte de Windows, le dossier de sauvegarde), et `s` le nom de l'archive de stockage.

Lorsqu'une sauvegarde est exécutée, la librairie de gestion des fournisseurs, **tsmman.dll**, analyse la chaîne de définition des paramètres de stockage et cherche dans le répertoire `($bin)\tsmdrivers` la librairie prenant en charge la chaîne de protocole rencontrée.

Une fois trouvé, la librairie est chargée et est exploitée pour stocker les flux que le service lui transmet.

Stockage des versions

Arborescence de répertoires

Les archives de sauvegarde générées par UltraBackup se composent d'un dossier comportant deux sous-dossiers. Le premier, nommé « Storage », contient les fichiers sauvegardés. Le second, nommé « System », contient des informations supplémentaires utilisées généralement lors de l'import de la sauvegarde. Ce format n'a pas été choisi au hasard : sans compression, il permet en cas de corruption d'une partie du disque d'accéder quand même aux éléments non endommagés, et comportant un fichier pour une version, il autorise le mirroring direct sur d'autres sites de stockage via FTP ou Rsync en n'imposant pas le transfert de fichiers non modifiés par le client. Par défaut, UltraBackup demande cependant au système de fichiers sous-jacent d'effectuer une compression à son niveau lorsque c'est possible (par exemple, les fichiers seront automatiquement compactés si vous utilisez NTFS qui supporte l'attribut `FILE_ATTRIBUTE_COMPRESSED`¹).

UltraBackup Netstation et UltraBackup 2007 stockent (via **tsmnative.dll**) chaque version dans un fichier différent, à l'intérieur du dossier « Storage ». L'arborescence cliente n'est pas reproduite et UltraBackup utilise l'identifiant unique de version pour déterminer le répertoire de stockage de la version. Cette méthode est rapide, immédiate, et ne nécessite pas de reproduire l'arborescence cliente ni d'appliquer les métadonnées (attributs, droits, dates de modification...) du fichier original.

¹ Il est d'ailleurs vivement recommandé d'utiliser NTFS comme système de fichiers pour la sauvegarde en raison de sa fiabilité et de sa forte tolérance aux pannes

Stockage de données avec UltraBackup Netstation/UltraBackup 2007

Pour uniformiser le nombre de fichiers dans les répertoires de stockage, UltraBackup crée une arborescence basée sur le numéro de version. UltraBackup utilise deux répertoires intermédiaires ; le premier correspond au numéro de version divisé par 1 000 000, le second au numéro de version divisé par 1000. Les versions validées sont stockées dans le répertoire associé, sous la forme « S<Numéro de version>.part ». Si d'autres fichiers ne validant pas ce motif sont présents dans le répertoire de stockage, c'est qu'il s'agit de fichiers temporaires qui n'ont pas été supprimés lors d'un crash du serveur.

Le dossier « System » contient quand à lui deux fichiers correspondant à un export de la base de données locale pour la sauvegarde :

- « *backupmd.bsys* » contient les informations sur la structure de la sauvegarde.
- « *backupcontent.bsys* » contient les informations sur le contenu de la sauvegarde. Dans ce fichier sont exportés toutes les informations de la base de données concernant les fichiers sauvegardés de la sauvegarde (correspondance entre numéros de version/fichier, attributs et méta données des éléments, etc.)

UltraBackup régénère ces fichiers à chaque fois qu'une sauvegarde a été exécutée et que des fichiers ont été mis à jour.

Lorsqu'une sauvegarde doit être importée, le service d'import ouvre et décode ces deux fichiers, extrait les informations et met à jour la base de données locale. Une fois ceci fait, la sauvegarde réapparaît dans le gestionnaire de sauvegardes et on peut continuer à l'exploiter.

Format de fichier

UltraBackup stocke dans le fichier les métadonnées de la version d'origine ainsi que les blocs de données reçus. Cette section détaille le format de stockage pour que vous puissiez éventuellement relire les fichiers écrits par UltraBackup.

Il existe deux versions de ce format : la version finale, utilisée dans les versions finales d'UltraBackup, et la version préliminaire, utilisée dans les produits basés sur le noyau utilisé par UltraBackup distribués en bêta-test depuis 2005 (incluant Simbo). Entre les deux versions, des champs ont été ajoutés dans l'en-tête de fichier : ceux-ci apparaissent en rouge dans le tableau ci-dessous :

Type et taille (octets) :	Rôle :
Non signé, 8 bits	Numéro de la version courante. Peut valoir 1 ou 2.
Entier signé, 64 bits	Numéro unique de la racine parente dans la base de données.
Non signé, 8 bits	Type de la racine parente. Pour l'instant, ce champ n'est pas utilisé.
Chaîne C ² , 512 car.	Chemin de la racine parente.
Chaîne C, 512 car.	Chemin du fichier original, relatif à la racine courante.
Chaîne C, 64 car.	Hash du fichier tel qu'il a été fourni par le client. Si le client n'a pas hashé le fichier, ce champ est vide.
Entier non signé, 32 bits	Attributs du fichier original renvoyés par l'API Windows GetFileAttributes .
Entier signé, 64 bits	Taille originale du fichier sur la plate forme cliente.
Entier signé, 64 bits	Dernière date de modification du fichier, au format Unix.

² Par chaîne C, on entend un chaîne ASCII terminée par le caractère nul (#0).

Stockage de données avec UltraBackup Netstation/UltraBackup 2007

Entier signé, 64 bits	Date de création du fichier, au format Unix.
Entier signé, 64 bits	Date de dernier accès, au format Unix.
Entier non signé, 32 bits	Droits Unix du fichier, si le client est un client Unix.
Entier non signé, 32 bits	UserID du fichier, si le client est un client Unix.
Entier non signé, 32 bits	GroupID du fichier, si le client est un client Unix.
Non signé, 8 bits	Informations supplémentaires réservées par le fournisseur de stockage.
Chaîne C, 16 car.	Code de l'algorithme ayant compressé les blocs du fichier. Si les blocs n'ont pas été compressés, ce champ est vide.
Chaîne C, 16 car.	Code de l'algorithme ayant chiffré les blocs du fichier. Si les blocs n'ont pas été chiffrés, ce champ est vide.
Entier signé, 32 bits	Nombre de blocs du fichier. Ce nombre ne peut pas valoir zéro.

Commencent ensuite les données sauvegardées. Comme on l'a dit précédemment, le découpage bloc est conservé et ceux ci se présentent ainsi :

Type et taille (octets) :	Rôle :
Entier signé, 32 bits	Taille du bloc. Le bloc ne peut excéder 2 Go, sa taille réelle est en général d'une dizaine de méga octets.
Chaîne C, 64 bits.	Hash MD5 du bloc. Si le client ne l'a pas haché, il est stocké ici une chaîne vide.
Les données occupent exactement la taille indiquée précédemment. Si le bloc n'est pas le dernier bloc sauvegardé du fichier, le champ suivant immédiatement le dernier octet de données indique la taille du nouveau bloc (et la structure présentée ici se poursuit dans le fichier).	

Comme dit précédemment, nous nous tenons à votre disposition pour toute information à propos de ce document via nos forums.