



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

PARIS

MEMOIRE

Présenté en vue d'obtenir

LE DIPLOME D'INGENIEUR CNAM

SPECIALITE : INFORMATIQUE

OPTION : SYSTEMES D'INFORMATION

Par

Gaëtan UVALLE

**Optimisation d'une infrastructure de sauvegarde :
étude, automatisation et déploiement du composant RMAN.**

Soutenu le 21 Mars 2011, à PARIS

JURY

Président: Monsieur Jacky AKOKA,
Professeur (CNAM, Paris)

Membres : Madame Isabelle WATTIAU,
Professeur (CNAM, Paris)

Madame Samira SI-SAID CHERFI,
Maître de Conférences (CNAM, Paris)

Monsieur Frédéric BEAUBOIS
Ingénieur, Expert Technique Sénior (ATOS-ORIGIN)

Madame Jamila DERRADJ
Responsable de l'équipe d'administration
des bases de données de production (SANOFI-AVENTIS)

RESUME

Résumé :

L'optimisation des sauvegardes de bases de données Oracle, oblige SANOFI-AVENTIS à introduire un nouvel outil appelé **R**ecovery **M**ANager (RMAN) dans l'infrastructure de sauvegarde existante.

Ce mémoire explique pourquoi RMAN est requis et comment il a été mis en œuvre au cours d'un projet réalisé en deux parties.

La première phase du projet a étudié l'existant c'est-à-dire les divers composants logiciels (RMAN, logiciel de sauvegarde DataProtector, et ordonnanceur VTOM) de l'infrastructure de sauvegarde. Cette étape a abouti à la conception d'une maquette de la solution. L'outil RMAN a ensuite été mis en œuvre sur un nombre limité de bases.

La deuxième phase du projet s'est attachée à généraliser la solution sur un périmètre aussi large que possible (environ 600 à 700 bases), puis à optimiser et fiabiliser le déploiement.

Mots clés : Recovery Manager, RMAN, sauvegarde, restauration, Oracle, sauvegarde incrémentale, sauvegarde différentielle, sauvegarde cumulative.

SOMMAIRE

REMERCIEMENTS.....	6
INTRODUCTION.....	7
Chapitre 1: ETAT DE L'ART SUR LES SAUVEGARDES.....	9
1.1 GENERALITES	9
1.1.1 Définitions	9
1.1.2 Utilité de la sauvegarde	10
1.1.3 Caractéristiques des sauvegardes.....	11
1.2 INFRASTRUCTURE DE SAUVEGARDE	16
1.2.1 Le matériel.....	16
1.2.2 Les logiciels.....	19
1.2.3 La chaîne de sauvegarde.....	23
1.3 ORGANISATION.....	25
1.3.1 Politique de sauvegarde	25
1.3.2 Les acteurs de la sauvegarde.....	27
1.4 RESUME DU CHAPITRE 1	30
Chapitre 2: SAUVEGARDES DE BASES ORACLE.....	32
2.1 PRINCIPES DE FONCTIONNEMENT D'UNE BASE ORACLE	32
2.1.1 Instance et base de données	32
2.1.2 Fichiers d'une base de données	34
2.1.3 Fichiers « Redolog » et mode archivage.....	34
2.2 METHODES DE SAUVEGARDES DES BASES ORACLE	35
2.2.1 Sauvegardes physiques	35
2.2.2 Sauvegardes logiques (exports).....	36
2.2.3 Sauvegardes avec l'outil RMAN.....	37
2.3 RESUME DU CHAPITRE 2	43
Chapitre 3 : PRESENTATION DU SUJET	44
3.1 PRESENTATION DE L'ENTREPRISE ET DU SERVICE.....	44
3.1.1 L'entreprise.....	44
3.1.2 Le département administration des bases de données.....	45
3.2 L'EXISTANT : LES SAUVEGARDES A SANOFI.....	47
3.2.1 Infrastructure au commencement du projet	47
3.2.2 Le matériel mis en oeuvre.....	47
3.2.3 Eléments de la politique de sauvegarde.....	48
3.3 DESCRIPTION DU PROJET	49
3.3.1 Problématique.....	49
3.3.2 Objectifs du projet	51

3.3.3 Périmètre du projet	51
3.3.4 Démarche et déroulement du projet.....	51
3.4 LE CONTEXTE	55
3.4.1 Contexte organisationnel	55
3.4.2 Les contraintes	57
3.5 RESUME DU CHAPITRE 3	59
Chapitre 4 : RMAN SUR UN PERIMETRE REDUIT	61
4.1 ETUDE DE L'EXISTANT	61
4.1.1 Présentation de la chaîne de sauvegarde existante.....	62
4.1.2 L'ordonnanceur VTOM.....	63
4.1.3 Le logiciel de sauvegarde DataProtector	71
4.1.4 Architecture du composant RMAN	77
4.2 MISE AU POINT DE LA MAQUETTE	81
4.2.1 Description et méthodologie de mise au point de la maquette	81
4.2.2 Préparation de l'outil RMAN	83
4.2.3 Mise en place de l'interface de communication DataProtector / RMAN	84
4.2.4 Mise au point des nouveaux modèles et spécifications de DataProtector.....	86
4.2.5 Mise au point des nouveaux enchainements VTOM.....	91
4.2.6 Tâches réalisées pour la mise au point de la maquette	96
4.3 PILOTE PHENIX	98
4.3.1 Objectifs du pilote PHENIX	98
4.3.2 Démarche utilisée pour le pilote PHENIX	99
4.3.3 Retour d'expérience sur le pilote PHENIX	100
4.4 RESUME DU CHAPITRE 4	102
Chapitre 5: GENERALISATION ET DEPLOIEMENT	104
5.1 OPTIMISATION	105
5.1.1 Démarche suivie pour optimiser la solution	105
5.1.2 Etude de l'organisation des catalogues.....	106
5.1.3 Etude de l'automatisation de la mise en œuvre	112
5.2 VALIDATION SUR L'ENSEMBLE DU PERIMETRE	115
5.2.1 Démarche pour valider la solution sur l'ensemble du périmètre	116
5.2.2 Extension aux autres versions Oracle et OS	117
5.2.3 Extension du périmètre aux serveurs mutualisés.....	118
5.2.4 Tâches réalisées pour étendre le périmètre	119
5.3 PREPARATION ET FORMATION.....	120
5.3.1 Formation et assistance aux changements	121
5.3.2 Préparation de l'infrastructure	121
5.4 PILOTE DE MIGRATION (SHIVA)	122
5.4.1 Objectifs du pilote SHIVA	122
5.4.2 Tâches réalisées lors du pilote SHIVA.....	123
5.4.3 Retour d'expérience du pilote SHIA	124
5.5 DEPLOIEMENT	125
5.5.1 Périmètre.....	125
5.5.2 Lotissement.....	125
5.5.3 Déploiement et retour d'expérience.....	126
5.5.4 Retour d'expérience de l'utilisation de RMAN	127

5.6 RESUME DU CHAPITRE 5	127
Chapitre 6 : BILAN, RESULTATS & PERSPECTIVES	129
6.1 RESULTATS DU PROJET	129
6.2 BILAN PERSONNEL.....	130
6.2.1 Ma contribution au projet	130
6.2.2 Bénéfices personnels tirés du projet	131
6.2.3 Limites du projet.....	131
6.3 PERSPECTIVES.....	132
BIBLIOGRAPHIE	134
LISTE DES FIGURES.....	138

REMERCIEMENTS

Je tiens à remercier vivement :

- Madame Isabelle WATTIAU pour sa disponibilité, pour m'avoir relu et m'avoir donné de précieux conseils.

- Madame Jamila DERRADJ, pour m'avoir soutenu lors de tout mon cursus d'ingénieur durant ces 5 années.

- Mademoiselle Françoise FARGIER, Madame Caroline PARDIEU, et Monsieur Jean MIART pour avoir pris de leur temps libre et m'avoir relu.

- Monsieur Sylvain HERVY qui a mis à ma disposition un environnement de test.

- Monsieur Frédéric BEAUBOIS toujours prêt à me faire partager son expertise.

- Mon épouse et ma famille pour leur patience, leur soutien et leurs encouragements.

INTRODUCTION

Pour remplir leurs objectifs de croissance, les entreprises cherchent à gagner en productivité et consomment de plus en plus de ressources informatiques, tant logicielles, que matérielles. Elles ont besoin en particulier, de plus en plus de bases de données. Sanofi-Aventis, n'échappe pas à cette règle. En 2002, l'équipe d'administration de bases de données (à laquelle j'appartiens) gérait 60 bases de données Oracle. Aujourd'hui elle administre un parc de plus 750 instances.

Les données stockées en bases sont d'une importance capitale pour l'entreprise. La perte d'une base de données pourrait la gêner, la paralyser, et même dans les cas extrêmes, lui être fatale. Il faut donc prévoir la perte inattendue des données et être en mesure de récupérer ces données si précieuses. C'est pourquoi, il est nécessaire de les sauvegarder, c'est-à-dire les dupliquer sur bandes pour les mettre ensuite en sécurité en lieu sûr.

Sanofi-Aventis dispose d'une solide infrastructure de sauvegarde.

Ce mémoire concerne un projet de modernisation de l'infrastructure des sauvegardes des bases de données Oracle. L'objectif est d'optimiser le service offert, en réduisant les volumes et les temps de sauvegarde. Ces améliorations seront obtenues grâce à l'introduction d'un nouveau composant, **Recovery MANager** (RMAN) dans l'infrastructure de sauvegarde existante :

Ce document a été scindé en 6 chapitres :

- **Le premier chapitre** est un état de l'art sur les sauvegardes, qui synthétise les connaissances utiles pour aborder les notions nécessaires dans ce mémoire.
- **Le deuxième chapitre** décrit l'architecture d'ensemble d'une base de données ainsi que les différentes méthodes de sauvegardes d'une base de données Oracle.

- **Le troisième chapitre** décrit le contexte du projet, pose la problématique, expose les limitations des sauvegardes actuelles et explique comment RMAN peut répondre à notre besoin.

- **Le quatrième chapitre** étudie l'existant ainsi que les composants impliqués dans la sauvegarde : le logiciel de sauvegarde DataProtector, l'ordonnanceur VTOM, puis l'outil RMAN lui-même. On verra comment une première maquette de la solution RMAN a été montée et comment RMAN a été mis en place pour un premier pilote.

- **La cinquième partie** étudie les moyens de rationaliser et généraliser la solution à l'ensemble des bases. Cette partie explique ensuite comment le déploiement a été automatisé, organisé puis mis en œuvre.

- **La sixième et dernière partie** tire le bilan du projet.

Chapitre 1: ETAT DE L'ART SUR LES SAUVEGARDES

Afin de faciliter la compréhension des chapitres qui suivent, nous commencerons par passer en revue les notions et les concepts importants de la sauvegarde.

1.1 GENERALITES

Nous démarrerons par quelques définitions et nous verrons le rôle et l'importance des sauvegardes. Nous traiterons ensuite des diverses caractéristiques de sauvegardes. Enfin, nous verrons les composants d'une infrastructure de sauvegarde, et comment cette dernière est organisée.

1.1.1 Définitions

Une **sauvegarde** (ou backup en anglais) est l'opération qui consiste à dupliquer et à mettre en sécurité les données contenues dans un système informatique sur un support (CD-ROM, DVD-ROM, bande, etc.) différent de celui qui les contient habituellement. La sauvegarde doit être vue comme une copie de documents et données à protéger. Toutefois, il ne s'agit pas seulement d'une simple copie, mais surtout d'une copie sûre.

La **Restauration** est l'opération inverse de la sauvegarde, qui consiste à récupérer les données originales sauvegardées à partir de leur support de sauvegarde. La restauration doit permettre une reprise d'activité de l'entreprise suite à un sinistre ou un incident et/ou doit permettre si nécessaire la reconstruction du système d'information, en partie ou en totalité.

1.1.2 Utilité de la sauvegarde

1.1.2.1 Pourquoi sauvegarder ?

Les données stockées dans le système d'information font partie de la richesse immatérielle d'une entreprise et représentent une valeur. A tel point qu'une loi de janvier 1978 rend obligatoire la protection de ces informations. Les sauvegardes doivent être perçues comme une assurance contre la perte de ces données. Suite à un incident, les données perdues pourraient entraîner l'indisponibilité de moyens informatiques, gêner l'activité de l'entreprise et dans les cas extrêmes, la paralyser. On estime que 50% des entreprises qui ont subi un sinistre informatique majeur disparaissent dans l'année qui suit. (*Source SNIA France - 2006*). On comprend alors l'importance qu'il faut accorder aux sauvegardes. La sauvegarde s'inscrit dans une démarche globale qui consiste à assurer la continuité d'activité d'un système informatique. Elles sont un élément indispensable de la sécurité du système d'information, au même titre que la lutte contre les virus, ou le contrôle des accès aux données de l'entreprise. Les sauvegardes sont donc nécessaires pour assurer la pérennité des activités et éviter des pertes financières significatives

1.1.2.2 Risques couverts par la sauvegarde

Les sauvegardes visent à se protéger de deux catégories d'incidents, ceux qui sont causés par l'activité humaine, et ceux qui sont la conséquence de dysfonctionnements informatiques, qu'ils soient logiciels ou matériels. Selon un rapport du CLUSIF (**CLU**b de la **S**écurité de l'**I**nformation **F**rançais), les sources de perte de données informatiques dans l'entreprise sont imputables à des défaillances matérielles dans 44 % des cas et à des erreurs humaines dans 32 % des cas, les autres causes de perte de données étant les virus, les vols, les incendies, les inondations...

Les erreurs de manipulation ou les actes de malveillance peuvent altérer ou supprimer les données. Dans ce cas la sauvegarde doit faciliter la restauration de tout ou partie d'un système

informatique (un fichier, un groupe de fichiers, un système d'exploitation, une donnée dans un fichier, etc.) suite à une suppression accidentelle ou à une modification non désirée.

Les pannes informatiques peuvent altérer ou rendre inaccessibles des données. Elles peuvent survenir suite à :

- une panne matérielle (Panne de disques, panne électronique, etc.)
- par un sinistre (Une inondation, un incendie, un vol, la foudre, etc.)
- un dysfonctionnement logiciel

1.1.3 Caractéristiques des sauvegardes

Nous pouvons caractériser une sauvegarde selon deux critères : fonctionnels et techniques.

1.1.3.1 Aspects fonctionnels des sauvegardes

Le Périmètre: En premier lieu, il faut pouvoir répondre à la question : quelles données doit-on sauvegarder ? Identifier le périmètre passe par l'inventaire des données à sauvegarder: documents, fichiers, bases de données, systèmes d'exploitation, applications, emails, etc.

L'importance et la criticité des données : Il faut classer les données selon leur importance. Les données vitales, c'est-à-dire celles qui par leur indisponibilité mettraient en péril l'entreprise ou paralyseraient momentanément l'activité, doivent être clairement identifiées.

Le type de données : Les sauvegardes peuvent être classées selon le type de données :

- **Systemes** : elles sont nécessaires au système d'exploitation. Elle nécessite des méthodes de sauvegarde particulière avec des utilitaires livré généralement avec le système d'exploitation (« Mksysb » pour UNIX AIX, « Ignite » pour UNIX HP-UX, Norton Ghost pour WINDOWS, etc.)

- **Applicatives** : elles sont nécessaire au fonctionnement de l'application. Elles peuvent nécessiter l'utilisation d'utilitaire qui s'interface généralement avec le logiciel de sauvegarde.
- **Fichiers autonomes**: Ils peuvent être copiés de manière indépendante. (Exemple un fichier texte)

La tolérance aux pertes: La perte d'information peut être tolérée ou non. On accepte de perdre les modifications d'une journée dans la plupart des cas. Si l'on restaure un système sans pertes, on parlera de restauration « *complète* ». Dans le cas inverse on parlera de restauration « *incomplète* ». La mise en place de solutions permettant des sauvegardes/restaurations complètes est généralement complexe à mettre en œuvre et réservée aux applications critiques ou sensibles.

L'accessibilité de données lors de la sauvegarde: Lors d'une sauvegarde, les données peuvent être ou non accessibles. Si les données sont sauvegardées en même temps qu'elles sont manipulées par les utilisateurs, on parlera indifféremment de sauvegardes « *en ligne* », « *online* », « *hot backup* » ou « *à chaud* ». Si les données ne sont pas accessibles, on parlera de sauvegardes « *hors ligne* », « *offline* », de « *cold backup* », ou de « *sauvegardes à froid* ». Les sauvegardes à froid ont l'avantage d'être simples à mettre en place et d'être faciles à restaurer, mais aucune requête ne peut interroger la base pendant toute la durée de la sauvegarde. Pendant les sauvegardes à chaud, les utilisateurs peuvent accéder aux données. En revanche, la mise en œuvre de ces sauvegardes et les restaurations associées sont techniquement complexes et, nécessitent un savoir faire important.

Contraintes d'exploitation et fenêtre de sauvegarde: Les *fenêtres de sauvegarde* sont les périodes pendant lesquelles les sauvegardes sont effectuées. Il est important de les connaître et de tenir compte des contraintes d'exploitation. S'il s'agit de sauvegarde offline où les applications sont inaccessibles, aucun traitement ne peut être lancé. Si, à l'inverse, il s'agit de sauvegarde online, on évite les modifications de données en masse pendant les sauvegardes, ce qui pourraient ralentir les traitements.

Temps de restauration : il faut être capable de restaurer dans un temps maximal déterminé à l'avance. Le temps maximum de restauration influe fortement sur le choix des techniques à mettre en place.

Rétention des sauvegardes : La rétention est le temps durant laquelle les données seront conservées sur support magnétique, et protégé contre l'écrasement. La rétention peut être dictée par 2 facteurs : les considérations réglementaires (Obligation fiscale par exemple) et le besoin des utilisateurs de conserver certaines informations (mail, ou documents de facturation par exemple). La durée de rétention peut varier de quelques jours à plusieurs années.

1.1.3.2 Caractéristiques techniques des sauvegardes

A partir des contraintes fonctionnelles fournies par les utilisateurs, on peut déterminer les caractéristiques techniques.

Sauvegarde locale versus sauvegarde réseau : Les sauvegardes **locales** se font directement sur la machine. Les données à copier ne transitent pas par un réseau mais sont directement envoyées sur le périphérique de sauvegarde connecté à la machine. Cette solution est viable pour un parc de machine à sauvegarder très limité. Elle est généralement peu onéreuse et facile à déployer mais reste peu évolutive et peu performante. A l'inverse des sauvegardes locales, les sauvegardes **distantes** sont effectuées au travers d'un réseau. Le réseau peut être le réseau local de l'entreprise ou, un réseau spécialisé pour le stockage. Depuis quelques temps, certaines entreprises spécialisées proposent à leurs clients des services de **sauvegardes en ligne par Internet**.

Périodicité / fréquence: C'est le nombre de sauvegardes qu'il faut faire pendant une période donnée. La fréquence dépend d'éléments fonctionnels tels que : le volume de données, la vitesse d'évolution des données, la quantité d'informations que l'on accepte de perdre, et éventuellement, la durée légale de conservation de l'information comme par exemple les factures.

Technique de copie des données : On distingue 2 catégories de copie de données selon qu'elles se fassent ou non au travers d'un outil intelligent.

- *Copies simples: dans ce cas les données à sauvegarder sont vues comme de simples séries de bits et donc une série de 0 et de 1.*
 - **Copie de fichiers ou d'arborescence de fichiers** : Il s'agit par exemple d'une simple copie d'arborescence de fichiers
 - **Copie physique** : il d'agit de sauvegarde physique de disque(s) entier(s), bloc par bloc, ou bit par bit.

- *Copies intelligentes : les données peuvent être interprétées et modifiées selon le contexte*
 - **Copie « bare metal »** : C'est une méthode de sauvegarde complète du système à l'aide d'un logiciel spécialisé tel que Norton Ghost de Symantec pour l'environnement Windows. Elle possède la particularité de pouvoir restaurer immédiatement un système et le rendre tout de suite opérationnel. Certains outils permettent de créer ces sauvegardes « bare metal » sur CD ou DVD et de les rendre amorçables (c'est-à-dire capable de lancer le système d'exploitation à l'allumage du système) pour une installation automatique et rapide en cas d'arrêt brutal indésirable.
 - **Copie de système d'exploitation** : elle se fait l'aide d'outils spécifiques mis à la disposition des éditeurs de système d'exploitation. (Par exemple Ignite pour UNIX HP-UX, mksysb pour UNIX AIX)
 - **Copie via l'applicatif** : certaines sauvegardes nécessitent l'intermédiaire d'un outil spécialisé fourni généralement par l'éditeur du logiciel applicatif. C'est le cas par exemple de SAP avec *BACHINT*, pour les bases Oracle avec RMAN. Ces outils s'interfacent entre le logiciel de sauvegarde et le logiciel applicatif .Ils sont plus intelligents, plus efficaces, plus rapides que le logiciel de sauvegarde seul.

Sauvegardes complètes ou incrémentales : Les différentes méthodes de sauvegarde logicielles peuvent être regroupées en 2 groupes, les sauvegardes **totales**, et les sauvegardes **incrémentales**. Elles sont généralement combinées entre elles afin d'assurer des stratégies de sauvegarde optimales.

- **Sauvegarde totale** : elle s'effectue en recopiant l'ensemble du système. L'objectif de la sauvegarde totale (en anglais *full backup*) est de réaliser une copie conforme des données à sauvegarder, sur un support séparé. Néanmoins, pour de gros volumes de données, la sauvegarde complète est longue. Elle peut poser des problèmes de lenteur sur la base elle-même, si les données de l'instance sont modifiées en cours de sauvegarde. En revanche, la restauration est relativement rapide et permet d'obtenir une image fidèle des données à un temps donné.

- **Sauvegarde incrémentale** (en anglais *incremental backup*): elle consiste à sauvegarder les fichiers créés ou modifiés depuis la dernière sauvegarde quel que soit son type (complète, incrémentale). Ce type de sauvegarde est donc plus rapide qu'une sauvegarde totale. Il existe deux types de sauvegardes incrémentales : les différentielles et les cumulatives.
 - **La sauvegarde incrémentale cumulative** : elle ne copie que les fichiers modifiés depuis la dernière sauvegarde totale. L'essentiel des données est donc sauvegardé lors de la 1^{ère} copie complète. Au fil du temps, le volume de sauvegarde augmente puis le nombre de modification depuis la dernière complète, augmente lui aussi. Pour restaurer l'intégralité des données, il faut se munir de la sauvegarde complète et de la dernière sauvegarde incrémentale cumulative.
 - **La sauvegarde incrémentale différentielle** : lors d'une sauvegarde différentielle, seuls les fichiers créés ou modifiés depuis la dernière sauvegarde (totale, cumulative ou différentielle) sont sauvegardés. Le temps de sauvegarde est plus court qu'une sauvegarde complète ou incrémentale cumulative.

Volumétrie des données: C'est le volume physique des données à sauvegarder. Il influe sur le temps de sauvegarde. Il faut aussi connaître la croissance de ce volume pour estimer les besoins futurs.

1.2 INFRASTRUCTURE DE SAUVEGARDE

Pour lancer une tâche de sauvegarde, une infrastructure logicielle et matérielle est nécessaire. Les médias de sauvegarde, les lecteurs et la robotique de sauvegarde constituent la partie matérielle. La partie logicielle est constituée du logiciel de sauvegarde bien sûr mais, aussi du logiciel chargé de lancer les sauvegardes, appelé ordonnanceur.

1.2.1 Le matériel

1.2.1.1 Les supports et lecteurs de bandes

Pour écrire sur un support il faut des périphériques spécialisés appelés des **lecteurs**, chargés d'écrire sur des **supports** de sauvegarde appelés aussi **médias** de sauvegarde.

Pour les petits volumes (un ordinateur personnel par exemple) les CD ou DVD réinscriptibles, les lecteurs Zip, etc. peuvent convenir. Mais, lorsque les volumes sont importants, le support couramment utilisé reste la bande magnétique, appelée aussi cartouche.

Il existe quatre technologies de bandes magnétiques et de lecteurs. Celles-ci sont par ordre de performance :

- **DDS/DAT** : DDS (Digital Data Storage) est le nom officiel de la DAT (Digital Audio Tape), qui stocke jusqu'à 320 Go de données stockées par bande , avec un débit de 6.9 Mo/s.

- **DLT** : bande magnétique offrant jusqu'à 320Go par bande. Débit de 120Mo/s. D'une capacité de 500.000 passages en lecture/écriture, soit une moyenne de 10.000 cycles d'utilisation, sa durée d'archivage est de 30 ans.

- **SDLT** : évolution de la norme DLT, ses bandes magnétiques offrent plusieurs To de stockage à un débit supérieur à 200 Mo/s, sa durée d'archivage est de 30 ans.

- **LTO (Linear Tape Open)** : débit de 270Mo/s. D'une capacité allant jusqu'à plusieurs téraoctets de données stockées sur une seule cartouche, sa durée d'archivage est de 30 ans ou de 1.000.000 de passages. C'est actuellement la technologie de sauvegarde la plus performante, tant en capacité qu'en vitesse de transfert.

Pour une entreprise de petite taille ayant une dizaine de Go à sauvegarder, ou pour une personne exerçant une profession libérale, les systèmes de bandes DDS peuvent convenir car les lecteurs de bande sont économiques et fiables.



Figure 1: Lecteur DAT et cartouches DDS-4

Pour une entreprise de plus grande taille, les systèmes de bandes DLT, SDLT ou LTO sont préférables même s'ils sont plus chers, car leurs performances sont nettement meilleures (meilleure capacité de stockage et meilleur taux de transfert des données).



Figure 2: Cartouche DLT

Le choix d'un type de support se fait en fonction de critères tels que :

- le prix de l'ensemble des médias
- le prix des périphériques de lectures
- la fiabilité
- la capacité totale par média
- la durée de vie des médias

1.2.1.2 Robotique de sauvegarde

Pour écrire ou lire sur médias, il faut des périphériques spécialisés, appelés des **lecteurs** de média. Un lecteur peut être autonome et directement relié à une machine ou au contraire piloté par une **librairie**.

Les librairies sont des périphériques spécialisées dans la sauvegarde sur bande comprenant un ou plusieurs lecteurs. On parle aussi de **robot de sauvegarde**. Les librairies sont capables d'écrire et/ou lire sur plusieurs bandes en même temps (si elles disposent de plusieurs lecteurs de bandes.) Les robots de sauvegarde possèdent un **bras mécanique** qui se charge de déplacer les médias. Lorsqu'une bande est pleine, le robot prend automatique la bande vierge suivante. Un robot de sauvegarde augmente donc la souplesse et l'autonomie d'utilisation.



Figure 3: Bras du robot prenant une bande de sauvegarde

Les robots peuvent ou non disposer de magasins. Sans **magasin**, il faut insérer ou extraire les bandes, une par une, ce qui peut s'avérer long et fastidieux. Un magasin permet d'extraire en une seule opération un groupe de media. Les magasins simplifient la rotation des bandes.

Il existe une gamme très large de robots, du plus petit gérant quelques cartouches, au plus gros capable d'en gérer des centaines.



Figure 4: Exemple de robot de sauvegarde (Storagetek)



Figure 5: Exemple de petit robot de sauvegarde (HP Surestore)

1.2.2 Les logiciels

Nous avons vu les composants matériels, les supports et périphériques tels que les lecteurs et robots de sauvegarde. Nous allons maintenant étudier la couche logicielle utilisée pour le lancement des sauvegardes.

1.2.2.1 Les logiciels de sauvegarde

Pour sauvegarder des données, on ne se contente pas généralement de commande du système d'exploitation, mais on préfère un outil plus intelligent : un logiciel de sauvegarde. C'est un programme spécialisé dans la copie et la récupération des données.

Certains sont uniquement capables de sauvegarder des données d'une seule machine. Mais, les solutions de sauvegarde d'entreprise, et donc celles qui nous intéressent dans ce mémoire, sont des solutions de type **client/serveur** où un seul serveur peut prendre en charge des centaines de clients.

Dans cette configuration, le **client** est la machine qui souhaite réaliser une opération de sauvegarde ou de restauration. Le client communique avec le serveur à l'aide d'un programme particulier appelé **agent**. Ce dernier est chargé de préparer les données et de les transmettre au serveur. Celui-ci se charge alors de toutes les opérations nécessaires à la réalisation de la sauvegarde. Un agent est généralement spécialisé par type de données. Par exemple, il existe un agent spécialisé pour la copie de systèmes de fichier, un autre pour les sauvegardes RMAN, un autre pour la messagerie Exchange, etc. Avant de choisir une solution de sauvegarde il est important de vérifier la compatibilité des agents avec les plateformes (UNIX, WINDOWS, etc.) et les types de données à sauvegarder.

Le logiciel **serveur** de sauvegarde est installé et s'exécute sur une machine dédiée. Il appelle des scripts ou des commandes de sauvegarde et/ou de restauration situés sur les clients. Le logiciel est généralement maintenu et paramétré grâce à une **console d'administration** unique centralisé et accessible à distance. Le serveur répond à toutes les demandes de sauvegarde / restauration soumises par les clients et gère toutes les tâches nécessaires aux sauvegardes et restaurations. Le logiciel de sauvegarde gère les périphériques (lecteurs de bandes et/ou bibliothèques) et les médias de sauvegardes (cartouches magnétiques par exemple). Les logiciels de sauvegarde disposent souvent d'une fonction d'ordonnancement lui permettant de déclencher des sauvegardes automatiques, selon un agenda défini à l'avance. Cependant cette fonction est peu utilisée, car des ordonnanceurs plus performants existent. (cf. plus loin, les ordonnanceurs).

Le logiciel stocke toutes les informations nécessaires à la sauvegarde dans une base de données appelée **référentiel**. Il peut s'agir d'une base de données spécifique à l'application ou une base de données type Oracle ou DB2. Dans le référentiel, on peut trouver de nombreuses informations concernant les sauvegardes : historique des sauvegardes et restauration, information sur les agents déployés, paramétrage, déclaration de composants installés, informations sur les médias et périphériques. Les logiciels de sauvegarde maintiennent constamment à jour ce référentiel. Remarquons qu'un logiciel de sauvegarde est une application comme une autre. Il ne faut donc pas oublier de le sauvegarder ni oublier de sauvegarder son référentiel.

Les logiciels de sauvegarde se distinguent par leurs caractéristiques techniques telles que :

- Le type de systèmes d'exploitation supportés
- Les types de données que l'on peut sauvegarder
- la performance, la fiabilité des techniques de copie
- la possibilité de sauvegarder des données en cours d'utilisation
- la création d'images amorçables
- les méthodes de planification
- la facilité de déploiement et la gestion des agents
- la facilité d'administration et les possibilités d'évolution
- les périphériques supportés (lecteur, robotique)
- etc.

Il existe de très nombreux outils pour la sauvegarde, payants ou non, limités à une seule machine ou au contraire capables de mettre en sécurité un parc de machines.

Parmi les logiciels libres on citera Amanda. Lorsqu'il s'agit de sauvegarde pour les grandes entreprises, le marché se répartit principalement entre:

- Netbackup et BackupExec de SYMANTEC
- Tivoli Storage Manager (TSM) d'IBM
- Networkers (EMC)
- ArcServ de Computer Associates.

Pour illustrer notre propos, la figure suivante compare les parts de marché des principaux outils de sauvegardes.

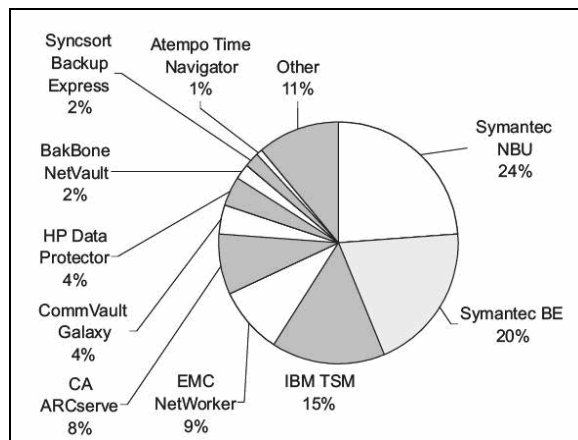


Figure 6: Répartition du marché des logiciels de sauvegarde (2005, source GARTNER)

Le logiciel DataProtector de Hewlett Packard, utilisé dans ce mémoire ne représentait en 2005 que 4% du marché.

Il est difficile de comparer objectivement ces logiciels de sauvegarde, car il s'agit de logiciels complexes à installer et mettre en œuvre, même si de l'avis des spécialistes, TSM semble sortir du lot. On peut trouver dans le livre « *Enterprise Systems backup and recovery, a corporate insurance policy* » de Preston de Guise, un tableau qui permet de recenser et comparer les fonctionnalités pour les solutions de sauvegardes. (Cf. la copie de ce tableau en annexe).

1.2.2.2 Ordonnanceur

Les tâches de sauvegardes ne se lancent pas toute seules. Bien sûr, on pourrait déclencher une copie manuellement. Mais lorsque le nombre de tâches de sauvegarde est important, l'automatisation devient nécessaire, il faut employer un outil particulier appelé **ordonnanceur**. Un ordonnanceur, est un programme (ou éventuellement un sous-programme) qui lance, contrôle et pilote l'exécution de scripts. Ce logiciel permet le lancement automatisé, planifié, contrôlé et centralisé de séries de tâches (appelé aussi des *batches*) sur l'ensemble du système d'information.

Dans certains ordonnanceurs, la mise en place des traitements se réalise dans une interface graphique. Chaque tâche se représente alors sous la forme d'une boîte, qu'il faut paramétrer ensuite. Il faut y indiquer le nom du script, les conditions et l'heure d'exécution, etc. Les boîtes sont ensuite reliées pour y définir des enchaînements, entre des traitements, à une heure définie et/ou à la suite d'autres traitements.

Généralement les logiciels de sauvegarde ont leur propre ordonnanceur. Mais ils sont trop spécialisés et dédiés aux sauvegardes ; ils sont donc limités et ne conviennent pas pour l'exploitation des entreprises. Les entreprises préfèrent lancer leur sauvegarde dans un logiciel tiers indépendant et beaucoup plus puissant.

Il existe de nombreux logiciels spécialisés dans l'ordonnancement des tâches d'exploitation. Les acteurs majeurs du marché de l'ordonnancement sont :

- AUTOSYS de la société Computer Associates
- CONTROL M de la société BMC software
- VTOM de la société ABSSYS
- DOLLAR UNIVERSE de la société ORSYP.

1.2.3 La chaîne de sauvegarde

Une infrastructure de sauvegarde s'appuie généralement sur trois éléments principaux :

- un logiciel de sauvegarde
- un ordonnanceur (intégré ou non au logiciel de sauvegarde)
- une robotique.

Le tout forme une chaîne, que l'on peut schématiser de la façon suivante :

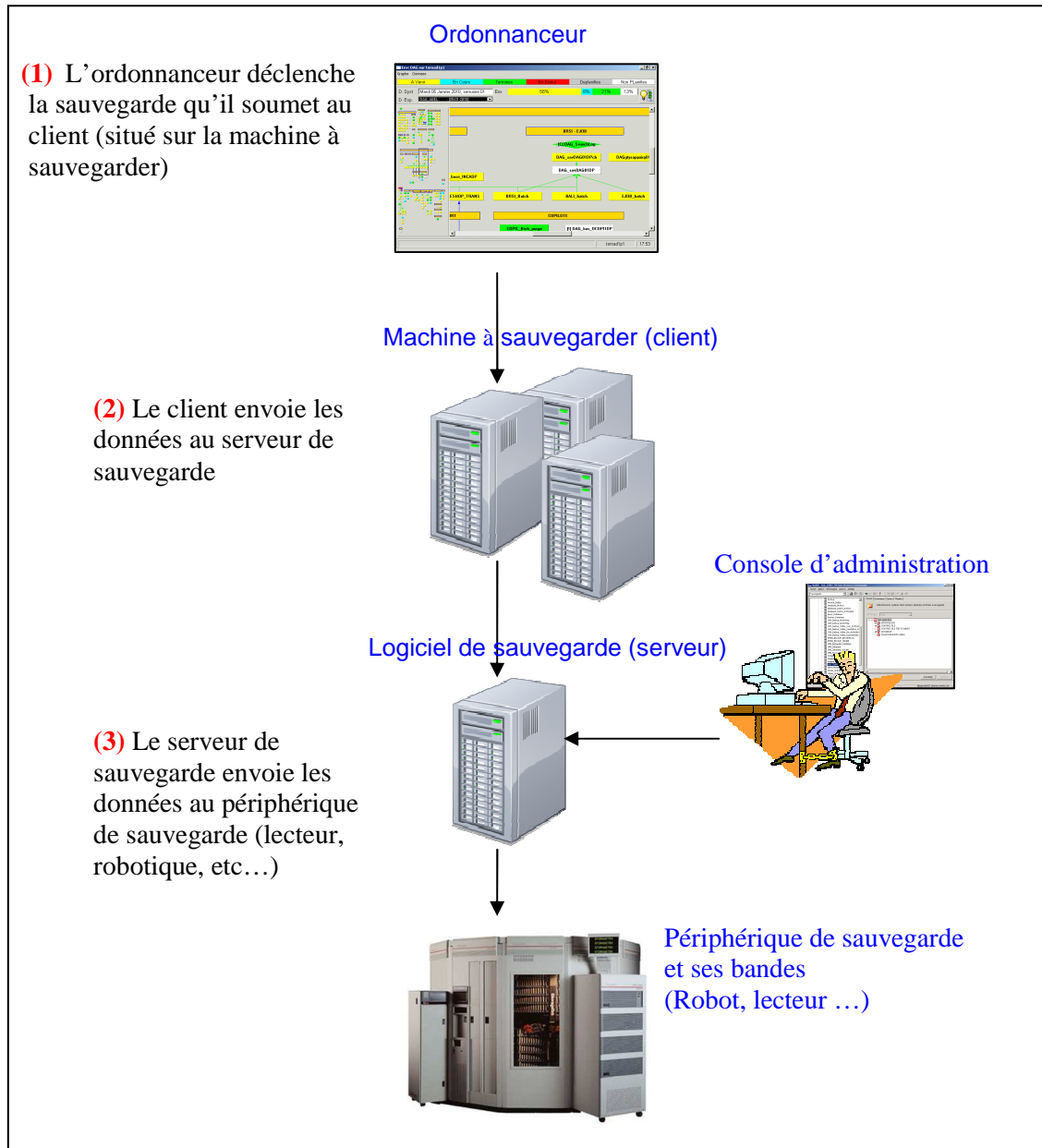


Figure 7: Schématisation d'une infrastructure de sauvegarde

1.3 ORGANISATION

L'infrastructure de sauvegarde ne peut fonctionner de manière aléatoire ou anarchique. Elle doit être organisée.

1.3.1 Politique de sauvegarde

Dans son ouvrage « La sécurité informatique », Jean François CARPENTIER affirme que pour mettre en place une architecture de sauvegarde opérationnelle, il faut inclure une politique de sauvegarde qui doit définir clairement :

- les ressources (humaines, matérielles et logicielles)
- les processus (de sauvegarde, de gestion des supports, de restauration)
- les technologies (infrastructure et moyens techniques).

Selon lui, la planification se déroule en 3 parties :

- la définition des contraintes
- la maîtrise des facteurs qui influencent la sauvegarde
- la préparation de la stratégie décrivant sa mise en œuvre.

La définition des contraintes tient compte de contraintes fonctionnelles, techniques et financières telles que :

- l'indisponibilité des données
- le type et la criticité des données
- le temps de restauration maximum
- la perte de données tolérées
- le mode et durée de conservation des médias
- le volume et la croissance des données à sauvegarder
- la fréquence et le temps de sauvegarde nécessaire
- le coût.

Les facteurs qui influencent la stratégie de sauvegarde sont :

- la politique de sécurité et les besoins en matière de sauvegarde et de stockage
- la configuration du réseau
- les ressources de l'entreprise (environnementales, humaines, matérielles et logicielles).

La politique de sauvegarde doit déterminer :

- *Quoi sauvegarder* : le périmètre, le type et la criticité des données ...
- *Quand* : périodicité, date, heure et fréquence de sauvegarde
- *Comment*: c'est-à-dire la manière :
 - de sauvegarder (mise en œuvre dans l'ordonnanceur)
 - de gérer et stocker les médias
 - de restaurer les données.
- *Par qui* : le rôle et la responsabilité de tous les acteurs
- *Avec quoi* : la description du système, et l'architecture des moyens de sauvegardes
- *Sur quoi* : le type de support
- *Où* : la description des locaux et des moyens de stockage des supports.

Pour qu'une politique de sauvegarde soit efficace, il est nécessaire de suivre quelques règles et bonnes pratiques qui consistent à :

- impliquer toutes les personnes qui participent de près ou de loin aux sauvegardes
- documenter et décrire toute l'organisation de sauvegarde comprenant l'architecture, les moyens, les rôles et responsabilités, et le processus de réalisation.
- effectuer des tests de restaurations et des supports régulièrement
- s'assurer que les supports sont transportés selon un cahier des charge puis entreposés dans un endroit séparé physiquement des salles informatiques, protégées des risques d'accidents et d'intrusions.

- Si les supports sont gérés par une société spécialisée dans le stockage des supports, il est nécessaire d'obtenir des garanties contractuelles en matière d'intégrité et de confidentialité des informations confiées.

1.3.2 Les acteurs de la sauvegarde

La protection des données n'est pas seulement un ensemble de machines et de logiciels, c'est aussi, et surtout, une affaire d'hommes et de femmes. Il est important de comprendre que les sauvegardes impactent de très nombreuses personnes. Certains d'entre eux n'en ont parfois même pas conscience. Les individus sont souvent le talon d'Achille de la sécurité informatique dans l'entreprise. C'est pourquoi, il faut les sensibiliser, les former, et les entraîner.

Les acteurs, leurs missions et responsabilités peuvent changer suivant la taille et l'organisation de l'entreprise, mais on peut classer ces acteurs en plusieurs groupes :

- les équipes spécialistes des sauvegardes
- les exploitants (chargé de la maintenance du système d'information : administrateurs, pilote, analyste d'exploitation, etc.)
- le management
- les utilisateurs.

Passons en revue, ces différents groupes.

1.3.2.1 Les spécialistes des sauvegardes

- ***Les opérateurs de sauvegardes*** réalisent le suivi des sauvegardes (au regard de la politique de sauvegardes). Ils gèrent les médias des sauvegardes, généralement des bandes. Ils interviennent en cas de problème relativement simple. (Relance de sauvegardes si nécessaire, etc.). Ils lancent aussi, à la demande, des sauvegardes et des restaurations.

- **Les administrateurs du logiciel de sauvegarde** sont chargés d'administrer les logiciels spécialisés de sauvegardes : installation, paramétrage et maintenance des logiciels de sauvegardes. Ils assistent les opérateurs de sauvegardes et sont capables de réaliser toutes les activités des opérateurs de sauvegardes
- **Les experts de sauvegardes** ont une connaissance approfondie de l'infrastructure de sauvegarde, tant logicielle que matérielle. Ils assistent les administrateurs de sauvegardes. Ils conçoivent les architectures de sauvegarde et étudient les évolutions et améliorations de l'outil de sauvegarde

1.3.2.2 Les exploitants

- **Les pilotes d'exploitation** : Ils sont chargés de suivre l'ensemble du système d'exploitation à l'aide d'outils de supervision. Lorsqu'un problème est détecté ils agissent en fonction de consignes.
- **Les techniciens d'exploitation** sont chargés d'ordonnancer et de suivre les traitements et donc aussi les tâches de sauvegardes. Ils interviennent en cas de problème.
- **Les administrateurs systèmes**: Ils sont chargés de l'installation, le paramétrage et la maintenance des systèmes d'exploitation tels que Windows, UNIX, ou autres. Ils interviennent pour donner des consignes de sauvegardes. Ils ont un rôle de conseil et participent aux tests de restaurations des systèmes d'exploitation.
- **Les administrateurs d'application** sont les spécialistes chargés d'installer et maintenir des applications complexes, telles que les bases de données (Oracle, SQL server, Sybase DB2, etc.), les logiciels ERP (SAP, etc.), la messagerie (Lotus Notes, Exchange, etc.). Ils conseillent et donnent les informations nécessaires aux sauvegardes et restaurations des applications dont ils ont la charge. Ils participent aux tests de restaurations

1.3.2.3 Le management

Comme nous l'avons vu, les informations stockées dans le système d'information sont une richesse immatérielle de l'entreprise. Il est du devoir du chef d'entreprise de manager et sécuriser ce patrimoine informationnel.

Les entreprises sont dans l'obligation de se soumettre aux lois et réglementations afférentes. En France, on citera la loi de 2003 sur la sécurité numérique. Pour l'étranger, on citera la loi américaine dite « Sarbanes Oxley Act ». Ces deux lois engagent la responsabilité pénale du dirigeant en cas de perte ou altération de données.

Le dirigeant doit informer son personnel des menaces réelles et de prendre les précautions nécessaires. Il délègue alors sa responsabilité qui se transmet tout au long de la chaîne de commandement. Les autres dirigeants, du plus haut niveau, aux chefs d'équipe en passant par le management intermédiaire peuvent aussi à leur tour, être concernés.

1.3.2.4 Les utilisateurs

- **Les utilisateurs clés** : ils font le lien entre les utilisateurs des applications et les différents administrateurs (d'application, de systèmes et sauvegardes). Ils déterminent les contraintes des utilisateurs sur les sauvegardes, s'il est possible d'arrêter ou non une application pendant une sauvegarde.
- **Les utilisateurs finaux** : Ce sont les consommateurs des applications mises à leur disposition. Ils sont informés des contraintes liées aux sauvegardes.

1.4 RESUME DU CHAPITRE 1

Les données de l'entreprise sont une richesse qu'il faut protéger contre les risques : pannes, accidents naturels, erreurs humaines, malveillances ... Pour se prémunir, on réalise des sauvegardes, c'est à dire des copies de données sur un support. Ces supports doivent être entreposés dans un endroit séparé physiquement des salles informatiques et être protégés des risques d'accidents et d'intrusions.

Une sauvegarde possède des caractéristiques fonctionnelles telles que le type de données, le temps de sauvegarde maximum toléré, les pertes acceptées, etc.

Une sauvegarde a aussi des caractéristiques techniques selon la technique de copie de données, le temps de sauvegarde, la périodicité... La méthode de copie la plus simple est la sauvegarde complète ou totale, qui consiste à copier tous les fichiers à sauvegarder qu'ils soient récents, anciens, modifiés ou non. Cette méthode considérée comme fiable, est longue et très coûteuse en temps et en volumétrie. Afin de gagner en temps de sauvegarde, il existe des méthodes qui procèdent à la sauvegarde des seules données modifiées et/ou ajoutées entre deux sauvegardes totales. Il s'agit des sauvegardes incrémentales (sauvegarde des seuls fichiers ayant été modifiés depuis la dernière sauvegarde complète ou incrémentale) et différentielles (sauvegarde des seuls fichiers ayant été modifiés depuis la dernière sauvegarde complète).

Pour lancer une tâche de sauvegarde, une infrastructure logicielle et matérielle est nécessaire. Les médias de sauvegarde, les lecteurs et la robotique de sauvegarde constituent la partie matérielle. La partie logicielle est constituée du logiciel de sauvegarde bien sûr mais aussi d'un logiciel chargé du lancement des tâches de sauvegarde appelé ordonnanceur.

L'infrastructure de sauvegarde ne peut fonctionner de manière aléatoire ou anarchique. Elle est organisée pour l'application d'une politique de sauvegarde qui définit clairement :

- les ressources (humaines, matérielles et logicielles)
- les processus (de sauvegarde, de gestion des supports, de restauration)
- les technologies (infrastructure et moyens techniques).

La protection des données ne se résume pas seulement un ensemble de machines et de logiciels, c'est aussi et surtout une affaire d'hommes et femmes. Il est important de comprendre que les sauvegardes impactent de très nombreuses personnes qu'il faut sensibiliser, former, et entraîner.

Chapitre 2: SAUVEGARDES DE BASES ORACLE

Dans le chapitre précédent, nous n'avons abordé que des généralités sur les sauvegardes. Ce chapitre se concentre sur des sauvegardes bien particulières : les sauvegardes de bases Oracle.

Or, avant de sauvegarder une base, il est absolument nécessaire de savoir comment elle fonctionne. C'est pourquoi, nous commencerons par décrire l'architecture d'ensemble d'une base de données. Nous ferons ensuite un tour d'horizon sur les différentes méthodes de sauvegardes d'une base de données Oracle.

2.1 PRINCIPES DE FONCTIONNEMENT D'UNE BASE ORACLE

2.1.1 Instance et base de données

2.1.1.1 Rappels

Une **base de données** est une collection d'informations stockées dans un système informatique.

Les informations d'une base de données sont manipulées au moyen d'un logiciel, le **S**ystème de **G**estion de **B**ase de **D**onnées (**SGBD**) et avec un langage particulier: le **SQL** (**S**tructured **Q**uery **L**angage) .

Oracle est un SGBD, de l'éditeur de logiciel qui porte le même nom (Oracle). Une base de données Oracle est une base de données créée et manipulée avec le SGBD Oracle.

2.1.1.2 Instance

Les informations d'une base de données Oracle sont stockées dans des fichiers. Chaque base de données Oracle est associée à une **instance** composée de processus et de mémoire. Pour permettre l'utilisation des données d'une base et les mettre à la disposition des utilisateurs, il faut préalablement ouvrir la base, c'est-à-dire allouer de la mémoire et démarrer des processus de l'instance. Une fois que l'instance est ouverte, le SGBD Oracle l'associe à la base de données.

2.1.1.3 Accès aux données

Les programmes utilisateurs de la base de données sont appelés des **clients**. Ces derniers dialoguent avec la base de données, à distance, via le réseau informatique en envoyant des requêtes SQL (Commandes de la forme « *insert* », « *delete* », « *update* », « *select* »...). Ces clients communiquent avec la base de données, par l'intermédiaire de l'instance. Les processus de l'instance sont créés par le SGBD Oracle sur le serveur qui héberge la base, lorsque la base est ouverte aux requêtes. En fonction des requêtes, les processus de l'instance lisent et écrivent des informations sur les fichiers de la base.

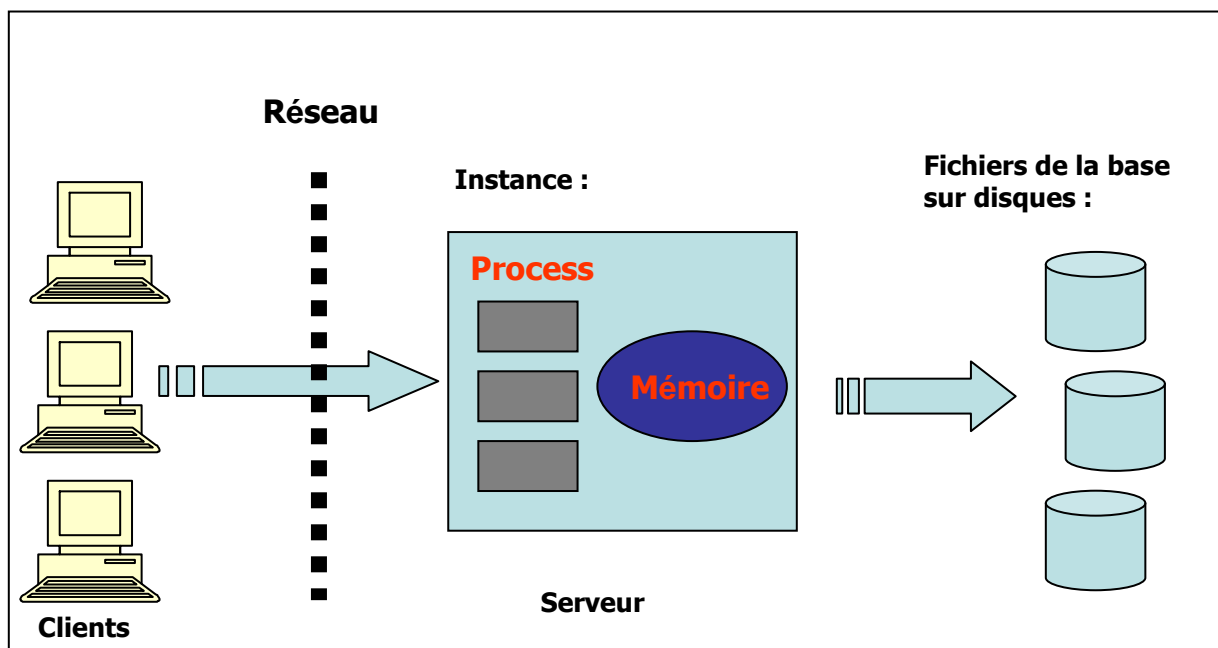


Figure 8: Architecture simplifiée d'une base Oracle

2.1.2 Fichiers d'une base de données

Nous l'avons vu, les informations d'une base Oracle sont stockées dans des fichiers. Il existe plusieurs types de fichiers. Une base de données Oracle est constituée des fichiers principaux suivants:

- « *Datafiles* » : contiennent les données de la base (stockées sous forme de tables)
- « *Redolog* » : ces sont les journaux de transactions qui contiennent l'historique des instructions effectuées sur la base
- « *Controlfiles* » : contient les données sensibles nécessaires au fonctionnement de la base. (liste des fichiers avec leur localisation, numéro de la dernière action sur la base...)

2.1.3 Fichiers « Redolog » et mode archivage

Les journaux de transactions (Appelés aussi fichiers Redolog ou redo) ont une importance particulière dans le mécanisme de sauvegarde et restauration. C'est pourquoi nous leur accordons un éclairage particulier dans ce paragraphe.

Lorsqu'une transaction vient modifier une base, les blocs de données ne sont pas immédiatement mis à jour dans ses fichiers de données. En revanche, les fichiers Redolog enregistrent, au fil de l'eau, l'historique des transactions passées (et non les données) dans la base de données.

En cas d'arrêt brutal et non prévu de la base, les informations contenues dans les fichiers de Redo, permettent de rejouer les transactions perdues.

Les fichiers de Redo sont aussi utilisés pendant les restaurations de base de données. Grâce aux informations stockées dans les fichiers redolog, la base de données Oracle est capable de rejouer l'ensemble des commandes n'ayant pas été sauvegardées pour récupérer le contenu de la base de données. Insistons sur le fait que les redolog ne contiennent pas les données modifiées, mais uniquement les commandes passées qui modifient les données. La base de

données dispose de plusieurs Redolog: au moins deux. Ils sont remplis séquentiellement et de manière circulaire. Lorsqu'un fichier Redolog est plein ou lors d'évènements particuliers, on passe au Redolog suivant (On parle de bascule ou de « Switch » de Redolog) , et lorsque le dernier est atteint, on repasse au premier.

Après un cycle complet de « Switch », le contenu est perdu. C'est pourquoi, il est possible d'activer un **mode archivage** qui recopie les fichiers Redologs avant leurs réutilisations. Ces fichiers d'archives sont importants car nécessaires aux restaurations. Précisions que le mode archivage doit être déclenché et paramétré avant la sauvegarde, si l'on veut pouvoir récupérer toutes les modifications de la base.

2.2 METHODES DE SAUVEGARDES DES BASES ORACLE

Nous avons vu la structure des bases de données Oracle. Nous allons voir comment les sauvegarder.

Il existe 3 moyens de sauvegarde :

- les sauvegardes physiques de fichiers de la base
- les sauvegardes logiques (exports)
- les sauvegardes avec l'outil RMAN.

2.2.1 Sauvegardes physiques

Les sauvegardes physiques d'une base de données Oracle consistent à recopier sur média les fichiers qui composent la base.

Il existe deux modes de sauvegardes physiques :

- **Les sauvegardes à froid** : Les sauvegardes des fichiers ont lieu alors que la base est fermée (pas d'accès à la base, et processus d'arrière plan stoppés). Dans ce cas il suffit de stopper la base et de copier simplement l'ensemble des fichiers sur bande.

Attention, il faut recopier l'ensemble des fichiers. Cette sauvegarde implique une perte de toutes les modifications de la base postérieures à la sauvegarde. La restauration est simple, il suffit, à partir des bandes de recopier les fichiers sur disques puis d'ouvrir la base.

- **Les sauvegardes à chaud** : la base est ouverte, c'est-à-dire que la base est utilisée par les clients et qu'elle est accédée par les processus de l'instance. Elle nécessite le mode archivage (cf. plus haut). Pour réaliser une sauvegarde dans ce mode, il faut mettre chaque fichier dans un état spécial appelé à l'aide d'une requête spéciale (commande SQL « *begin backup* »), avant de les copier sur bandes. En fin de sauvegarde il est nécessaire d'enlever ce mode « *begin backup* ». Il faut ensuite sauvegarder les archives générées. Dans cette configuration, la restauration consiste à récupérer à partir des bandes, l'ensemble de tous les fichiers sur disque, puis à appliquer toutes les transactions contenues dans les fichiers d'archivage.

2.2.2 Sauvegardes logiques (exports)

Auparavant, nous avons parlé de sauvegarde de fichiers. Il s'agit dans ce cas de sauvegardes appelée sauvegardes physiques. Ici nous allons parler de **sauvegardes logiques**.

En réalité le terme « sauvegarde » est un abus de langage et Oracle appelle ces sauvegardes logiques des **exports**, ou des **dump**. Un export (ou dump) est un fichier généré par un utilitaire Oracle tel que « **exp** » (ancienne version de l'utilitaire d'export) ou « **expdp** » (nouvel outil d'export de base appelé datapump), et ne fonctionne qu'avec des bases oracle. Le format d'un export est propriétaire. Un tel fichier contient les données, mais également la définition des objets de tout ou partie de la base. Pour lancer un export il est nécessaire que la base soit ouverte. On peut exporter une base entière, un objet de la base (table par exemple), un ensemble logique d'objets (objets appartenant à un propriétaire par exemple).

Les exports sont très utiles puisqu'ils ont l'avantage d'être indépendants de la plateforme et de son système d'exploitation. Ils sont donc très utilisés pour les migrations de bases entre plateformes différentes, par exemple pour déplacer une base d'un serveur Windows vers un

serveur UNIX. Pour importer les informations contenues dans ce fichier dans une base de données, ou utilise la commande « **imp** » (si l'export a été généré avec « exp ») ou « **impdp** » (si l'export a été réalisé avec « expdb »).

Les exports ne remplacent pas les sauvegardes physiques de bases. Ils sont complémentaires et très utiles. Ils permettent de restaurer facilement une table, ou un ensemble logique d'une base. Ces récupérations « logiques » seraient difficiles avec d'autres méthodes.

Quelques remarques : les exports sont générés sur disques. Il faut généralement les copier sur bandes à l'aide du logiciel de sauvegarde.

2.2.3 Sauvegardes avec l'outil RMAN

Nous avons vu qu'il était possible de faire des sauvegardes des fichiers de la base et aussi comment il était possible de générer des fichiers d'export pour réaliser des sauvegarde de données logiques. L'éditeur Oracle offre un troisième moyen de faire des sauvegardes : l'utilitaire RMAN.

L'outil RMAN (**R**ecover **MAN**ager) est livré en standard avec le moteur de la base de données Oracle. Par conséquent, il ne nécessite pas d'installation particulière, ni de supplément de licence. (Attention, l'agent permettant la communication avec un outil de sauvegarde tiers est quant à lui payant). C'est un outil spécialisé pour des sauvegardes et restaurations aisées, fiables et optimisées. Il ne fonctionne qu'avec les bases Oracle.

Concrètement, RMAN se présente sous la forme d'un programme exécutable qui se lance en ligne de commande comme le montre l'écran ci-dessous. L'invite de commande ouverte, il suffit de taper les commandes de configuration et/ou de sauvegarde / restauration.

Notons que RMAN ne sait pas écrire tout seul sur une bande. Il faut donc interfacer RMAN avec un logiciel de sauvegarde, qui lui est capable de dialoguer avec les lecteurs de média.

```
gtaxdetvm32.pharma.aventis.com - PuTTY
gtaxdetvm32:oracle#
gtaxdetvm32:oracle#
gtaxdetvm32:oracle# rman target / catalog urman/urman@RMANTST

Recovery Manager: Release 11.1.0.7.0 - Production on Wed Jan 6 12:07:28 2010

Copyright (c) 1982, 2007, Oracle. All rights reserved.

connected to target database: ORCL02 (DBID=3376000402)
connected to recovery catalog database

RMAN>
```

Figure 9: Exemple de lancement de RMAN à l'aide de l'invite de commande UNIX

RMAN offre de nombreuses possibilités et avantages :

- il permet de sauvegarder les bases à chaud et à froid,
- il permet de sauvegarder des bases entières ou des fichiers,
- il permet des sauvegardes incrémentales et différentielles,
- il réduit les volumes de sauvegardes en évitant les blocs vides,
- il s'interface avec de nombreux logiciels de sauvegarde,
- il permet de faire des sauvegardes sur disques et banques,
- il est indépendant du système d'exploitation. (Un script RMAN fonctionne quelle que soit la plateforme),
- il contrôle les sauvegardes et restaurations (vérification des blocs corrompus),
- il est capable de simuler des restaurations,
- Il sait restaurer des blocs de données corrompus à chaud.

A l'opposé des sauvegardes physiques qui réalisent une copie des fichiers de la base, RMAN opère au niveau des blocs de données. RMAN réalise une sauvegarde intelligente : il ne copie pas les blocs vides ou inutiles ; il est capable d'exclure de la sauvegarde les blocs non modifiés depuis la dernière sauvegarde, ce qui réduit le temps et le volume de sauvegarde.

Notons que RMAN ne sait pas écrire directement sur bande. Il lui faut un logiciel de sauvegarde tel que DataProtector (**DP**) de la société HP, ou Tivoli Storage Manager (**TSM**) de la société IBM. Nous verrons plus en détail, dans un prochain chapitre, l'architecture de l'outil RMAN.

Sauvegardes totales et partielles sous RMAN

Avec RMAN il est possible de faire des sauvegardes totales ou partielles de la base :

Les sauvegardes totales consistent à copier toutes les données que celles-ci soient récentes, anciennes, modifiées ou non. RMAN exclut de sa sauvegarde les blocs vides. Pour réaliser sous RMAN une sauvegarde totale, il suffit d'entrer dans l'invite RMAN, une commande du type « backup database ».

Les sauvegardes partielles ne sauvegardent qu'un sous ensemble de la base, telles que les sauvegardes de fichiers, ou de tablespaces. Il convient alors d'utiliser les syntaxes suivantes : « *backup datafile....* » ou « *backup tablespace ...* ».

Sauvegardes incrémentales sous RMAN

Une **sauvegarde incrémentale** nécessite une sauvegarde de référence dite de **niveau 0** (appelée aussi « **baseline** »). Il s'agit d'une sauvegarde dite **totale** qui contient tous les blocs utilisés de la base. Cette sauvegarde sert de référence aux autres sauvegardes de niveau 1 qui ne comprendront que les blocs modifiés lors d'une des sauvegardes précédentes. Sous RMAN, pour réaliser une sauvegarde de niveau 0, il suffit de taper une commande du type: « *backup incremental level 0 database* ».

Les sauvegardes incrémentales qui suivent une sauvegarde de niveau 0, sont dites de niveau 1.

Les deux types de sauvegardes incrémentales sous RMAN sont :

- **Les sauvegardes incrémentales différentielles** qui sauvegardent tous les blocs modifiés depuis la dernière sauvegarde (complète ou incrémentale.) Pour effectuer une sauvegarde différentielle de niveau 1, il suffit de taper dans l'invite de commande RMAN: « *backup incremental level 1 database* ».

Dans la figure qui suit, des sauvegardes incrémentales différentielles sont réalisées, chaque dimanche, une sauvegarde de niveau 0 (totale) est réalisée. Le reste de la semaine on ne sauvegarde que les blocs de données modifiés depuis la veille.

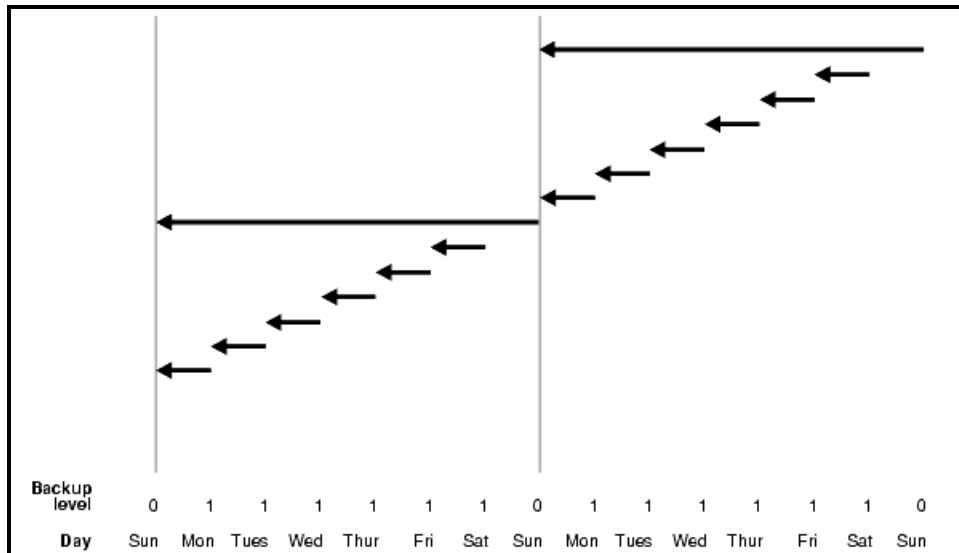


Figure 10: Exemple d'utilisation de sauvegardes incrémentales de type différentiel

- **Les sauvegardes incrémentales cumulatives** qui consistent à sauvegarder tous les blocs modifiés depuis la sauvegarde complète. Pour effectuer une sauvegarde incrémentale cumulative de niveau 1, il suffit de taper dans l'invite de commande RMAN: « *backup incremental level 1 cumulative database* ».

Dans la figure qui suit, des sauvegardes incrémentales cumulatives sont réalisées : chaque dimanche, une sauvegarde de niveau 0 (totale) est lancée. Le reste de la semaine, on ne sauvegarde que les blocs de données modifiés depuis la dernière sauvegarde de niveau 0, donc le dernier dimanche.

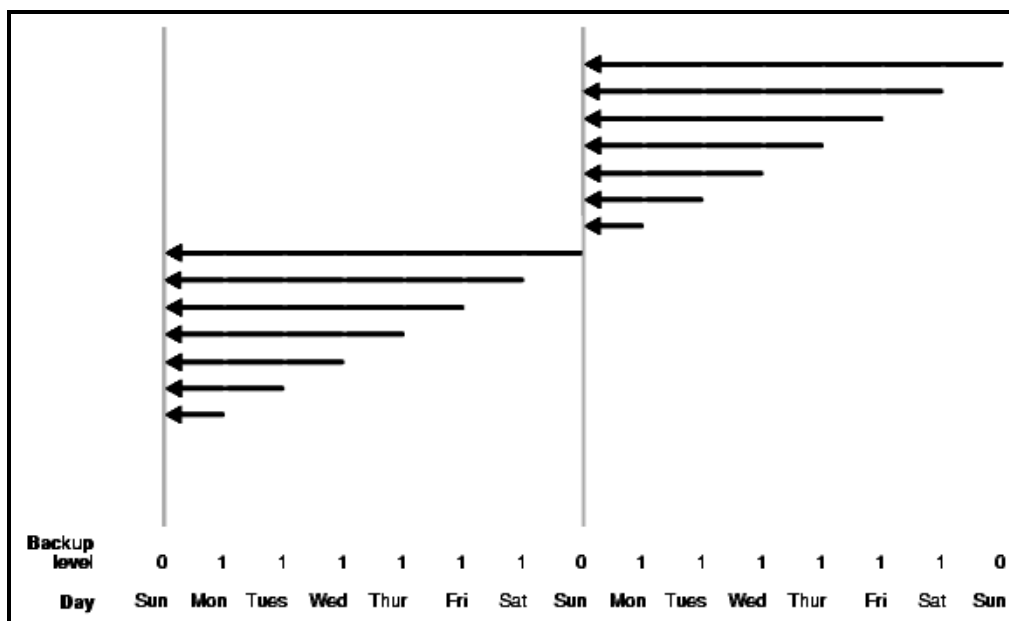


Figure 11: Exemple d'utilisation de sauvegardes incrémentales de type cumulatif

Sauvegardes à chaud et froid sous RMAN

Qu'il s'agisse d'une sauvegarde complète ou incrémentale, avec RMAN il est possible de réaliser des sauvegardes à chaud ou à froid :

- **A chaud**, la base est accédée par les utilisateurs. (On remarquera qu'il n'est pas nécessaire de passer les fichiers de la base dans le mode « *begin backup* », comme c'était le cas dans l'ancien mode de sauvegarde). Le **mode d'archivage** est alors obligatoire.
- **A froid**, la base n'est pas accessible par les utilisateurs. On remarquera, qu'à froid, dans l'ancien mode de sauvegarde sans RMAN l'instance était complètement stoppée, et aucun processus de l'instance ne tournait. Avec RMAN, une sauvegarde à froid entraîne des processus qui accèdent à la base, mais dans un mode particulier, le mode « **mount** » qui autorise la lecture/écriture des informations de sauvegarde dans les fichiers de contrôle.

Le temps et le volume de sauvegarde peuvent être considérablement allégés avec des sauvegardes incrémentales.

En revanche, la restauration peut s'avérer bien plus longue et complexe : elle nécessite plusieurs sauvegardes, la sauvegarde de niveau 0 (complète) et une ou plusieurs sauvegardes incrémentales. Bien sûr, il faut aussi penser à appliquer les fichiers d'archivage pour compléter le processus de restauration, si nécessaire.

La sauvegarde différentielle est plus rapide que la sauvegarde cumulative. Cependant, la restauration à partir de sauvegardes différentielles est plus longue qu'une restauration à partir de sauvegardes cumulatives.

2.3 RESUME DU CHAPITRE 2

Ce chapitre aborde le fonctionnement des bases Oracle et les divers moyens de les sauvegarder

Les informations d'une base de données Oracle (tables et autres objets) sont stockées dans trois types de fichiers : des fichiers de données qui contiennent les informations telles que les tables, des fichiers Redo qui contiennent l'historique des instructions effectuées sur la base, des fichiers de contrôle qui contiennent les données sensibles nécessaires au fonctionnement de la base. (Liste des fichiers avec leur localisation, numéro de dernière action ...). Chaque base de données Oracle est associée à une instance c'est à dire un ensemble de processus et de la mémoire.

Il existe trois moyens de sauvegarder une base Oracle:

- La copie de fichiers, à froid (base arrêtée) ou à chaud (base en fonctionnement) consiste à dupliquer sur bande l'ensemble des fichiers de la base (fichier de données, redo, fichiers de contrôle), un à un. A chaud, il faut lancer une commande SQL spéciale pour mettre le fichier dans un mode particulier appelé « begin backup ».
- La génération d'un fichier d'export de la base que l'on recopie ensuite sur bande ou sur tout autre média.
- L'utilisation de l'outil RMAN.

RMAN est l'outil spécialisé dans les sauvegardes et restaurations de bases Oracle aisées, fiables et optimisées. Il se présente sous la forme d'un programme exécutable qui se lance en ligne de commande. RMAN ne sait pas écrire tout seul sur bande, c'est pourquoi il faut l'interfacer à un logiciel de sauvegarde. Un logiciel de sauvegarde (DataPtotector, Netbackup, TSM...), sans RMAN réalise ses sauvegardes au niveau des fichiers des bases. RMAN permet au logiciel de sauvegarde d'opérer au niveau des blocs de données. Avec RMAN, le logiciel de sauvegarde devient capable d'exclure les blocs de données vides, inutiles ou inchangés. Il sait alors réaliser de « vraies » sauvegardes incrémentales différentielles ou cumulatives.

Chapitre 3 : PRESENTATION DU SUJET

Ce chapitre présente le projet, objet de ce mémoire.

Nous commencerons par décrire l'entreprise et l'équipe. Nous décrirons succinctement le contexte dans lequel s'est déroulé le projet.

Nous découvrirons la problématique du projet pour comprendre pourquoi il est judicieux de faire évoluer la méthode de sauvegardes actuelles vers RMAN. Nous déterminerons les objectifs (mettre en place une solution de sauvegarde basée sur RMAN) ainsi que les contraintes du projet (réglementation pharmaceutique, réglementation financière, contraintes techniques, respect des processus...).

3.1 PRESENTATION DE L'ENTREPRISE ET DU SERVICE

3.1.1 L'entreprise

Sanofi-Aventis est le troisième groupe pharmaceutique mondial, présent, dans plus de 100 pays. Il compte aujourd'hui plus de 105 000 collaborateurs dont 30 000 en France.

Sanofi-Aventis commercialise dans le monde entier une large gamme de médicaments issus de sa recherche. Sanofi-Aventis est un géant qui dispose d'un des plus importants portefeuilles de molécules du secteur, concentré sur sept axes thérapeutiques principaux : cardiovasculaire, thrombose, système nerveux central, oncologie, maladies métaboliques, médecine interne et vaccins.

Sanofi-Aventis est le fruit d'une série de fusions et d'acquisitions avec en particulier la « méga » acquisition par le groupe français Sanofi-Synthélabo du groupe franco-allemand Aventis.

Avec un chiffre d'affaire d'environ 28 milliards d'euros, Sanofi-Aventis est le numéro 1 en France, numéro 1 en Europe et le numéro 5 mondial de l'industrie pharmaceutique pour 2009, après Pfizer - Wyeth, Merck & Co.- Schering-Plough, Roche-Genentech et Novartis.

3.1.2 Le département administration des bases de données

L'entreprise est divisée en 4 entités principales: la Recherche et le Développement, la Production, la Distribution et les Fonctions Centrales. Cette dernière rend services aux trois autres entités.

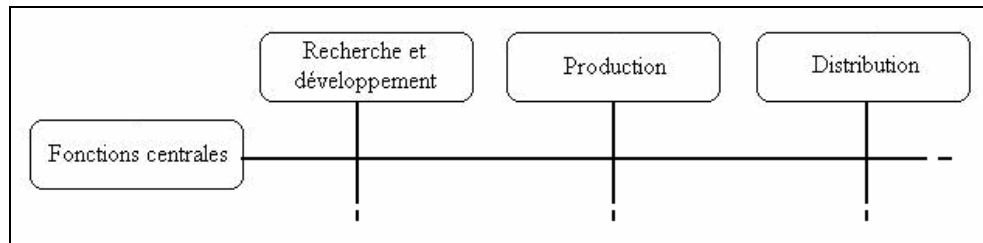


Figure 12: Organisation fonctionnelle de Sanofi-Aventis

Le Système d'Information appartient aux fonctions centrales. Il est divisé en plusieurs entités dont les Services d'Infrastructure Globaux (SIG) qui sont répartis dans plusieurs pays (France, Allemagne, Amérique du Nord et Japon). SIG assure la continuité de services et des opérations.

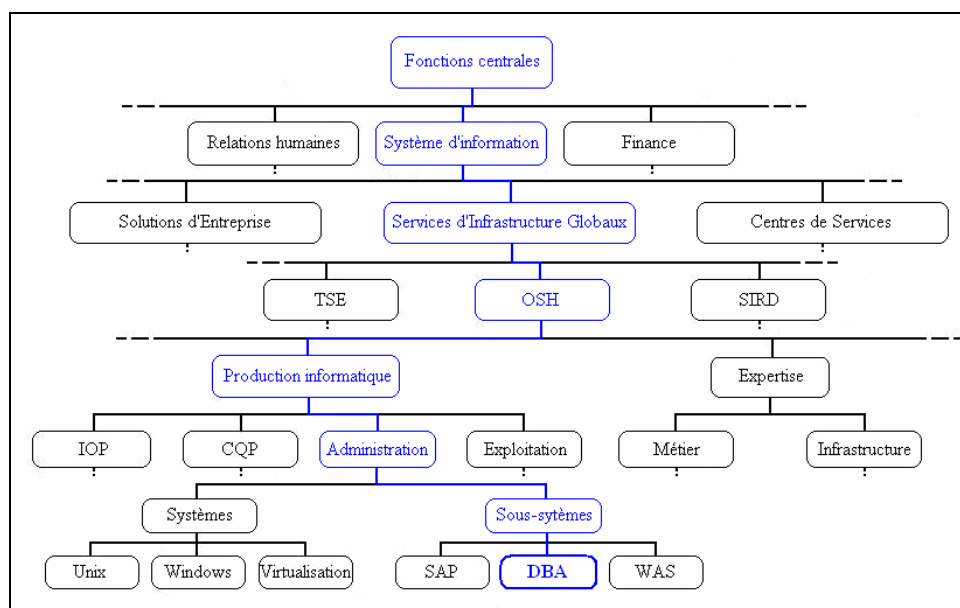


Figure 13: L'équipe DBA chez Sanofi-Aventis

C'est dans les Services d'Infrastructure Globaux que se trouve le département OSH (Opérations et Service d'Entreprise) divisé en deux groupes la Production Informatique (PI) et l'Expertise, basés tous les deux sur le site de Gentilly en France, aux portes de Paris.

En tant qu'adjoint au responsable de l'équipe d'administration de bases de données, je travaille au sein du département de la production informatique. Cette dernière a pour vocation, l'étude, l'installation, le suivi et l'exploitation du système d'information pour le groupe. J'y exerce la fonction d'administrateur de bases de données depuis septembre 2002.

Au cœur de cette entité, j'appartiens au pôle bases de données, dans une équipe de 10 personnes, dont 3 internes et 7 prestataires. Il est dirigé par un chef d'équipe qui rend compte au responsable de la production informatique. Notre équipe est chargée de l'administration de bases de données Oracle et Microsoft SQL Server (sauf les bases du progiciel SAP) : installation des moteurs SGBD, création d'instance, optimisation, mise en production

3.2 L'EXISTANT : LES SAUVEGARDES A SANOFI

3.2.1 Infrastructure au commencement du projet

Chez Sanofi-Aventis l'infrastructure de sauvegarde est composée des éléments suivants :

- Visual TOM ordonnanceur de la société Absys (VTOM) qui lance les tâches de sauvegardes selon un planning et des règles définies à l'avance.
- Le logiciel de sauvegarde HP DataProtector. (De la société Hewlett Packard). Ce dernier lit les fichiers à sauvegarder, et les écrits sur bande par l'intermédiaire d'un robot de sauvegarde.
- Une robotique.

L'existant sera étudié plus en détail au chapitre 4.

3.2.2 Le matériel mis en oeuvre

Les sauvegardes sont réalisées par deux robots de la marque Storagetek (de type 9310), situés sur deux sites différents et disposant de 5200 bandes de sauvegarde et d'une trentaine de lecteurs de bandes chacun.



Figure 14: Extérieur robot de sauvegarde 9310



Figure 15: Robot prenant une bande de sauvegarde

3.2.3 Eléments de la politique de sauvegarde

Les bases sont généralement sauvegardées à chaud ou à froid une fois par jour sur bandes. Celles-ci sont externalisées selon un cahier des charges précis, et mises en lieu sûr tous les jours.

3.3.3.1 Stratégies de sauvegarde et ordonnancement

Deux stratégies de sauvegarde de bases Oracle sont déployées :

- sauvegarde à froid (base arrêtée) chaque jour.
- sauvegarde à chaud (base en fonctionnement) en semaine, à froid (base arrêtée) le week-end.

Généralement, les sauvegardes de fichiers de la base sont précédées d'exports.

3.3.3.1 Gestion des bandes

Les bandes sont consommées selon :

- une stratégie de rotation des supports (ce qui permet de les réutiliser de manière cohérente.)
- une stratégie de rétention (qui définit le temps durant lequel les données seront conservées sur supports magnétiques et protégées contre l'écrasement.)

La politique de rétention (par défaut) est la suivante:

Fréquence de sauvegardes	Rétention des sauvegardes
Quotidienne	10 jours
Hebdomadaire	35 jours
Mensuelle	13 mois
Annuelle	4 ans
Archives log quotidienne	10 jours

Figure 16: Politique de rétention

3.3.3.1 Chiffres clés

Quelques chiffres clés concernant les sauvegardes :

- 24000 sauvegardes par mois soit environ 800 par jour
- 731 Téra de données sauvegardées, dont 135 pour Oracle (hors SAP) par mois.

3.3 DESCRIPTION DU PROJET

3.3.1 Problématique

Au commencement du projet, sauvegarder une base revient à recopier tous ses fichiers sur bandes. Il n'est pas possible de faire des sauvegardes incrémentales. Pourtant, le logiciel DataProtector est pourvu de mécanismes permettant des sauvegardes incrémentales et différentielles. Mais, il opère au niveau du fichier : si un seul bit est modifié dans un fichier, DataProtector considère qu'il faut le sauvegarder à nouveau et complètement. Ce mode est inadapté si les fichiers sont constamment en modification. C'est le cas pour une base Oracle,

où chaque fichier est modifié en permanence. Le mode incrémental offert par DataProtector n'apporte donc aucun avantage : les fichiers évoluent en permanence, il faut sauvegarder systématiquement tous les fichiers de chaque base (fichiers de données, fichiers de contrôle, fichiers Redolog, fichiers d'archivage ...), ce qui entraîne :

- une forte volumétrie de données à sauvegarder,
- un nombre élevé de bandes et de lecteurs de bandes,
- des temps de sauvegardes conséquents
- une indisponibilité importante lors des sauvegardes à froid
- une possible saturation des lecteurs et robots de sauvegardes

Pour mettre en œuvre de véritables sauvegardes incrémentales, il est nécessaire de trouver un nouvel outil qui ne doit plus opérer au niveau des fichiers, mais à un niveau plus fin. Pour une base de données Oracle, il est préférable d'opérer au niveau logique, celui des blocs de données Oracle. En travaillant à ce niveau, il devient possible d'exclure les blocs vides, inutiles ou déjà sauvegardés. De vraies sauvegardes incrémentales et différentielles deviennent possibles.

Le seul outil capable de résoudre notre problème s'appelle RMAN car c'est le seul outil spécialisé dans la sauvegarde de base Oracle qui sache opérer au niveau du bloc de données Oracle. RMAN (**R**ecover **MAN**ager) a été développé par Oracle précisément pour répondre aux besoins d'une sauvegarde optimisée :

- Il ne sauvegarde que les blocs utilisés
- Il ne sauvegarde pas les blocs inutiles (blocs temporaires par exemple)
- Il propose divers types de sauvegardes notamment différentiels, cumulatifs.
- RMAN est fiable
- RMAN dispose de fonctionnalités avancées telles que le clonage et la simulation de restauration.

Rappelons que RMAN ne sait pas écrire tout seul sur bande, il lui faut l'aide d'un logiciel de sauvegarde. RMAN ne remplace donc pas le logiciel de sauvegarde. RMAN ajoute de l'intelligence en permettant à un logiciel de sauvegarde d'opérer au niveau des blocs de données.

3.3.2 Objectifs du projet

Les objectifs du projet et les contraintes ont clairement été énoncés dans le cahier des charges. L'objectif principal est l'optimisation du service de sauvegarde des bases ORACLE en introduisant l'outil RMAN, seul outil capable de répondre à notre problématique. Grâce à la réalisation de vraies sauvegardes incrémentales, il sera possible de réduire la durée de sauvegarde, la consommation de média, et celle de la robotique tout en offrant une meilleure fiabilité des restaurations. Enfin, ce projet vise à mettre en place le mode de sauvegarde recommandé par l'éditeur Oracle. L'introduction de RMAN devra cependant conserver autant que possible les scénarios de sauvegardes, et les ordonnancements déjà existants.

3.3.3 Périmètre du projet

Seules les bases de données de l'éditeur Oracle sont éligibles aux modifications de sauvegarde. En effet, RMAN ne fonctionne qu'avec les bases Oracle. Les bases utilisées pour l'ERP SAP ne sont pas incluses dans le périmètre du projet actuel. Au total il est prévu de migrer les sauvegardes d'environ 500 bases.

3.3.4 Démarche et déroulement du projet

La démarche mise en œuvre dans ce projet est à la fois prudente et professionnelle.

Au démarrage du projet, les avantages de RMAN n'étaient pas évidents pour tous les acteurs. C'est pourquoi, le chef de projet, et moi-même, avons cherché une méthodologie qui puisse rapidement produire des résultats pour lever d'éventuelles résistances au changement.

Le projet a donc été coupé en deux phases :

- La première partie visait à montrer qu'il était possible de mettre en place RMAN sur un périmètre réduit, de quelques bases au total, avec un coût et une charge de travail modérée. L'objectif était alors de démontrer la faisabilité ainsi que plus value de la solution.
- Dans une deuxième partie, on a cherché à étendre et à déployer la solution sur un périmètre aussi large que possible. (Plusieurs centaines de bases).

3.3.4.1 Phase du projet

Le projet s'est donc déroulé en 2 parties distinctes.

- 1^{ère} phase du projet (périmètre réduit) :

La première partie du projet (d'août 2009 à mars 2010) a consisté à évaluer la solution RMAN, puis à la déployer sur un périmètre réduit. L'objectif fut de prouver que l'outil RMAN pouvait être mis en place dans l'environnement DataProtector / VTOM existant avec un coût raisonnable. La solution envisagée a dû bien sûr, préserver au mieux l'architecture initiale (VTOM, DP).

Nous avons étudié la solution et avons mis en place une maquette qui nous a permis de rédiger une série de documents de mise en œuvre et d'installation.

Ensuite, nous avons profité opportunément d'un projet (PHENIX) et de son besoin de sauvegarde particulier pour y intégrer et valider notre solution basée sur RMAN. Ce projet PHENIX a servi de pilote pour les bases nouvelles non sauvegardées.

Dans cette 1^{ère} partie les principales difficultés analysées furent :

- **techniques** : Chaque composant de la chaîne de sauvegarde (DP : VTOM, RMAN) disposait de ses propres contraintes. Les divers composants devaient communiquer entre eux de manière cohérente : le planificateur VTOM, le logiciel de sauvegarde DataProtector, la robotique de sauvegarde, le logiciel RMAN et accessoirement la base de données.
 - **organisationnelles** : de nombreux intervenants participaient à la mise en œuvre des sauvegardes. Il a donc été nécessaire d'analyser les impacts sur l'organisation et de définir le nouveau mode de fonctionnement ainsi qu'un plan de formation, d'assistance et de support lorsque cela fut nécessaire.
- 2^{ème} phase du projet (généralisation de la solution):

La seconde partie s'est déroulée de mars 2010 à février 2011 et avait pour but de généraliser la solution de manière optimisée, pour le périmètre le plus large possible (Au maximum 500 bases).

Il a fallu d'abord valider la solution sur les versions de bases et de systèmes d'exploitation qui n'avaient pas encore été mises en œuvre ni en phase maquette ni pendant le pilote PHENIX. Nous avons dû faire cohabiter RMAN avec l'ancien mode de sauvegarde lorsque plusieurs bases se trouvent déjà sur un serveur. Nous avons recherché la meilleure organisation de catalogues possibles puis cherché à optimiser le coût et le temps de déploiement.

Le déploiement a été précédé, d'une phase de préparation de l'infrastructure (mise en place du cluster hébergeant les catalogues), et d'actions d'accompagnement au changement (information, support, et formation des acteurs impliqués). Cette phase a été suivie d'un pilote nommé « SHIVA ». Il s'agissait de passer des sauvegardes Oracle et déjà existantes, vers RMAN. Finalement le déploiement de la solution a pu commencer. Il continue à ce jour.

En multipliant le nombre de bases, de nouveaux problèmes sont apparus. Le système de catalogue RMAN devenait un élément critique puisqu'il devenait nécessaire aux sauvegardes de 750 bases. De plus, le temps de déploiement de la solution devient une contrainte importante. De nombreuses questions se sont alors posées. Comment automatiser la mise en place des sauvegardes RMAN ? Comment structurer le ou les catalogues ? Comment mettre en place une infrastructure robuste ? Comment exploiter et maintenir les catalogue(s) ? Comment surveiller et prévenir les pannes ? Comment gérer efficacement les mots de passe ?

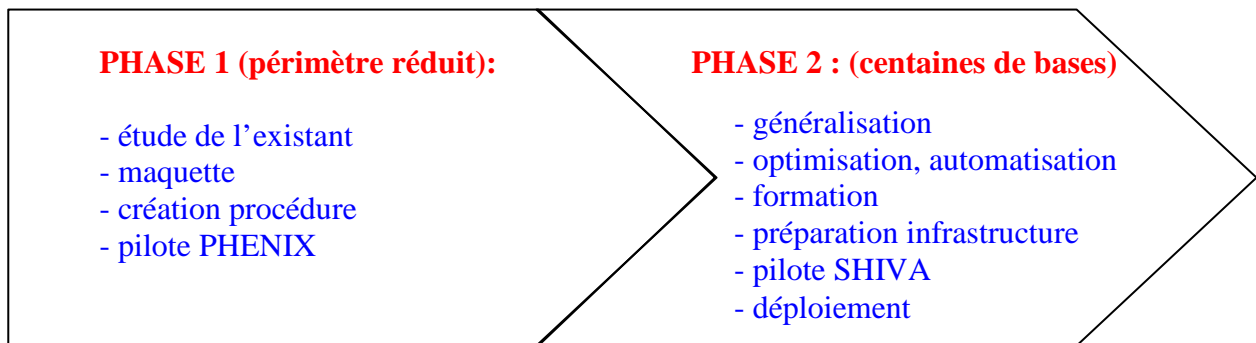


Figure 17: Phases du projet, en deux parties.

3.3.4.2 Dates prévisionnelles

Lot 1 : *Fourniture du service sur un périmètre réduit (Projet PHENIX)*

- *Etude de l'existant et Maquette : 15 Octobre 2009*
- *Pré-Production PHENIX : 31 Décembre 2009*
- *Production PHENIX : 31 mars 2010*

Lot 2 : *Généralisation (Mars 2010 – aujourd'hui) :*

- *Généralisation aux autres bases et aux autres systèmes d'exploitation*
- *Etudes d'architecture / Rationalisation / automatisation*
- *Mise en place de l'infrastructure*
- *Formation, information, assistance au changement*

- *Elaboration d'un pilote nommé SHIVA (migration vers RMAN dans un environnement mutualisé)*
- *Déploiement*

3.4 LE CONTEXTE

Pour mieux appréhender le déroulement du projet, il est important de restituer le contexte pour en mesurer les contraintes et les difficultés.

Bien sûr, la première contrainte est de s'intégrer au mieux dans l'infrastructure de sauvegarde existante (VTOM / DataProtector). Notons que la solution basée sur RMAN pourra être mise en œuvre en remplacement mais aussi en complément des moyens de sauvegarde actuels.

3.4.1 Contexte organisationnel

3.4.1.1 Les acteurs de la sauvegarde

La mise en place d'une sauvegarde et/ou d'une restauration implique de nombreux acteurs

- Les **DBA** (administrateurs de bases de données ou en anglais **D**ata**B**ase **A**dministrator): ils sont responsables des bases et du paramétrage RMAN,
- L'équipe **ADS** (**A**dministration **D**es **S**auvegardes): elle est responsable de la mise en place et du suivi des sauvegardes,
- **IOP** (**I**nfrastructure **O**util **P**roduction): elle est responsable de l'infrastructure de sauvegarde (robotique, administration DP),

- Le **BT** (Bureau technique): il met en place et gère les traitements dans l'ordonnanceur. (Pour notre projet les traitements de sauvegarde),
- Le **CQP** (**C**ontrôle **Q**ualité **P**roduction): équipe qui surveille la qualité,
- L'équipe **DRP** (**D**isaster **R**ecovery **P**lan): elle organise les tests de restauration.

3.4.1.2 Les acteurs du projet

Un chef de projet a été nommé pour mener le projet. Un expert Oracle et un architecte (pour DataProtector), fournissent un support.

Pour les besoins du mémoire, j'ai bénéficié d'une large autonomie. Mon rôle dans le projet a été principalement d'assister le chef de projet et plus particulièrement :

- L'aide à la coordination avec le chef de projet
- La participation à l'étude et la réalisation de tests
- L'étude de l'intégration de l'outil RMAN dans l'exploitation existante
- L'étude et l'aide à l'industrialisation
- La formation / information des équipes
- La mise en œuvre de la solution pour le compte l'équipe de DBAs à laquelle j'appartiens.

J'avais à ma disposition un environnement de test complet comprenant :

- deux serveurs de test UNIX (AIX et HPUX)
- un environnement de maquette DataProtector et une robotique de sauvegarde.

De nombreux acteurs sont impliqués dans la mise en place de sauvegardes (et restaurations). C'est donc un travail d'équipe et chaque évolution dans la chaîne de sauvegarde demande un effort de coordination et de communication entre les équipes.

3.4.2 Les contraintes

3.4.2.1 La réglementation pharmaceutique (GxP)

Sanofi-Aventis en tant qu'entreprise pharmaceutique est soumise à des obligations réglementaires contraignantes dites « GxP » (Good Practices), et doit suivre une série de procédures très rigoureuses pour produire les médicaments conformément à des critères qualitatifs stricts. L'objectif est de minimiser les risques liés à la production de médicaments.

La réglementation dite « GxP » couvre toutes les étapes de la production : des appareils de production jusqu'à la formation du personnel, sans oublier l'informatique. Des procédures écrites et détaillées sont essentielles pour chaque étape du processus impliquée dans la qualité du produit final. Un système permettant de prouver la régularité du suivi de ces procédures au cours de toutes les productions et à chaque étape du processus de production, doit être mis en place.

Pour notre projet, la norme GxP va se traduire par trois contraintes principales:

- Il est nécessaire de mettre au point les **procédures** avant toute mise en place de notre solution.
- Il faudra fournir des **rapports** d'installation détaillée afin de prouver que la mise en place est conforme aux procédures prévues.
- Enfin il faudra **tracer toutes modifications** réalisées.

Des audits réguliers, internes et externes, sont régulièrement effectués pour veiller au respect des règles. Nous verrons que cet aspect réglementaire aura un impact important tout au long du projet.

3.4.2.2 La réglementation financière (SOX)

Sanofi-Aventis est coté à la bourse de New York et, à ce titre, elle est soumise à la loi Américaine dite « Sarbanes Oxley Act » ou SOX, votée en juillet 2002 après plusieurs

scandales financiers (Par exemple la faillite frauduleuse de la société ENRON spécialisée dans la production et le négoce d'énergie).

Cette loi américaine vise à combattre la fraude comptable et accroît à la fois les responsabilités des entreprises et celle de leurs dirigeants dans leur devoir d'information. Elle impose aux directions d'établir et de maintenir une structure de contrôle interne et des procédures relatives aux communications des informations financières.

L'impact de Sarbanes-Oxley et des contrôles internes sur l'informatique et la gestion de l'information est énorme :

- Mise en place des processus de contrôles mensuels et annuels.
- Gestion de la traçabilité et gestion des logs.
- Gestion des identités.
- Sécurité des ERP : étanchéité des profils, revues des comptes, etc.
- Sauvegardes, Restaurations et Plan de secours.
- Gestion des données confidentielles.
- Maîtrise du risque informatique : mise à jour de sécurité, antivirus ...
- Maîtrise du parc : assurances, inventaires, vols.

Il faut intégrer ces diverses contraintes dans le projet. En pratique nous aurons les mêmes obligations que pour Gxp, à savoir :

- rédiger des procédures de mise en place de la solution,
- rédiger des rapports d'installation,
- tracer les modifications.

Comme pour GxP nous sommes audités régulièrement.

3.4.2.3 Contraintes liées au nombre de bases

En multipliant le nombre de bases, de nouveaux problèmes sont prévisibles et de nombreuses questions se poseront parmi lesquelles :

- Comment déployer à moindre coût la mise en place des sauvegardes RMAN ?
- Quelle architecture choisir ?
- Comment la rendre robuste, fiable et maintenable ?

3.4.2.4 Contraintes d'exploitation

Les contraintes d'exploitation et la continuité du service de sauvegarde devront être prises en compte. Il faudra s'assurer que la mise en œuvre de nouvelles sauvegardes avec les composants RMAN ne perturbe pas le système d'information existant.

3.5 RESUME DU CHAPITRE 3

Le projet a été mené à Sanofi-Aventis, leader de l'industrie pharmaceutique. J'ai participé à ce projet en tant qu'administrateur de bases de données de production.

Le projet consiste à modifier l'infrastructure de sauvegarde existante. Celle-ci est composée de plusieurs composants. Un ordonnanceur est chargé de soumettre les demandes de sauvegardes/restauration au logiciel de sauvegarde DataProtector. Ce dernier réalise ensuite les écritures/lectures sur bande via une robotique.

Le projet vise à moderniser et optimiser le service de sauvegarde des bases oracle grâce à l'introduction d'un nouveau composant de l'éditeur Oracle : **Recovery MAN**ager (RMAN) qui permet des sauvegardes « intelligentes ». A l'instar du logiciel DataProtector qui opère au niveau du fichier, RMAN opère au niveau du bloc de données Oracle: il ne sauvegarde que les blocs réellement utilisés et sait exclure les blocs déjà copiés sur bande. RMAN autorise de nouveaux types de sauvegarde différentiels (type incrémental et cumulatif) et offre des fonctionnalités avancées telles que le clonage ou la simulation de restauration.

Le projet s'est déroulé en 2 phases distinctes. Une première phase a consisté à étudier à l'aide d'une maquette la solution sur un nombre limité de bases puis à mettre en œuvre un premier pilote PHENIX. La deuxième phase du projet s'est attachée à généraliser la solution sur un

périmètre aussi large que possible. Elle a abouti au déploiement général de la solution sur plusieurs centaines de bases.

Or, avec un nombre potentiel élevé de bases (600 à 700 bases), il faut prévoir une solution optimisée automatisée et robuste.

Comme dans tout projet, il est salubre de prendre en compte les réalités de l'entreprise. D'abord, le projet doit s'intégrer dans une l'infrastructure de sauvegarde existante sans en perturber le fonctionnement. Le projet nécessite la coordination de nombreux acteurs. Le projet devra se plier à la réglementation en vigueur pharmaceutique (Gxp) et financière. (SoX).

Chapitre 4 : RMAN SUR UN PERIMETRE REDUIT

Dans ce chapitre nous nous limitons à mettre en œuvre RMAN que sur un périmètre réduit, c'est-à-dire à ne sauvegarder (avec RMAN) qu'un nombre limité de bases Oracle. (Nous étudierons la généralisation de RMAN à l'ensemble des bases, et l'automatisation de la mise en œuvre plus tard, au chapitre qui suivra.)

Ce chapitre se décompose en trois temps :

- Nous commencerons par une phase d'étude de l'existant, c'est-à-dire des divers composants logiciels. Nous chercherons à comprendre le fonctionnement de la chaîne de sauvegarde (VTOM et DataProtector), ainsi que celui de l'outil RMAN.
- Dans un deuxième temps, nous analyserons le moyen d'intégrer RMAN dans un environnement de maquettes. Nous mettrons ensuite au point les procédures nécessaires au déploiement RMAN.
- Enfin, nous validerons ces procédures au cours d'une phase pilote sur deux bases en utilisation.

4.1 ETUDE DE L'EXISTANT

Cette partie analyse le fonctionnement des divers composants. Tout d'abord, nous étudierons, un à un, les composants de la chaîne de sauvegarde existante. Ensuite nous étudierons l'outil RMAN lui-même. Nous nous concentrerons sur les points importants pour le projet.

4.1.1 Présentation de la chaîne de sauvegarde existante

Nous avons déjà vu que l'infrastructure de sauvegarde est constituée des éléments suivants :

- Visual TOM ordonnanceur de la société Abssys (VTOM) qui lance les tâches de sauvegardes selon un planning et des règles définies à l'avance.
- Le logiciel de sauvegarde HP DataProtector. (De la société Hewlett Packard). Ce dernier lit les fichiers à sauvegarder, et les écrits sur bande par l'intermédiaire d'un robot de sauvegarde.
- Une robotique.

La chaîne de sauvegarde initiale (sans RMAN) peut être schématisée de la façon suivante :

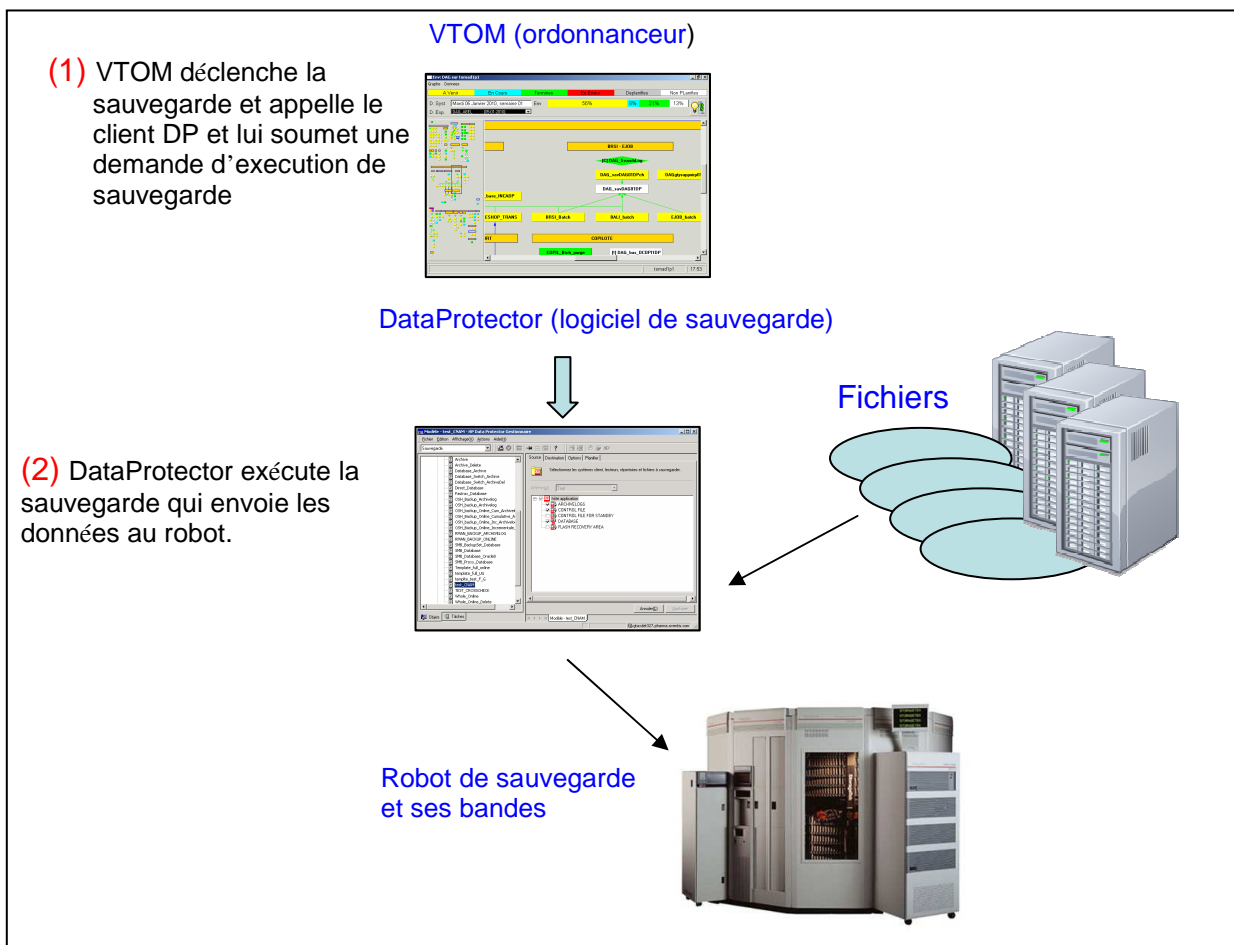


Figure 18: Description simplifiée de la chaîne de sauvegarde (sans RMAN)

Il n'est pas nécessaire d'étudier la robotique, car celle-ci ne sera pas impactée par le projet.

Seuls VTOM et DataProtector seront étudiés. Le fonctionnement de ces logiciels est relativement complexe si bien qu'il a été nécessaire de consulter les diverses documentations fournies par les différents éditeurs (ABSSYS pour l'Ordonnanceur VTOM et HEWLETT PACKARD pour DataProtector). J'ai eu la chance de manipuler et tester le logiciel de sauvegarde grâce à un environnement de test composé de DataProtector, et d'une robotique mise à ma disposition.

Entrons à présent dans le détail.

4.1.2 L'ordonnanceur VTOM

VTOM (Visual TOM) est un logiciel d'ordonnancement de la société ABSSYS qui permet le lancement planifié et contrôlé de séries de tâches (appelé aussi des « *batch* ») sur l'ensemble du système d'information. Pour les besoins de notre projet, il n'est pas utile de décrire l'architecture du logiciel d'ordonnancement VTOM car nous cherchons simplement à comprendre les enchaînements utilisés pour remplir le service de sauvegarde des bases de données Oracle. Nous cherchons à entrevoir ce qui va changer pour intégrer RMAN.

La mise en place des traitements sous VTOM se réalise dans une interface graphique et revient à définir ou modifier des traitements représentés par des « boîtes ». Celles-ci sont ensuite reliées pour y définir des enchaînements entre des traitements, à une heure définie et/ou à la suite d'autres traitements, comme le montre l'exemple ci-dessous :

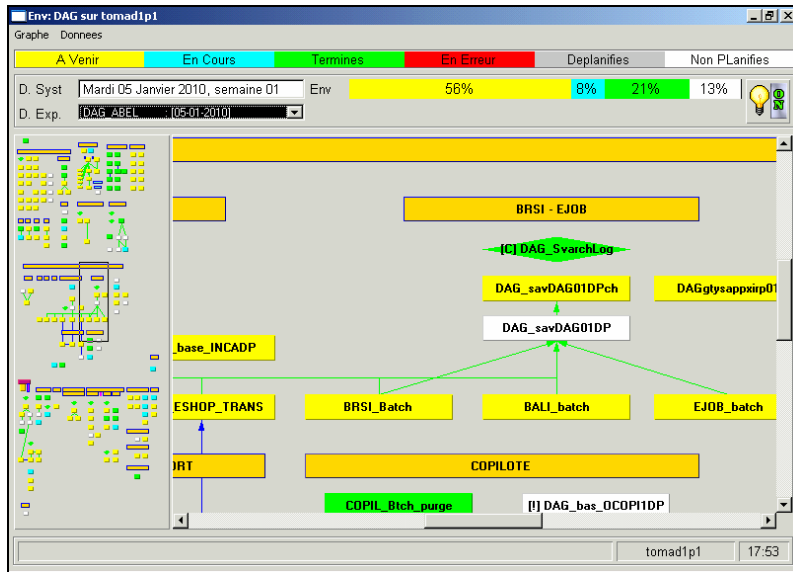


Figure 19: Exemple d'enchaînement dans une fenêtre graphique de VTOM

Deux stratégies de sauvegarde de bases Oracle type y sont déployées :

- Des sauvegardes à froid chaque jour (ou chaque semaine)
- Des sauvegardes à chaud en semaine complétées par des sauvegardes à froid le week-end.

A chacune de ces deux stratégies, correspond un scénario type pour l'ordonnancement, comme le montrent les 2 figures ci-dessous.

Remarque : La flèche dans la figure ci-dessous, signifie que la tâche B dépend de conditions sur la tâche A.

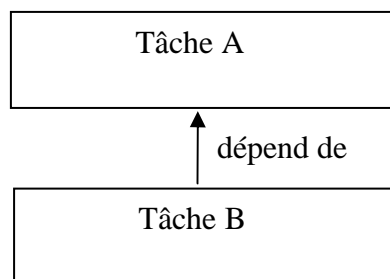


Figure 20: Ordonnancement de 2 tâches A et B avec B dépend de A

4.1.2.1 Ordonnancement des sauvegardes à froid sous VTOM

Les sauvegardes à froid sont ordonnancées de la façon suivante :

- réaliser un export de la base (facultatif)
- Arrêter la base
- Sauvegarde des fichiers (selon le cas, journalière, hebdomadaire ou mensuelle)
- Relancer la base

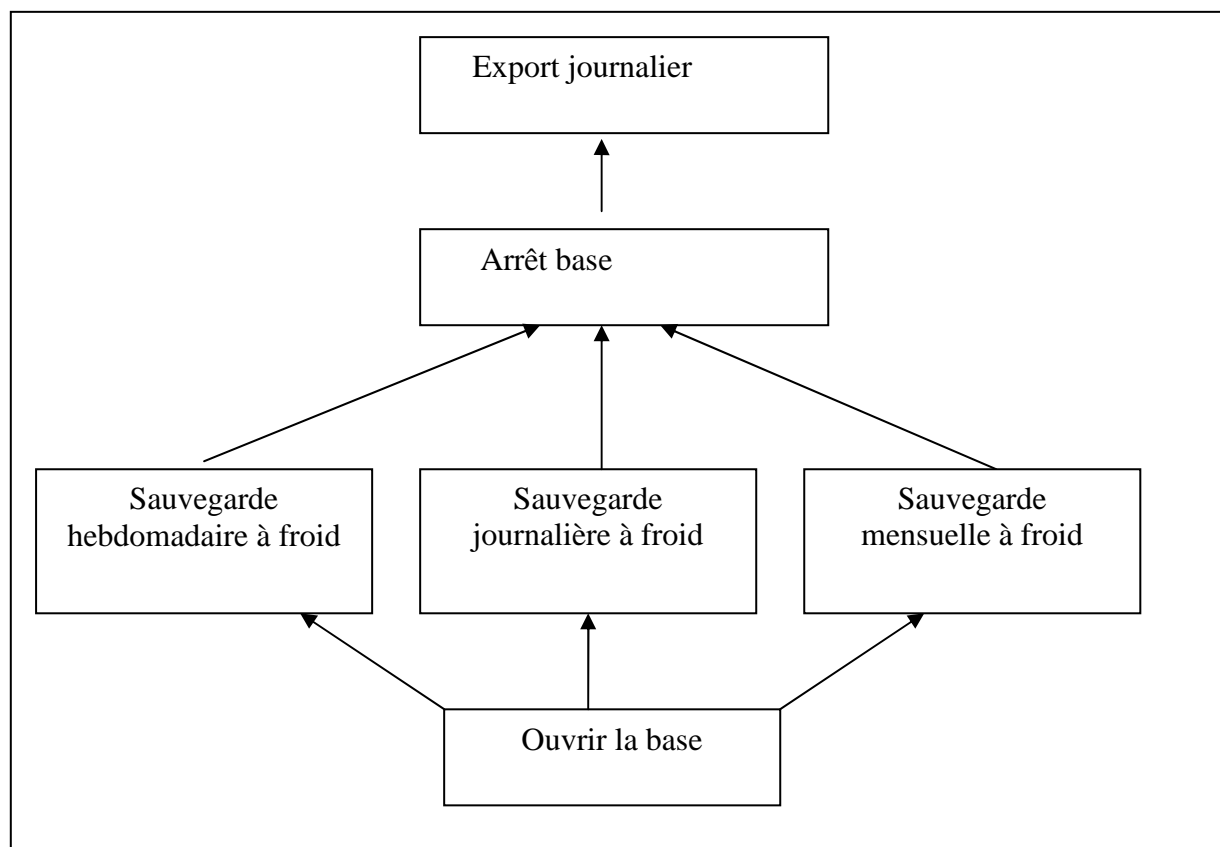


Figure 21: Ordonnancement type pour les sauvegardes à froid

4.1.2.2 Ordonnancement des sauvegardes à chaud sous VTOM

Pour la sauvegarde à chaud, deux enchaînements distincts sont nécessaires :

- l'un pour la sauvegarde des fichiers, en mode « begin backup ». (cf. chapitre 2)
- l'autre pour la sauvegarde des fichiers d'archives.

a) Sauvegarde à chaud : sauvegarde des fichiers de la base

Les sauvegardes à chaud des fichiers de la base ne sont lancées que durant la semaine. Le week-end une sauvegarde à froid est lancée. Nous avons un l'enchaînement type suivant. :

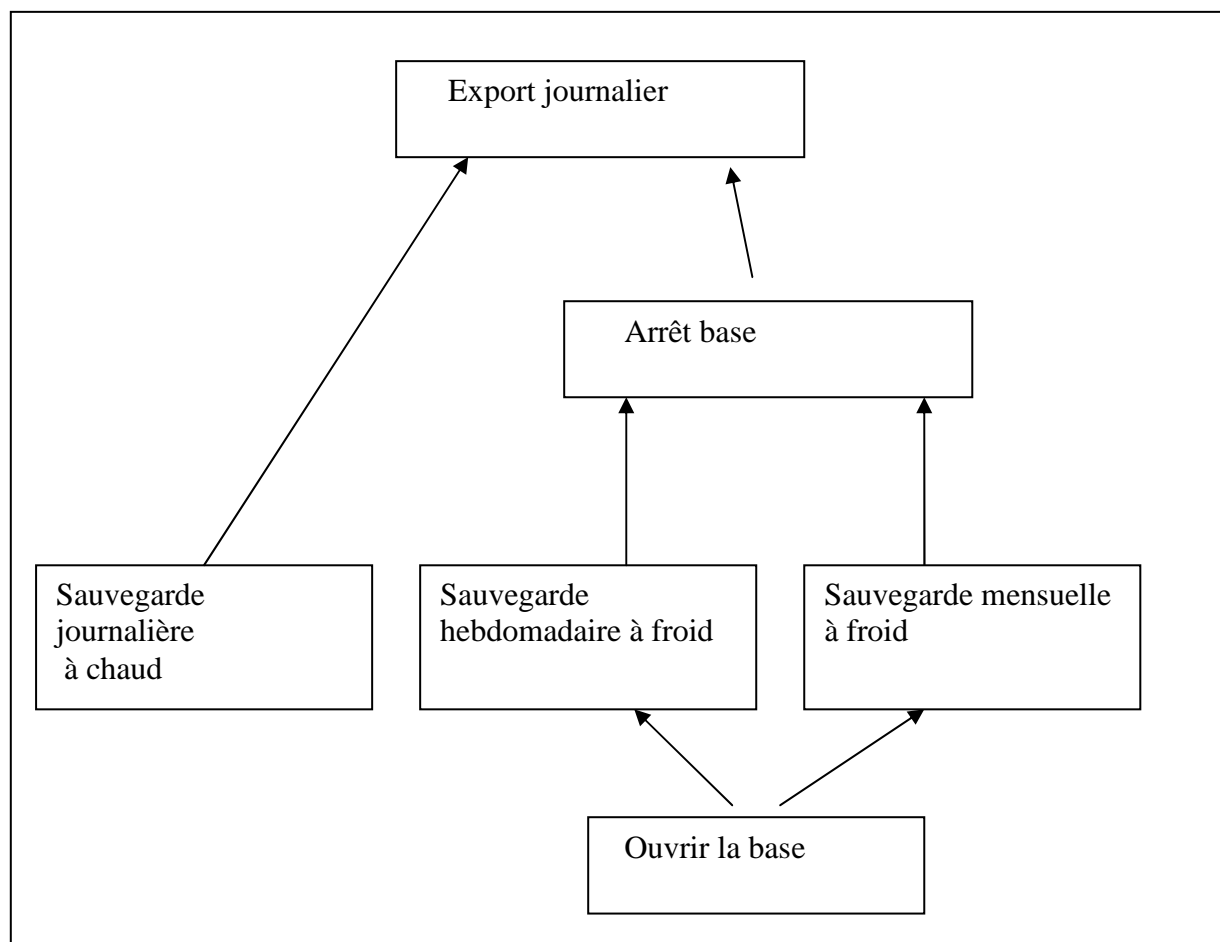


Figure 22: Ordonnancement type pour les sauvegardes à froid des fichiers de base

Les tâches de sauvegarde elle-même se composent de tâches plus petites :

- changement de fichier de redolog (« switch » de log, cf. chapitre 2, partie 2.1.3)
- recopier les archives dans un répertoire temporaire
- sauvegarder les archives depuis ce répertoire temporaire
- passer des fichiers en mode « Begin backup » (Cf. chapitre 2)
- recopier des fichiers sur bande
- changer de fichier de log (« switch » de log, cf. chapitre 2, partie 2.1.3)
- recopier les fichiers de contrôle sur disque
- sauvegarder sur bande du fichier de contrôle recopié précédemment sur bande

b) Sauvegarde à chaud : sauvegarde des fichiers d'archives

Les sauvegardes à chaud nécessitent une sauvegarde régulière des fichiers d'archives, car ceux-ci sont générés de manière continue sur un espace disque limité. Pour la sauvegarde des fichiers d'archivage, le scénario type consiste à exécuter successivement les actions suivantes:

- Vérifier le remplissage du système de fichier dédié aux archives. Si le seuil est atteint (moins de 50% d'espace libre), on passe à l'étape suivante. (Sinon, on ne fait pas de sauvegarde des archives, et on replanifie la prochaine exécution.)
- Réaliser un changement (« switch ») de fichier de redolog, puis déplacer les archives dans un répertoire temporaire. (Cette action vise à séparer les archives à sauvegarder de celles qui vont être générées pendant la sauvegarde, pour empêcher la sauvegarde de fichier en cours d'archivage).
- Copier les archives sur bande
- Effacer l'exemplaire des archives mises sur bandes.

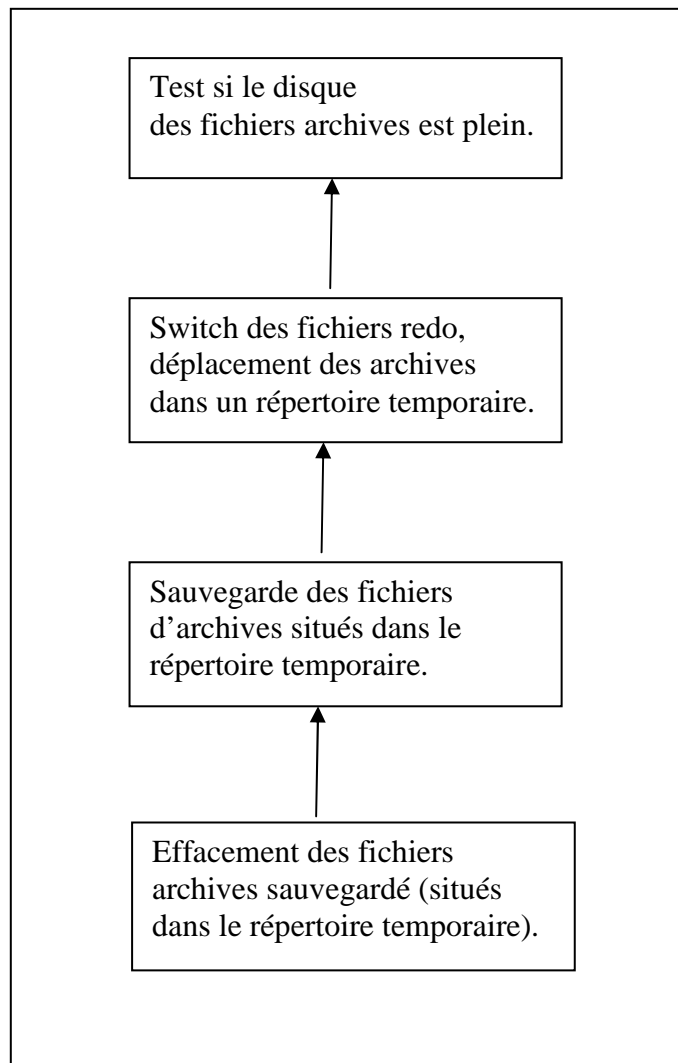


Figure 23: Ordonnancement type pour les sauvegardes des fichiers d'archives

4.1.2.3 Le script « shob2sauv » appelé par les tâches de sauvegarde VTOM

Toutes les sauvegardes sur bandes lancent un script unique (*appelé SHOB2SAUV*) chargé de soumettre les diverses sauvegardes à DataProtector, qu'il s'agisse de sauvegardes de fichiers, Exchange, SAP, ou autres.

Ce script est l'unique script appelé par VTOM pour les sauvegardes. Seul son paramétrage diffère. L'utilisation d'un seul script simplifie la mise en place des sauvegardes.

Le script lance la commande DataProtector « *omnib* » qui exécute la sauvegarde, avec les paramètres adéquats. La commande « *omnib* » a besoin de paramètres tels que le nom de spécification du type de sauvegarde (SAP, RMAN, exchange ...), ainsi que du nom de spécification de sauvegarde (créé plus haut).

L'ossature du script shob2sauve est simple : on teste (avec une instruction « case ») le type de sauvegarde (SAP, exchange, système de fichiers ...) puis, selon ce dernier, les paramètres adéquats sont positionnés. Le script lance ensuite la commande qui déclenche la sauvegarde, à savoir « *omnib* » en lui passant tous les paramètres requis.

Le script est de la forme suivante :

```
#####
#
# Procedure : shob2sauve
#
# Fonction : Sauvegarde par OB2
#
# Parametre :
# 1 : Type de save ( SAP , EXCHANGE , LOTUS , ORACLE , FS )
# 2 : Nom de la datalist OB2
# 3 : Type de la sauvegarde
# 4 : Type de retention (Days, Weeks, Month, none)
# 5 : Temps de retention
#
#####
.
.
.
case $TYPE in
  SAP|sap)    MODE=barmode
              TYPE=sap_list
              case ${MODELIST} in
                FULL) MODELIST=FULL ;;
                INCREMENTAL) MODELIST=INCREMENTAL ;;
                *) echo "incoherence entre le type de Sa
uve $TYPE et le mode $MODELIST"
              exit 2 ;;
              esac
              ;;
  EXCHANGE|exchange)  MODE=barmode
              TYPE=msese_list
              case ${MODELIST} in
                FULL) MODELIST=full ;;
                INCR) MODELIST=incr ;;
                *) echo "incoherence entre le type de Sa
uve $TYPE et le mode $MODELIST"
              exit 2 ;;
              esac
              ;;
  LOTUS|lotus)    MODE=barmode
              TYPE=lotus_list
              case ${MODELIST} in
                FULL) MODELIST=FULL ;;
                INCR) MODELIST=INCR ;;
                *) echo "incoherence entre le type de Sa
uve $TYPE et le mode $MODELIST"
              exit 2 ;;
              esac
              ;;

```

```

HOST|host)  MODE=mode
            TYPE=datalist
            case ${MODELIST} in
                FULL) MODELIST=FULL ;;
                INCREMENTAL[1-9]) MODELIST=${MODELIST} ;
            ;
            *) echo "incoherence entre le type de Sa
            uve $TYPE et le mode $MODELIST"
            exit 2 ;;
            esac
        ;;
    esac
    ◦
    ◦
    ◦
    omnib -${TYPE} ${DATALIST} -${MODE} ${MODELIST} -protect ${RETENT} ${NUMBER} ${
    OPT_SUP}

```

Figure 24: Script de lancement des sauvegardes « shob2sauv »

La signification des paramètres nécessaires au lancement du script shob2sauve est la suivante:

- 1 : Type de données (SAP, EXCHANGE, LOTUS, ORACLE, FS)
- 2 : Nom de la spécification de sauvegarde (créé préalablement sous DataProtector)
- 3 : Type de la sauvegarde (Complète, incrémental, etc.)
- 4 : Rétention (Jour, semaine, mois, sans objet).
- 5 : Temps de rétention (en fonction du type de rétention)

On remarque que peu de modifications sont nécessaires pour y intégrer le nouveau type de sauvegarde RMAN. Il sera facile d'ajouter, le type de sauvegarde RMAN, en plus de ceux qui existe déjà.

4.1.2.4 Conclusion : que retenir de l'étude du logiciel VTOM ?

Il existe 2 scénarii types pour les sauvegardes :

- Un pour les sauvegardes à froid, avec un seul enchaînement de tâches ordonnancées.
- Un autre pour les sauvegardes à chaud, nécessitant 2 enchaînements de tâche VTOM: un premier pour les sauvegardes fichiers de la base, un deuxième pour les sauvegardes des fichiers d'archive.

Pour lancer une sauvegarde, VTOM lance un script unique, nommé « *SHOB2SAUV* » quel que soit le type de sauvegarde (SAP, fichiers, Exchange ...). Ce dernier exécute la commande DataProtector nommée « *omnib* », avec les bons paramètres. (Le nom de la sauvegarde, type de sauvegarde notamment).

Pour le projet, il nous faudra donc répondre aux questions suivantes :

- Quels sont les éventuelles modifications à apporter au script « *SHOB2SAUV* » ?
- Comment faut-il définir les nouveaux enchaînements VTOM pour pouvoir lancer des sauvegardes incrémentales en plus des sauvegardes complètes?

4.1.3 Le logiciel de sauvegarde DataProtector

Le logiciel DataProtector permet des sauvegardes de toutes sortes (fichiers, systèmes de fichiers, messagerie Exchange, SAP, RMAN, etc.), sur divers types de système d'exploitation (UNIX, Windows...).

Nous n'allons pas détailler le fonctionnement complet du logiciel, mais nous concentrer sur :

- son fonctionnement général,
- la création de sauvegardes et de modèles.

4.1.3.1 Principe de fonctionnement de DataProtector

DataProtector est composé des éléments principaux suivants ;

- **Le gestionnaire de cellule** (en anglais *Cell Manager*) : c'est le cœur de l'infrastructure applicative DataProtector qui centralise et contrôle toute les opérations liées aux sauvegardes : gestion et lancement des sauvegardes et restaurations, gestion des agents, gestion des sessions de sauvegarde et restauration, etc. Cette application centrale est hébergée par un serveur ou un cluster, c'est-à-dire 2 ou plusieurs serveurs redondants assurant la continuité du service. Dans notre cas, le logiciel DataProtector est installé sur un cluster composé de 2 serveurs UNIX.
- **Le référentiel DataProtector** : le logiciel DataProtector stocke diverses informations dans une base de données propriétaire interne (appelée aussi IDB pour *Internal Database*) : spécifications et déclaration des sauvegardes, historique des sauvegardes et restauration, information sur les agent déployés....Ces informations sont nécessaires aux sauvegardes: il peut s'agir de l'historique des sauvegardes/restaurations, paramétrage des serveurs et composants déclarés dans DataProtector, des informations sur les bandes et lecteurs...
- **Des spécifications de sauvegarde** stockées sur le Cell Manager. Il s'agit de fichiers qui décrivent et détaillent ce qu'il est nécessaire de sauvegarder. Le contenu de ces fichiers diffère selon le type de sauvegarde (fichiers, RMAN, messagerie Exchange, etc.).
- Des interfaces déposées sur le(s) serveur(s) à sauvegarder, capables de communiquer avec les divers composants à sauvegarder. Lors d'une sauvegarde ou restauration, DataProtector communique avec les serveurs à sauvegarder et les robots de sauvegarde grâce à des agents qu'il faut préalablement installer. Dans notre cas il s'agit des bibliothèques RMAN pour DataProtector.
- Des interfaces capables de communiquer avec la robotique. (Appelées aussi **agent**)

Toute action sous **Data Protector** (DP) peut être déclenchée par l'exécutable « omnib », ou bien par l'intermédiaire d'une console graphique (cf. figure ci-dessous) . Cette dernière est accessible à distance sur un ordinateur personnel (Nécessite l'installation d'un client) et permet toutes les opérations d'administration des sauvegardes telles que :

- La création de modèles de sauvegarde,
- La création de spécifications de sauvegarde
- Le déploiement d'agents Data Protector
- Les sauvegardes/restaurations
- La gestion des sessions
- Etc.

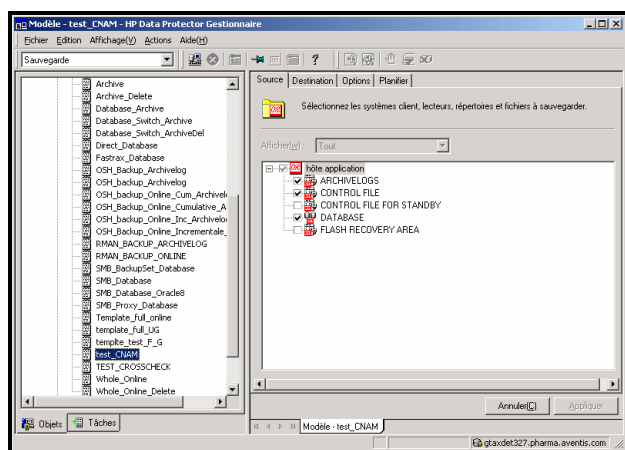


Figure 25: Interface graphique du logiciel de sauvegarde Data Protector

L'architecture générale du logiciel DataProtector se présente de la façon suivante :

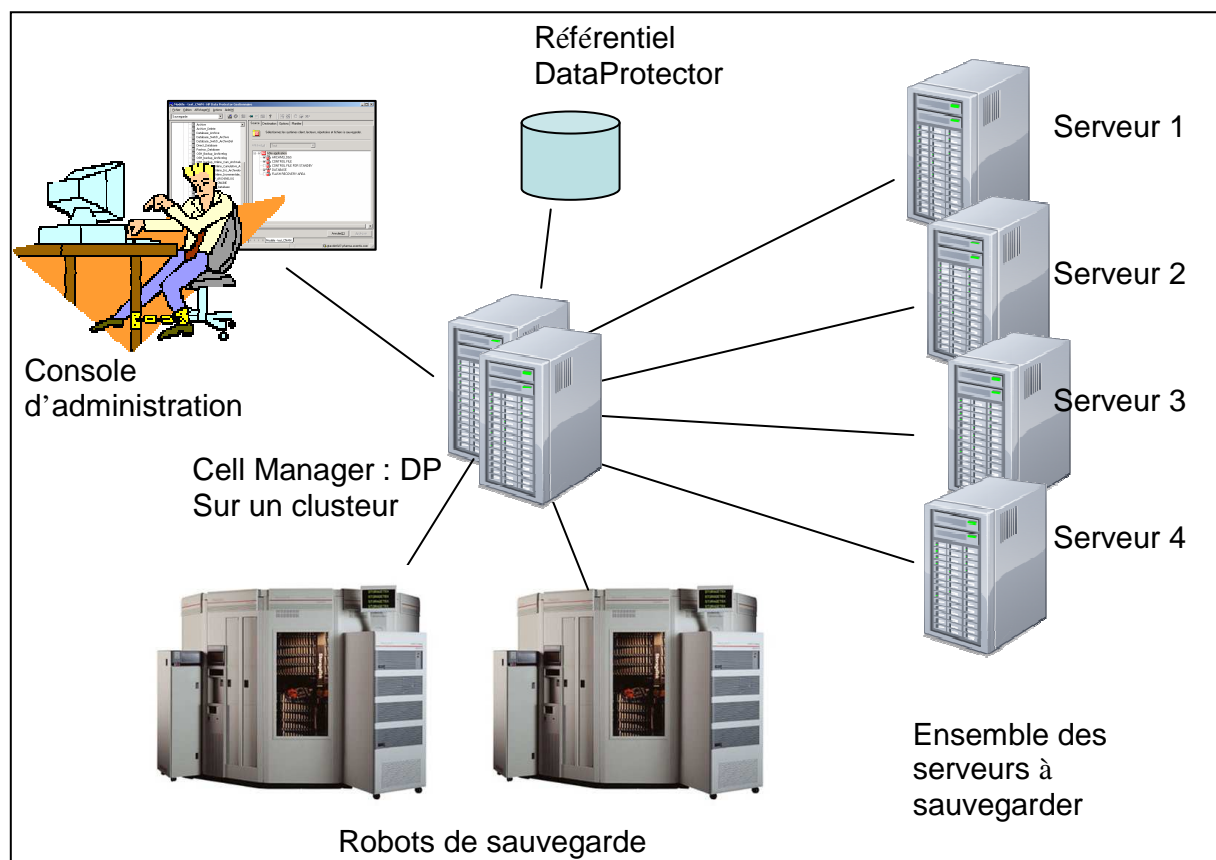


Figure 26: Architecture simplifiée DataProtector

4.1.3.2 Principe pour définir les sauvegardes sous Dataprotector

Nous avons vu précédemment que VTOM soumet à DataProtector les demandes de sauvegardes en invoquant commande « *omnib* » avec les bons paramètres.

Le paramètre le plus important est la spécification de sauvegarde qui décrit précisément ce qu'il faut sauvegarder.

Nous allons voir plus bas, comment créer une spécification de sauvegarde et que cette dernière est précédée de la création d'un modèle.

Mise en place des sauvegardes sous DP

La mise en place d'une sauvegarde sous DP (appelée par VTOM par la commande « *omnib* ») se réalise en 2 temps :

- Tout d'abord, il est nécessaire de créer une « **Template** », c'est-à-dire un **modèle de sauvegarde**. DataProtector dispose de modèles prédéfinis qu'il est possible de modifier et d'adapter aux besoins.
- Ensuite, à partir d'un modèle, on définit une **spécification de sauvegarde**. Le modèle sert de référence pour créer d'autres sauvegardes et contiendra toutes les caractéristiques d'une sauvegarde. (Ici a chaud, à froid, archive log, ...).

Création d'un modèle

Pour créer un modèle, on utilise l'outil d'administration DP en cliquant sur le type de modèle correspondant au type de sauvegarde souhaitée (SAP, Exchange, SQL server ...).

Il faut ensuite suivre l'assistant et remplir les différents champs demandé.

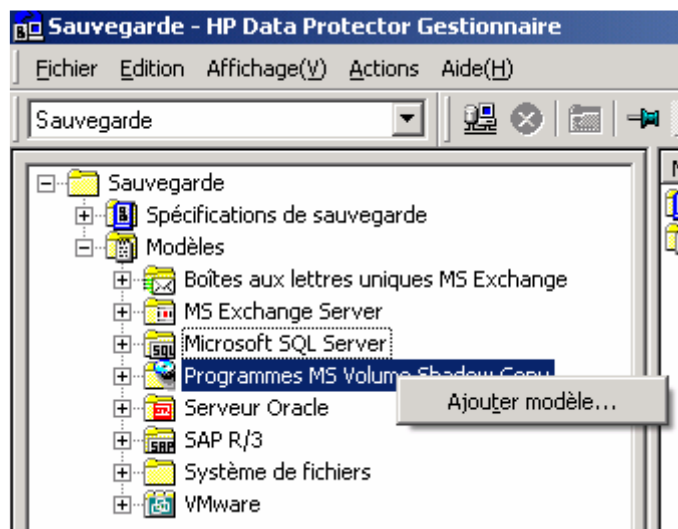


Figure 27: Création d'un modèle de sauvegarde

Une fois toutes les informations nécessaires saisies, il ne reste plus qu'à donner un nom à notre modèle, et à le sauvegarder.

Création d'une spécification à partir d'un modèle de sauvegarde

Le modèle de sauvegarde créé, nous allons nous en servir pour déclarer une sauvegarde qui va hériter de toutes les caractéristiques du modèle, c'est que l'on appelle une spécification de sauvegarde.

Une spécification de sauvegarde est un objet qui porte un nom unique et qui est un ensemble de paramètres de sauvegarde, stockée dans un fichier sur le « Cell Manager ». Lorsqu'on lance une sauvegarde particulière, on lit cette spécification dans le fichier adéquat.

Avec la console d'administration, il suffit de se positionner dans le menu de sauvegarde et de cliquer sur « ajouter sauvegarde », puis sélectionner un modèle, comme l'illustrent les copie d'écrans ci-joints. La sauvegarde va hériter de toutes les propriétés du modèle.

4.1.3.3 Que retenir de l'étude du logiciel DataProtector ?

Le logiciel DataProtector est organisé autour d'un composant majeur le « Cell manager » qui pilote l'administration et les opérations de sauvegardes et restaurations. Il utilise un référentiel propriétaire. Toutes les opérations (création de sauvegarde par exemple) avec le logiciel DataProtector peuvent s'effectuer avec la console d'administration graphique.

Pour dialoguer avec RMAN et pour dialoguer avec la robotique, DataProtector a besoin de programme particulier appelé des « agents » qu'il faut préalablement installer sur chaque machine à sauvegarder.

Pour que DataProtector sache précisément ce qu'il doit sauvegarder, il faut au préalable créer des spécifications de sauvegardes. Ces dernières sont créées dans la console d'administration graphique. Les spécifications sont créées à partir de modèles préétablis et modifiables.

Pour le projet, il nous faudra répondre aux questions suivantes :

- comment permettre la communication entre DataProtector et RMAN ?
- comment créer les modèle et spécifications de sauvegardes RMAN ?

4.1.4 Architecture du composant RMAN

Nous allons nous concentrer, dans ce qui suit sur l'architecture de RMAN. Cette étude sera nécessaire pour intégrer RMAN dans l'infrastructure de sauvegarde.

Catalogue RMAN versus fichier de contrôle

Pour fonctionner, RMAN a besoin d'informations telles que :

- les paramètres de sauvegardes (les informations de communication avec le logiciel de sauvegarde, la rétention, le mode de compression, le média etc.)
- la description de la base à sauvegarder
- les informations sur les sauvegardes en cours
- l'historique des sauvegardes déjà réalisées
- etc.

RMAN enregistre systématiquement ces informations directement dans le fichier de contrôle de la base à sauvegarder. Mais, les informations n'y restent que quelques jours, par défaut 7 jours. (Mais cette durée peut être augmentée en modifiant le paramètre de base. CONTROL_FILE_RECORD_KEEP_TIME.)

Heureusement, il existe un autre moyen de pérenniser ces informations : dans un **catalogue**. Concrètement, ce catalogue est un schéma dans une base, (c'est-à-dire un ensemble d'objets

tels que des tables/indexes, des procédures, des vues, etc. associés à un utilisateur donné) qui va héberger toutes les données utiles à RMAN, aussi longtemps que désiré. Pendant une sauvegarde, RMAN reste connecté au catalogue.

Il est parfaitement possible de ne pas employer un tel catalogue. Les informations relatives aux sauvegardes RMAN sont alors contenues dans les seuls fichiers de contrôle. Cependant, ces derniers deviennent alors extrêmement critiques puisque nécessaires à toute restauration et dans ce cas, il ne faut pas oublier de les sauvegarder à part. Il vaut mieux mettre ces données nécessaires à une sauvegarde ailleurs que sur la base à restaurer, et donc dans un catalogue. Notons que les fichiers de contrôle contiennent moins d'informations et offrent moins de fonctionnalités qu'un catalogue stocké en base. Par exemple, il n'est pas possible de stocker des scripts précompilés dans les fichiers de contrôle, alors que cela est possible avec un catalogue.

C'est pourquoi, ce projet privilégie l'utilisation d'un référentiel stocké dans un schéma de base de données. C'est d'ailleurs une bonne pratique fortement recommandée par l'éditeur Oracle.

Pour utiliser cette configuration, il est impératif de créer le nouveau schéma dans une base de données différente de la base de données cible. A noter que RMAN synchronise automatiquement les informations du fichier de contrôle avec le catalogue, dès lors que certaines opérations sont invoquées sous RMAN.

Il faut cependant garder en tête que l'utilisation d'un catalogue implique aussi des contraintes :

- un catalogue augmente la complexité de l'architecture si bien qu'il faut penser à organiser le(s) catalogue(s)
- il faut penser à l'administration du catalogue (place disque, purge, etc.),
- lors de changements de versions de la base à sauvegarder, il faut penser à l'éventuelle migration du catalogue associé.

Ces aspects seront évalués plus tard dans ce projet.

Architecture de l'outil RMAN

Pour utiliser RMAN et écrire sur bande, il faut disposer des éléments principaux suivants :

- Une base Oracle à sauvegarder.
- L'exécutable RMAN lui-même qui se présente sous forme d'un outil en ligne de commande
- La librairie de gestion des supports, ou **MML** (En anglais **M**edia **M**anagement **L**ibrary): c'est un programme qui réalise l'interface entre RMAN et le logiciel de sauvegarde tiers tel que DataProtector, Net Backup, ou TSM.... C'est cette interface qui permet à RMAN de lire et écrire sur bande. La MML est généralement fournie par l'éditeur du logiciel de sauvegarde. Elle est généralement payante. Nous verrons plus loin, comment la déployer.
- Un catalogue, puisque nous avons choisi d'en utiliser un. (Cf. paragraphe précédent)

Lors d'une sauvegarde ou restauration RMAN établit une connexion simultanée entre la base cible et son catalogue. En outre, le logiciel de sauvegarde et RMAN communiquent via la MML

Dans le dessin qui suit, illustrons notre propos :

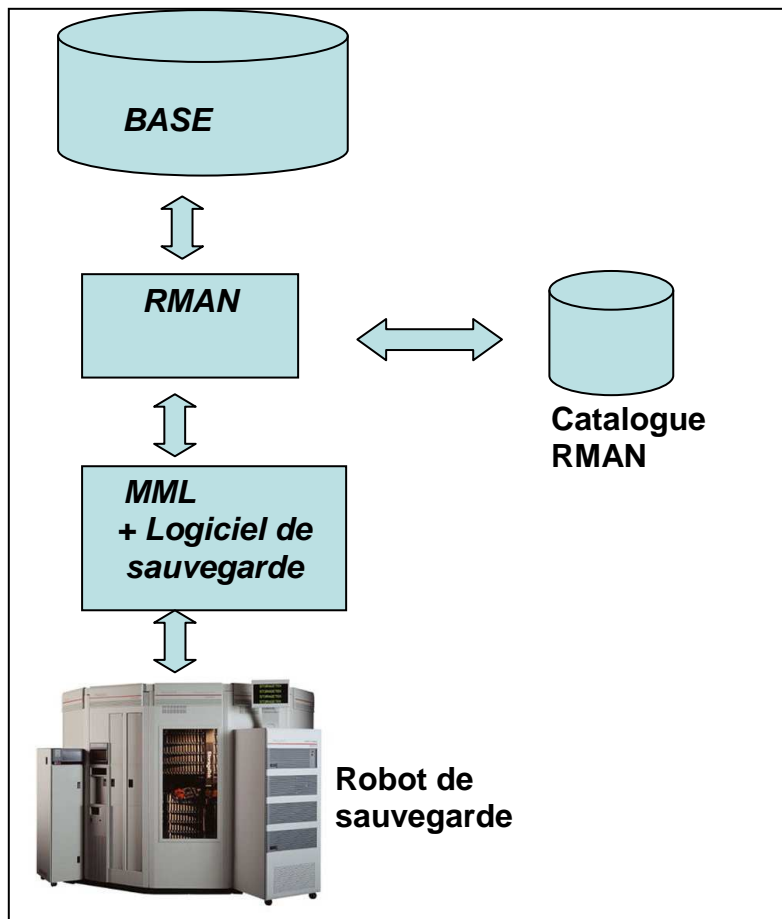


Figure 28: Composants principaux de l'outil RMAN

Que retenir de l'architecture de l'outil RMAN ?

L'étude met en évidence 2 point importants :

- RMAN communique avec un logiciel de sauvegarde (ici DataProtector) via la MML, c'est à dire un programme appelé agent. Il va donc falloir apprendre à déployer puis mettre en œuvre cet agent.
- Dans notre projet, il a été décidé d'utiliser systématiquement les catalogues RMAN car ils offrent de multiples avantages (utilisation de scripts stockés, sécurité accrue) qui vont contenir toutes les informations utiles aux sauvegardes et restauration. En

conséquence, il va falloir créer et administrer une base pour y créer et y paramétrer le(s) catalogue(s).

Pour le projet, il nous faudra répondre aux questions suivantes :

- comment permettre la communication entre DataProtector et RMAN ?
- comment créer et paramétrer et organiser le(s) catalogue(s) nécessaire(s) ?

4.2 MISE AU POINT DE LA MAQUETTE

Dans la partie précédente, nous avons collecté de précieuses informations, qui ont permis d'entrevoir le fonctionnement de chaque composant.

Nous allons maintenant répondre à la question : que modifier pour permettre l'intégration de l'outil RMAN dans l'infrastructure existante ?

4.2.1 Description et méthodologie de mise au point de la maquette

4.2.1.1 Composition et intérêt d'une maquette

Pour se prémunir de toute perturbation de la chaîne de sauvegarde lors de tests, il a été décidé de partir d'une maquette de sauvegarde déjà existante, comprenant un environnement de test complet VTOM et DataProtector ainsi qu'une robotique.

La maquette va permettre d'évaluer les impacts pour introduire RMAN et de rédiger les procédures nécessaires à la mise en œuvre de RMAN

Le schéma simplifié de notre maquette correspondant à la nouvelle infrastructure cible avec RMAN :

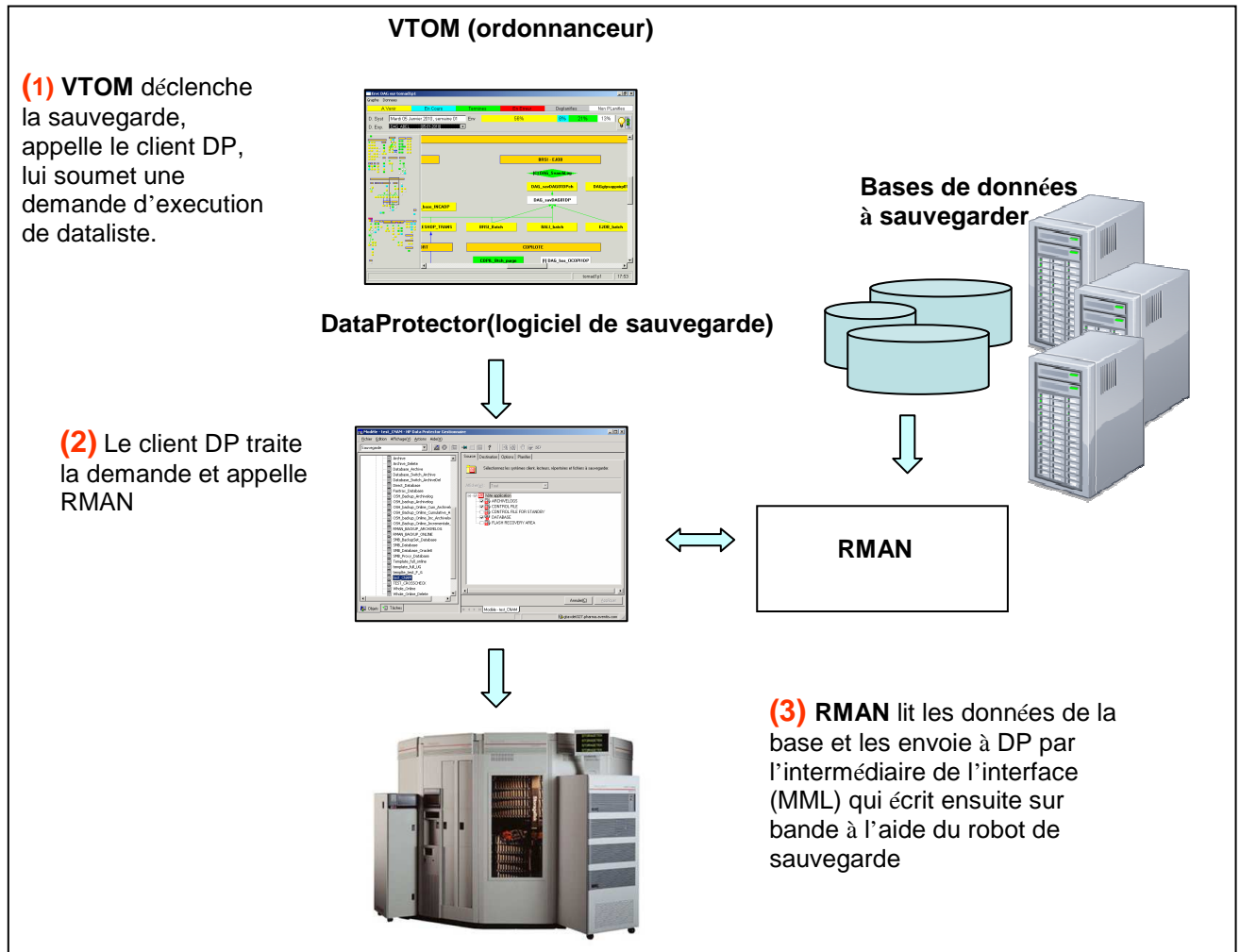


Figure 29: Description simplifiée de la chaîne de sauvegarde cible (avec RMAN)

4.2.1.2 Démarche de construction de la maquette

Pour construire notre environnement de maquette, une démarche pragmatique a été suivie. Chaque composant (RMAN, DataProtector, et VTOM) a été étudié séparément, puis les éléments ont été intégrés, un à un, pour aboutir au final à la mise au point d'une chaîne de

sauvegarde complète. Les modifications nécessaires pour passer à RMAN ont été évaluées et testées directement sur la maquette et ont aboutit à la rédaction de procédures. Ces procédures ont finalement été mises en œuvre et évalué lors d'un pilote et sur des bases Oracle réellement utilisées.

4.2.1.3 Modifications à apporter pour intégrer RMAN

Pour intégrer RMAN, de nombreuses modifications de l'existant sont nécessaires. Il faut successivement :

- Préparer l'outil RMAN c'est-à-dire créer et paramétrer le ou les catalogues RMAN
- Permettre la communication entre RMAN et DP en installant la MML.
- Créer de nouveaux modèles et nouvelles spécifications de sauvegarde dans le logiciel DataProtector
- Revoir les scénarios pour permettre les sauvegardes incrémentales, reprendre le script de lancement « shob2sauv », ainsi que les enchaînements de tâches ordonnancées sous VTOM.

4.2.2 Préparation de l'outil RMAN

Nous avons choisi d'utiliser un catalogue RMAN. (cf. plus haut), il faut donc savoir en créer un. Selon la documentation mise à disposition par l'éditeur Oracle, il faut procéder de la façon suivante :

- tout d'abord disposer d'une base de données Oracle.
- créer un schéma qui va héberger le catalogue grâce au lancement d'une requête SQL de la forme « `create user ...` ». Il est nécessaire d'attribuer à ce schéma des droits particulier

- créer le catalogue (sous rman la commande « *create catalog...* ») puis le paramétrer (sous RMAN avec les commandes « *configure ...*»)
- Enfin, il va falloir enregistrer la base dans ce catalogue. (Sous RMAN avec la commande « *register database* »)

4.2.3 Mise en place de l'interface de communication DataProtector / RMAN

Une fois le catalogue prêt à l'emploi, il faut mettre en place le module d'interface (nommé aussi selon les cas, librairie, agent DP, ou MML) qui permette le dialogue entre DP et RMAN.

La mise en place de cette interface est décrite en détail dans le manuel « HP Openview Storage DataProtector ». Mais, concrètement, le déploiement se fait en 2 parties principales :

- La distribution de l'agent spécifique RMAN/DP via la console d'administration DP.
- L'interfaçage Oracle à Data Protector grace à une commande UNIX « *ln -s* »

Déploiement de l'agent RMAN pour DP

Le déploiement de l'agent consiste à copier les fichiers nécessaires à DP sur le serveur qui héberge les instances à sauvegarder, de fait à partir de la console d'administration DataProtector. Le déploiement se fait aisément au moyen de quelques clics de souris. L'opération de déploiement consiste principalement à :

- Sélectionner les serveurs cibles
- Sélectionner l'agent « intégration Oracle », requis pour RMAN

On lance ensuite le déploiement qui ne dure que quelques secondes.

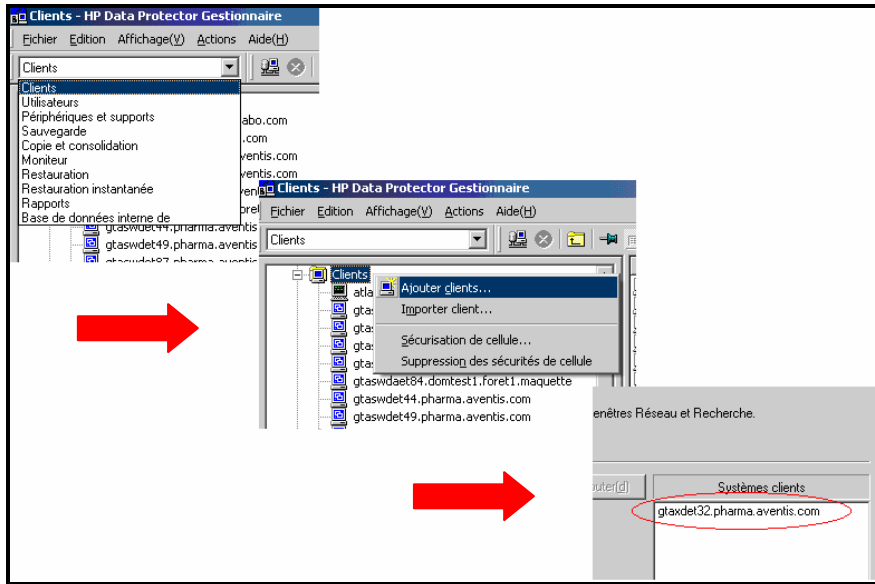


Figure 30: Déploiement de l'agent RMAN pour DP avec la console d'administration

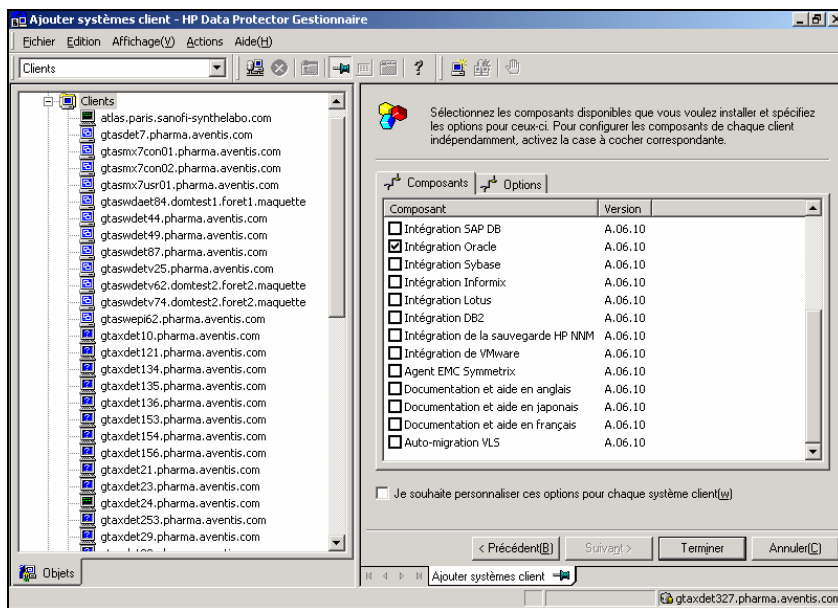


Figure 31: Illustration du déploiement de l'agent RMAN pour DataProtector (suite)

L'opération de déploiement de l'agent est donc une opération simple, et offre l'avantage de permettre un déploiement de multiples agents sur plusieurs serveurs en une seule fois.

Interfaçage DP/RMAN

Le seul déploiement de l'agent RMAN pour DP ne suffit pas. Il faut aussi indiquer à Oracle et RMAN où se trouve la librairie précédemment déposée, avant de pouvoir l'utiliser. Concrètement, il suffit de créer un lien symbolique UNIX, en utilisant la commande « ln -s ».

Pour mettre en place ce lien (sous UNIX) : il suffit de

- Arrêter toutes les instances Oracle
- Créer le lien symbolique, par une commande UNIX de la forme:

```
ln -s /opt/omni/lib/libob2oracle8_64bit.sl libobk.sl
```

- Redémarrer les instances.

Notons que, le nom exact de la librairie, le répertoire hébergeant la librairie, ainsi que l'extension du fichier peuvent changer selon la version du SGBD, ou du système d'exploitation.

Lier Oracle à DP est donc une opération simple, mais on remarque que la mise en place du lien implique un arrêt et une relance des bases Oracle. Ce point peut être problématique, lorsque l'on souhaite migrer les sauvegardes vers RMAN car il implique l'accord des utilisateurs des applications utilisant ces bases.

4.2.4 Mise au point des nouveaux modèles et spécifications de DataProtector

Le logiciel DP est prêt à être utilisé avec RMAN, nous allons maintenant mettre en place des sauvegardes. Dans DP, comme pour les autre type de sauvegarde (SAP, exchange ...), la mise en place d'une sauvegarde sous DP se réalise en 2 temps :

- Tout d'abord, il est nécessaire de créer une « **Template** », c'est-à-dire un **modèle de sauvegarde**. DataProtector dispose de modèles prédéfinis RMAN qu'il est possible de modifier et d'adapter aux besoins.
- Ensuite, à partir du modèle RMAN choisi, on définit une **spécification de sauvegarde** qui héritera des caractéristiques du modèle.

Création d'un modèle de sauvegarde RMAN

Pour créer un modèle, on utilise à nouveau l'outil d'administration DP en cliquant sur « Serveur Oracle » puis « Ajouter modèle ». On obtient la fenêtre qui suit, et on choisit ce que l'on souhaite sauvegarder. (Ici fichiers d'archivage, fichiers de contrôle, fichiers de la base). On peut aussi renseigner d'autres informations telles que les lecteurs de sauvegarde.

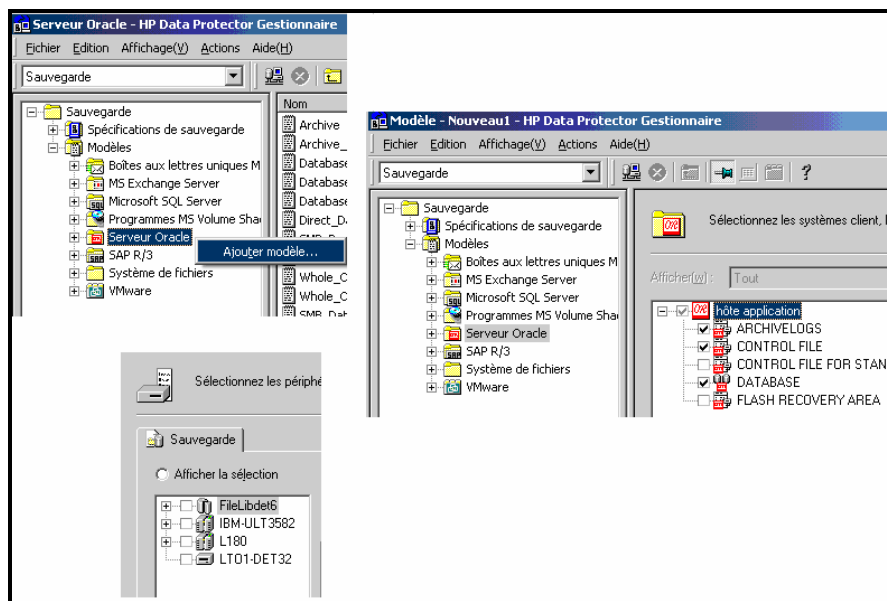


Figure 32: Création d'un modèle de sauvegarde

Dans l'un des onglets, il est possible de vérifier le code RMAN qui est lancé par DP pendant la sauvegarde. Plus exactement, il ne s'agit pas d'un vrai script RMAN, mais d'un pseudo code très proche du code qui sera effectivement soumis à RMAN. Il est possible de le modifier, mais la syntaxe acceptée est bien plus limitée : certaines instructions RMAN ne sont pas reconnues.

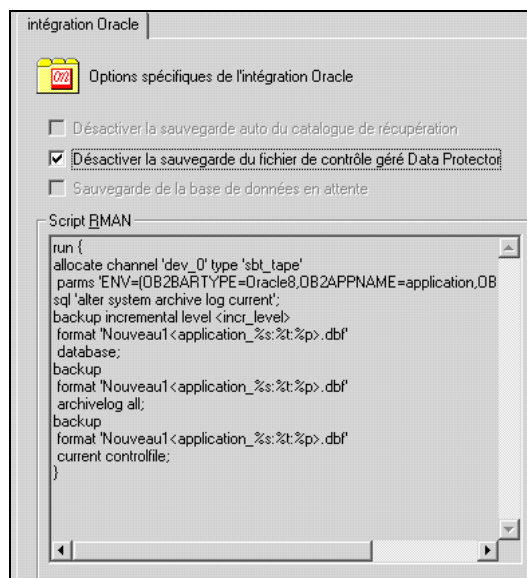


Figure 33: Pseudocode RMAN du modèle

Une fois toutes les informations nécessaires saisies, il ne reste plus qu'à donner un nom à notre modèle, et à le sauvegarder.

Création d'une spécification de sauvegarde

Le modèle de sauvegarde RMAN créé, nous allons nous en servir pour déclarer une spécification de sauvegarde qui va hériter de toutes les caractéristiques du modèle.

Comme pour les autres modèles (SAP, système de fichiers ...), à partir de la console d'administration, il d'exécuter l'assistant de création, puis de se laisser guider.

Pour cela, Il suffit de se positionner dans le menu de sauvegarde et de cliquer sur « ajouter sauvegarde », puis sélectionner un modèle, comme l'illustrent les copies d'écrans ci-joints.

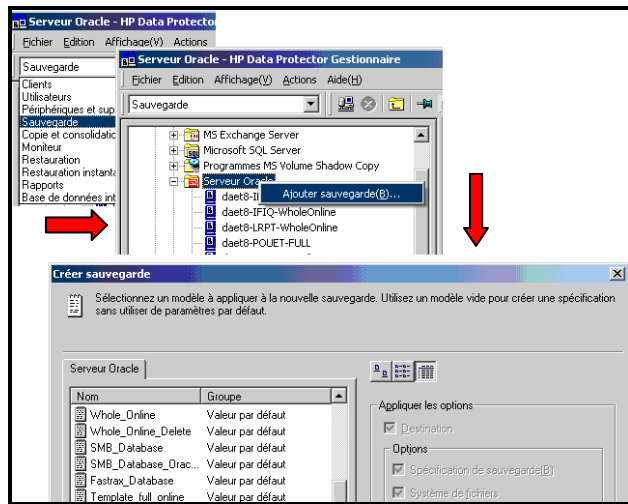


Figure 34: Création d'une spécification de sauvegarde avec la console d'administration

Lors des sauvegardes, RMAN doit se connecter à la fois à son catalogue et à la base à sauvegarder.

Il faut renseigner les informations de connexion au catalogue RMAN et les informations nécessaires à la connexion de la base (User SYS).

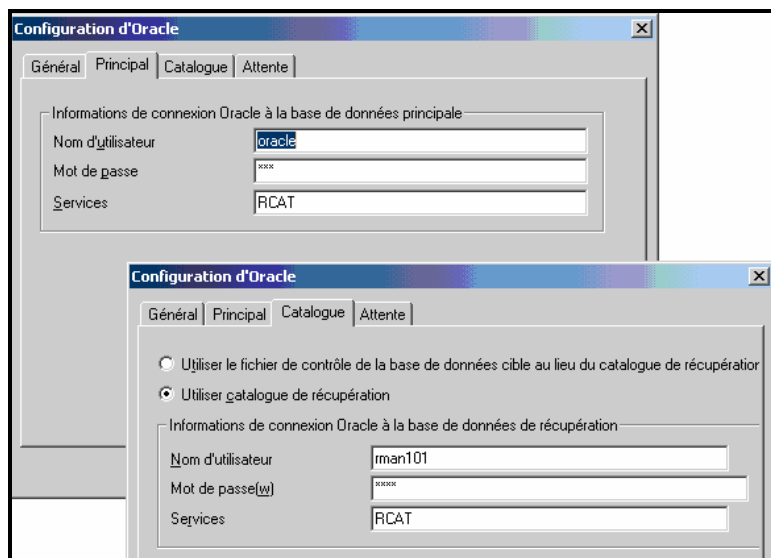


Figure 35: Informations de connexion nécessaires à la spécification de sauvegarde

Comme à la création du modèle, il est possible de visualiser et modifier le « pseudo-code » RMAN, très proche du code compréhensible RMAN, mais plus limité dans sa syntaxe.

Pour les besoins du projet, nous avons développé 3 modèles à partir de modèle préexistants :

- un pour les sauvegardes à chaud
- un pour les sauvegardes à froid
- un pour la sauvegarde des archives.

Notons que ces modèles permettent de lancer des sauvegardes incrémentales ou complètes.

Contraintes et limitations du logiciel DataProtector

On notera quelques limitations et particularités de DataProtector :

- Les modèles et les spécifications ne supportent pas l'ensemble de la syntaxe RMAN.
- DataProtector n'autorise pas un mélange de sauvegardes RMAN en mode catalogue, avec d'autres en mode fichier de contrôle. (Heureusement, il a été décidé que seul le mode catalogue serait utilisé.)
- Après chaque sauvegarde RMAN, DataProtector réalise un export du catalogue, puis l'écrit sur bande. En conséquence, le temps de sauvegarde total d'une base est le temps de sauvegarde de la base elle-même, plus le temps d'export du catalogue, auquel on ajoute le temps d'écriture sur bande de l'export. L'export automatique du catalogue induit plusieurs conséquences: plus le catalogue est volumineux, plus le temps de sauvegarde total augmente. Ce qui signifie aussi que si la base à sauvegarder est petite, l'export et la sauvegarde du catalogue peuvent représenter une part importante dans le temps total de sauvegarde d'une base.

On notera aussi quelques bugs de DataProtector, dont il faut tenir compte :

- Le mode de sauvegarde automatique des fichiers de contrôle provoque des erreurs.
- Certains caractères spéciaux dans les mots de passe, par exemple le point virgule « ; » entraînent des erreurs et génèrent des messages d'avertissement non pertinents.

4.2.5 Mise au point des nouveaux enchaînements VTOM

L'arrivée de RMAN entraîne de nombreux changements au niveau de l'ordonnancement :

- Le script lanceur des sauvegardes doit être mis à jour pour prévoir le nouveau type de sauvegarde RMAN.
- Il faut redéfinir les scénarios pour prendre en compte le mode de sauvegarde incrémental.
- Enfin, il va falloir mettre au point des nouveaux enchaînements ordonnancés tenant compte des changements de scénarios.

4.2.5.1 Modifications des scénarios de sauvegardes

Nous avons déjà vu, que quel que soit le type de données à sauvegarder (Fichiers, ERP SAP, messagerie Exchange, etc.), les traitements de sauvegarde appellent invariablement un seul et même script UNIX, nommé « *shob2sauve* », chargé de lancer des sauvegardes.

Pour permettre le lancement de sauvegarde RMAN, la modification du script est très simple puisqu'elle consiste à ajouter un bloc de quelques lignes, qui ajoute un nouveau type de sauvegarde. La modification permet le lancement la commande « *omnib* » avec les paramètres adéquats pour RMAN.

Le bloc suivant a été ajouté :

```
RMAN[rman)  MODE=mode
            TYPE=datalist
            case ${MODELIST} in
              FULL) MODELIST=FULL ;;
              INCREMENTAL1) MODELIST=INCREMENTAL1 ;
              *) echo "incoherence entre le type de Sauve $TYPE et           le mode $MODELIST"
                exit 2 ;;
            esac
```

Figure 36: Lignes de code à ajouter dans le script UNIX qui lance les sauvegardes DP

4.2.5.1 Modifications des scénarios de sauvegardes

Avant de procéder aux modifications des tâches de sauvegardes dans l'ordonnanceur, il est nécessaire de revoir les scénarios de sauvegardes.

J'ai proposé au chef de projet les modifications de scénarios qui me semblaient nécessaires, puis à l'équipe administration des sauvegardes. Nous les avons fait valider par l'équipe chargée de l'ordonnancement des tâches de sauvegarde.

Cette action n'a pas été longue, car au final, les modifications nécessaires sont peu nombreuses et découlent immédiatement du cahier des charges :

- Le nouveau scénario de sauvegarde à froid RMAN se présente de la façon suivante : on réalise un export base fermée et une sauvegarde de la base chaque jour. Les sauvegardes hebdomadaires et mensuelles sont complètes, les sauvegardes journalières peuvent être complètes ou incrémentales cumulatives.
- Le scénario de sauvegarde à chaud change de la manière suivante avec RMAN : on réalise un export et une sauvegarde de la base chaque jour, à chaud. Les sauvegardes hebdomadaires et mensuelles sont complètes, les sauvegardes journalières peuvent être selon les besoins complètes ou incrémentales cumulatives.

On remarque un seul véritable changement : en semaine les sauvegardes peuvent être lancées avec le mode incrémental selon les besoins.

4.2.5.1 Modifications de l'ordonnancement

Pour tenir compte des modifications de scénarios (les sauvegardes à chaud sont désormais cumulatives la semaine), les anciens enchaînements de tâches de sauvegardes (cf. paragraphe 4.1.2.1) ont dû revus de la manière suivante :

Nouveau scénario de sauvegarde RMAN à chaud

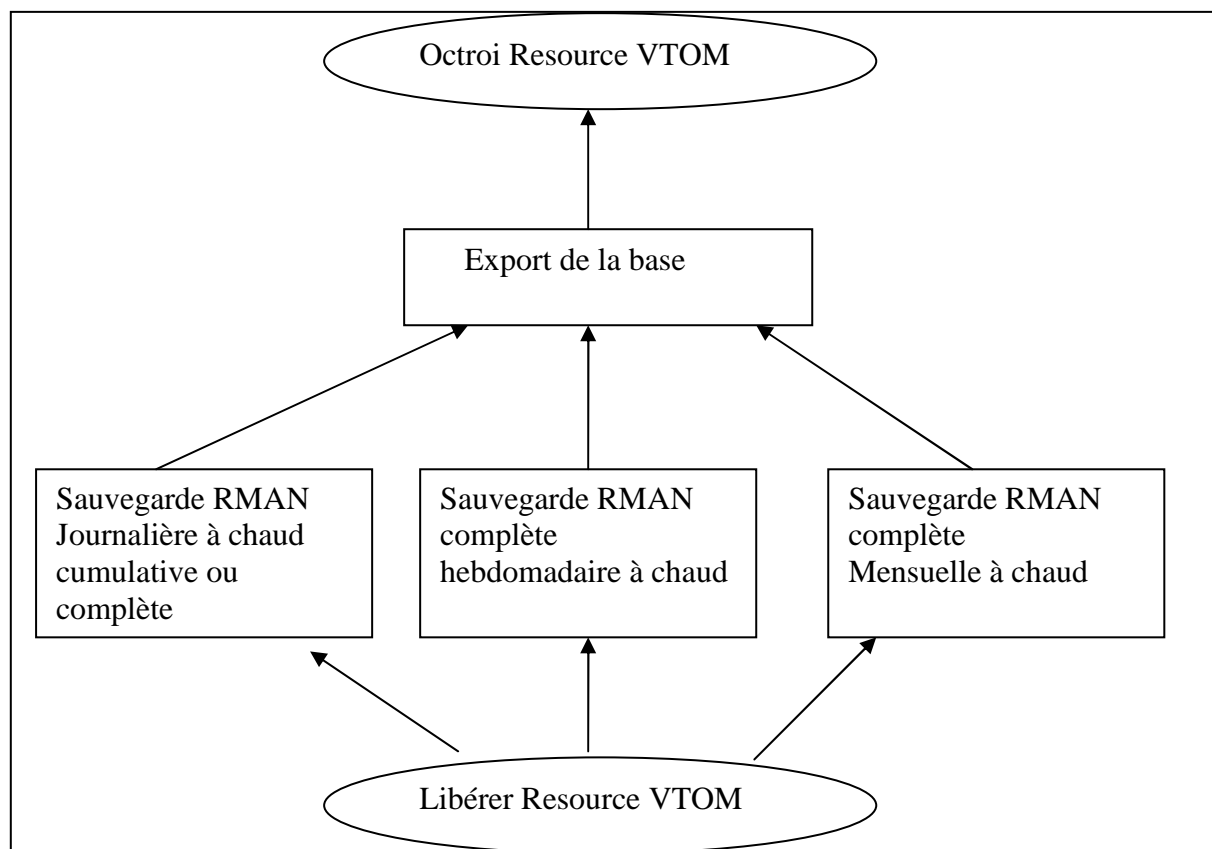


Figure 37: Scénario de sauvegarde à chaud RMAN

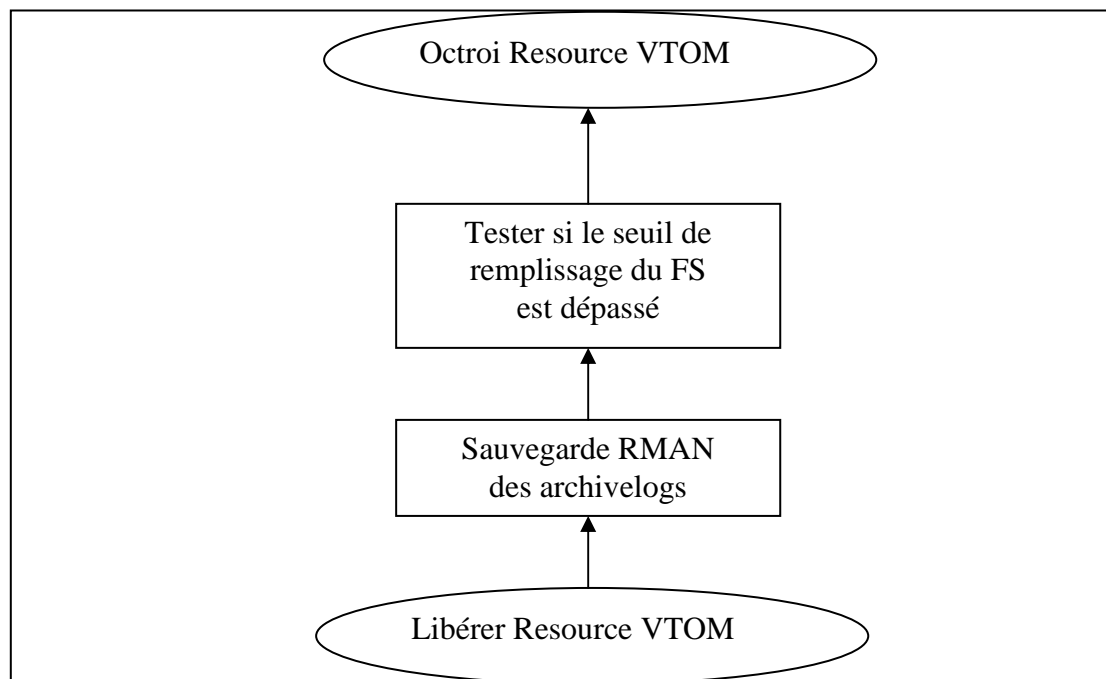


Figure 38: Scénario de sauvegarde des archives avec RMAN

Nouveau scénario de sauvegarde RMAN à froid

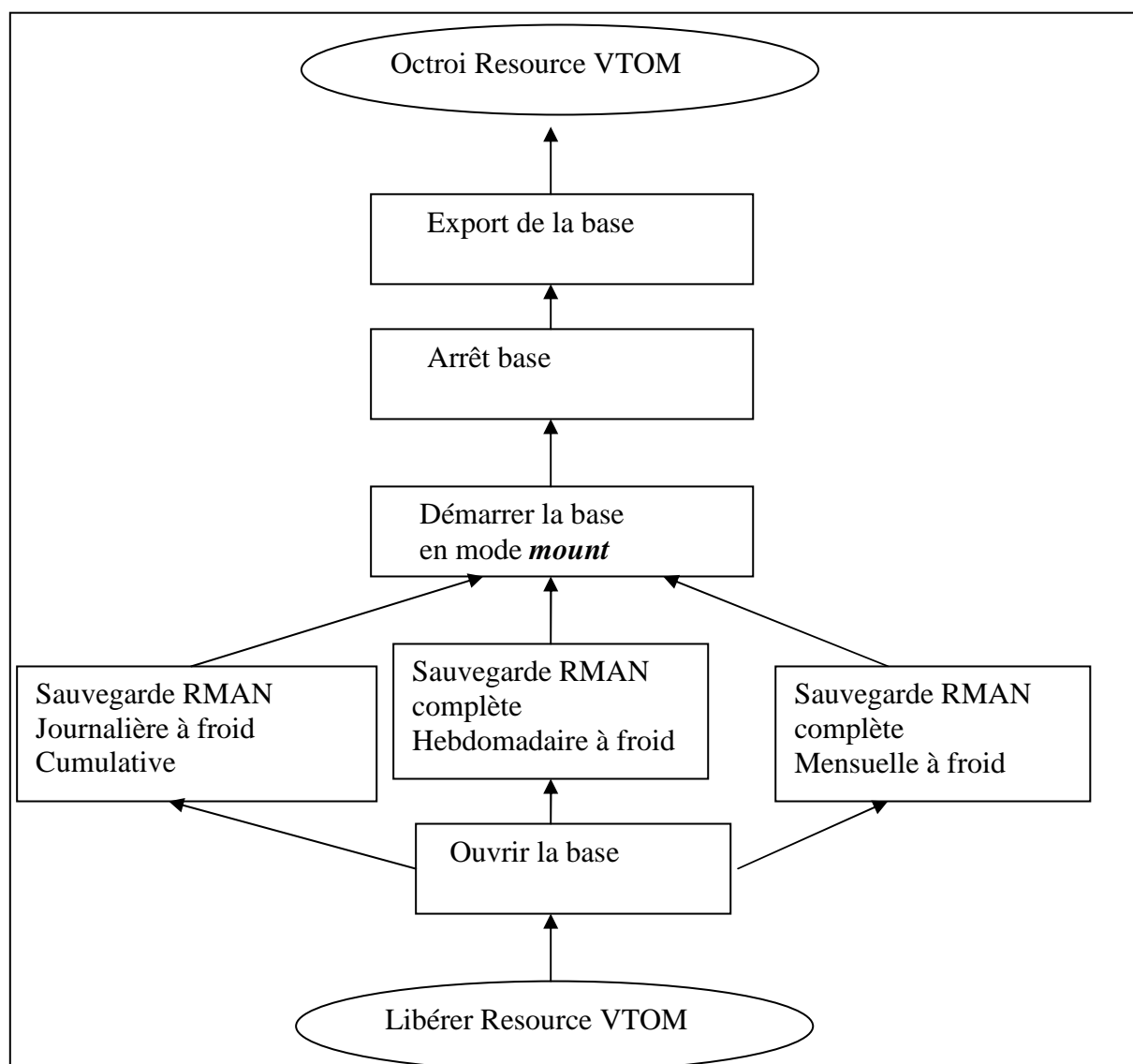


Figure 39: Scénario de sauvegarde à froid RMAN

Quelles sont les modifications apportées à l'ordonnancement des tâches de sauvegarde ?

Les sauvegardes RMAN ne peuvent se faire que si son catalogue RMAN est disponible. C'est pourquoi des « ressources VTOM », sortes de verrous applicatifs, ont été ajoutés. Ils bloquent temporairement une sauvegarde RMAN pendant la sauvegarde de l'instance qui héberge le catalogue.

Avec RMAN, la sauvegarde à froid fonctionne uniquement en mode « *mount* » (la base de données n'est donc plus complètement stoppée). Le scénario de sauvegarde à froid, prend désormais en compte un redémarrage ce mode « *mount* » avant chaque sauvegarde à froid.

RMAN simplifie l'ordonnancement :

- La sauvegarde des archives nécessite moins de tâches. (cf. ancien et nouvel enchaînement des fichiers d'archivage).
- Dans l'ancien mode de sauvegarde (sans RMAN), il était nécessaire de régénérer avant chaque sauvegarde la liste de fichiers de la base à sauvegarder car cette liste pouvait évoluer, par exemple dans le cas d'ajout ou suppression de fichiers de bases. Un traitement ordonnancé était donc lancé avant chaque sauvegarde pour modifier les spécifications de sauvegarde DataProtector ; et y inclure la liste des fichiers. RMAN ayant déjà toutes les informations nécessaires dans le fichier de contrôle de la base, ainsi que dans son catalogue, il n'est donc plus nécessaire de régénérer automatiquement les spécifications de sauvegarde.

4.2.6 Tâches réalisées pour la mise au point de la maquette

Pour mettre au point la maquette, toute la documentation nécessaire a été préalablement récupérée. Les composants logiciels d'infrastructure de sauvegarde ont ensuite été étudiés puis testés sur la maquette.

Il a fallu que j'interface la librairie RMAN pour DataProtector, puis que je crée un catalogue comme cité plus haut. J'ai dû installer plusieurs moteurs Oracle puis créer plusieurs bases de données Oracle de test (en version, 8, 9, 10, et 11.) à sauvegarder.

Le fichier « *shob2sauv* » a été modifié pour pouvoir accepter la sauvegarde RMAN. Les scénarios de sauvegardes ont été revus pour permettre le mode de sauvegarde incrémentale. Les trois nouveaux enchaînements VTOM ont été modifiés (chaud, froid, fichiers d'archivage).

Enfin, trois modèles de sauvegarde RMAN et des spécifications de sauvegarde ont été créés dans DataProtector.

Pour tester la nouvelle chaîne de sauvegarde, la collaboration avec les diverses équipes (Exploitation, sauvegarde notamment) a été nécessaire. J'avais à ma disposition un environnement complet du logiciel de sauvegarde DataProtector, et un environnement VTOM. J'étais quasiment autonome sur la partie DataProtector : je pouvais lancer des créations de sauvegarde, et des tests de sauvegardes/restaurations. Cependant, je ne disposais que de droits en lecture sur VTOM, si bien que les tests d'ordonnancement ont été réalisés par l'équipe en charge de l'outil.

Les premières sauvegardes et restaurations ont d'abord été lancées depuis l'invite de commande RMAN. (Sans passer par la console DataProtector). Ces tests ont permis le paramétrage de l'outil RMAN.

Ensuite, les sauvegardes ont été déclenchées depuis la console d'administration de DataProtector. Cela a permis de tester l'intégration de RMAN à DP et a conduit à la mise au point de trois modèles et spécifications des sauvegardes (sauvegardes à chaud, à froid et des fichiers d'archives).

De nombreux tests sur la maquette ont été réalisés pour en vérifier le fonctionnement de bout en bout. Ces tests de sauvegarde et restauration sont des tests longs et fastidieux à réaliser compte tenu du nombre et de la durée des sauvegardes.

Pour valider l'ensemble de la nouvelle chaîne sauvegarde (RMAN / DP /VTOM), des sauvegardes déclenchées par VTOM ont été lancées.

Les tests de restauration ont montré qu'il n'était pas aisé de restaurer depuis la console d'administration DataProtector mais qu'il était préférable de restaurer depuis l'invite de commande RMAN. Compte tenu de la complexité de l'outil, seul un administrateur de bases de données qualifié sur l'outil RMAN, est en mesure de restaurer efficacement.

Finalement, ces tests ont permis de rédiger toutes les procédures nécessaires à la mise en place de sauvegarde RMAN. Compte tenu des exigences réglementaires, le temps consacré à

la documentation a été particulièrement long. Les documents ont du suivre un processus d'écriture, validation, correction, publication particulièrement rigoureux contrôlé par un membre de l'équipe chargée du contrôle qualité.

4.3 PILOTE PHENIX

En phase maquette, nous avons mis au point les procédures nécessaires à l'intégration du composant RMAN dans notre chaîne existante. Avant de déployer la solution RMAN sur un large périmètre, il est important de les tester et de les vérifier sur un périmètre réduit. C'est pourquoi on a procédé à une phase préliminaire dite pilote pour une application nommée « *PHENIX* ». Cette étape a consisté à sauvegarder avec RMAN sur des nouvelles bases oracle pour deux environnements distincts : la pré-production et la production.

4.3.1 Objectifs du pilote PHENIX

L'objectif principal de ce pilote est de prouver qu'il est possible d'intégrer RMAN dans l'exploitation existante, avec un coût raisonnable, sans effet indésirable. Le projet PHENIX a été choisi car il offrait un certains nombres d'avantages :

- C'est un nouveau projet, il n'y a donc pas de modifications à apporter aux sauvegardes existantes.
- C'était un projet qui prenait en compte de nouvelles normes d'exploitation. Cela a permis de valider l'impact de RMAN avec ces nouvelles normes.
- Des nouvelles technologies ont été mises en place (Le mode flashback et le double archivage)
- Enfin, le projet est soumis à la réglementation « GxP », ce qui nécessite une procédure rigoureuse et la qualification des environnements.

En revanche, ce pilote utilise un catalogue dédié, car le serveur qui devait héberger les catalogues RMAN n'était pas encore disponible.

4.3.2 Démarche utilisée pour le pilote PHENIX

Le pilote PHENIX a consisté à déployer et corriger les procédures écrites lors de la phase de maquette.

Le pilote s'est déroulé en 3 étapes :

- Dans un premier temps, l'infrastructure a été préparée : il faut permettre la communication entre DataProtector et RMAN, puis créer les catalogues RMAN.(En environnement de pré-production puis en production)
- Dans un deuxième temps les sauvegardes ont été définies dans DataProtector et ordonnancées dans VTOM.
- Dans un dernier temps les sauvegardes et restaurations ont été validées.

Etape préparatoire :

- L'agent RMAN pour DataProtector a été déployé.
- Deux instances ont été créées héberger les catalogues : une pour loger le catalogue de la base de production, une deuxième pour loger le catalogue de pré-production. Ces bases ont été sauvegardées avec l'ancienne méthode. (Les catalogues seront migrés plus tard dans une base qui hébergera l'ensemble des catalogues RMAN.)
- Les bases de données y ont été enregistrées dans leurs catalogues respectifs, et le paramétrage RMAN a été mis en place.
- Les trois modèles de sauvegardes prévus pour DataProtector ont été mis en place : à chaud, froid et archivage. Ils permettent le lancement de sauvegardes RMAN de type « cumulatifs » ou « complets».

Etapes de préparation des sauvegardes

- Les spécifications de sauvegarde nécessaires ont été définies à partir des modèles.
- Le scénario de sauvegarde à chaud, ainsi que celui de l'archivage, ont été ordonnancés sous VTOM. Le week-end, et en fin de mois, des sauvegardes complètes sont lancées, et la semaine des sauvegardes cumulatives.

Etape de tests et validation

- Des rapports d'installation ont été rédigés, et les procédures ont pu ainsi être validées et publiées. Des cas de test ont été réalisés : test de l'installation, test de sauvegardes test de restauration. Le projet PHENIX étant soumis à la réglementation GPX, toutes les actions réalisées ont été tracées et communiquées à l'équipe projet. Toutes les procédures utilisées ont été publiées dans l'outil de gestion de document.
- Un test de restauration de l'instance a été fait, sur un serveur de test et s'est effectué sans encombre.

4.3.3 Retour d'expérience sur le pilote PHENIX

D'une manière générale, le pilote PHENIX s'est globalement bien déroulé selon les prévisions.

Une relative simplification des sauvegardes a été notée, alors qu'au contraire les restaurations sont devenues plus complexes. Ces dernières ne peuvent plus être réalisées par l'équipe en charge des sauvegardes. En effet, pour être efficace elle doit se réaliser dans l'invite de commande RMAN. Elle nécessite une très bonne connaissance d'Oracle, et une formation spécifique à l'outil RMAN.

L'intégration de RMAN dans la chaîne de sauvegarde n'a pas posé de problème technique majeur. Il a suffi de dupliquer le travail fait en maquette.

Cependant, quelques validations et vérifications et ajustements ont été nécessaires.

Un double archivage (redondance des fichiers d'archivage) a été mis en œuvre pour la première fois dans ce projet. Pour notre projet, RMAN supporte ce mode de manière transparente. Les tests ont montré que RMAN sauvegarde et efface correctement les fichiers d'archivage (fichiers d'archivage duplexés), comme attendu.

Ensuite, une fonctionnalité de récupération rapide appelée « *flashback* » a aussi été activée. Elle permet, sans restauration préalable, de ramener instantanément la base à l'état où elle était à un moment du passé. Les tests de restauration ont montré que ce mode devait être provisoirement désactivé pendant la récupération de la base. Les procédures de restauration ont été modifiées en conséquence.

Un temps de sauvegarde anormalement long des fichiers d'archivage a été constaté. Une analyse a donc été faite et a conduit à retoucher les modèles de sauvegardes. En effet, à la demande de l'équipe projet la base a été paramétrée pour générer un grand nombre de fichiers d'archive alors que la spécification de sauvegarde était paramétrée pour sauvegarder les archives par lot contenant un petit nombre de fichiers d'archive. Il a fallu modifier un paramètre « *filperset* » dans les modèles DataProtector. En effet, le logiciel DataProtector sauvegarde par groupes de fichiers, et le paramètre « *filperset* » désigne le nombre de fichiers sauvegardé simultanément. Lorsque le mode archivage est activé, il suffit d'augmenter ce paramètre pour accroître le nombre d'archives dans ces lots, ce qui permet alors de réduire les va-et-vient sur le lecteur de bande. Concrètement, dans la console d'administration DataProtector, il a fallu modifier le pseudo code RMAN du modèle et de la spécification de sauvegarde.

Le test de restauration a mis en évidence la difficulté de déterminer les bandes nécessaires à une restauration, et plusieurs bandes ont provisoirement manqué pendant le test. Une assistance supplémentaire à l'équipe d'administration des sauvegardes a été nécessaire. De plus les formulaires de demande de restaurations ont été modifiés et séparés en deux parties distinctes : la restauration du catalogue et la restauration de la base de données.

Quelques tests empiriques de comparaison de l'ancien mode de sauvegarde (sans RMAN), avec le nouveau (RMAN) ont été entrepris. Le temps de sauvegarde RMAN n'est pas meilleur et même légèrement plus long. En effet, DP sauvegarde automatiquement le catalogue, après avoir sauvegardé les données de la base. Si bien que le temps de sauvegarde total doit inclure le temps de sauvegarde RMAN, le temps d'export du catalogue et aussi le temps de sauvegarde du catalogue. C'est pourquoi, RMAN n'améliore pas le temps de sauvegarde des petites bases. Pour cette raison, la base de notre pilote PHENIX étant très petite au démarrage du projet, RMAN n'a pas amélioré son temps de sauvegarde.

Le pilote a montré que la sauvegarde des catalogues devait être étudiée : créer, maintenir et sauvegarder une base pour chaque catalogue RMAN peut s'avérer une tâche lourde et fastidieuse. Il va falloir réfléchir au moyen d'optimiser l'organisation des catalogues.

Le pilote a permis de valider les procédures développées en phase maquette, et a prouvé que RMAN pouvait être intégré à la chaîne de sauvegarde. Seuls quelques réglages mineurs ont été nécessaires. On notera que suite à un contretemps, le catalogue RMAN a été provisoirement temporairement hébergé dans une base.

4.4 RESUME DU CHAPITRE 4

Nous avons réuni la documentation nécessaire et une maquette a été réalisée pour valider que RMAN pouvait être introduit dans la chaîne de sauvegarde existante.

Chaque composant (RMAN, DataProtector, et VTOM) a été étudié séparément, puis les éléments ont été intégrés, un à un, pour aboutir au final à la mise au point d'une chaîne de sauvegarde complète.

Les deux scénarios et enchaînements existants VTOM (sauvegarde à chaud, froid), ont été modifiés pour prendre en compte les sauvegardes RMAN cumulatives. La modification d'un seul script, nommé « shob2sauv » a été nécessaire. Il a suffi d'ajouter un bloc de quelques lignes pour prendre en compte le nouveau type de sauvegarde RMAN.

La phase de maquette a donné lieu à une série de tests de sauvegardes et restaurations, et à la rédaction de la documentation (procédures d'installation, des sauvegardes, restauration, cas de tests, etc.).

La maquette a été suivie d'un premier pilote nommé PHENIX qui a conduit à la mise en place de RMAN pour de nouvelles bases, ce qui a permis de valider que RMAN pouvait être intégré dans la chaîne de sauvegarde existante, sans perturbation. Toutes les procédures y ont été validées.

Chapitre 5: GENERALISATION ET DEPLOIEMENT

En phase maquette, et lors du pilote PHENIX, les sauvegardes RMAN n'ont été mises en œuvre que sur des environnements ne comprenant qu'un petit nombre de bases nouvellement créées. Dans ce chapitre, nous déploierons la solution sur un large périmètre composé de centaines de bases, dans une large gamme de versions de base, de systèmes d'exploitation, et de serveurs. Toutes ces bases sont déjà existantes, et il va falloir migrer ces sauvegardes vers RMAN.

Ce chapitre est décomposé en 5 parties :

- Dans la première partie, nous verrons comment optimiser la solution en organisant les catalogues, en automatisant le déploiement et en cherchant à améliorer la robustesse en cas de panne.
- Ensuite, nous verrons comment ont été réglés les problèmes de validation de la solution sur un périmètre aussi important, avec de multiples versions de bases, de systèmes et de serveurs.
- Dans une troisième partie, nous réaliserons la préparation de l'infrastructure puis la formation des divers acteurs.
- Nous évoquerons ensuite un pilote, nommé SHIVA visant à valider la migration vers RMAN de sauvegardes existantes.
- Enfin nous verrons comment le déploiement final a été préparé puis piloté.

5.1 OPTIMISATION

L'étape d'optimisation est nécessaire car elle permet de gagner du temps, de l'argent. Elle augmente la fiabilité de la solution et celle du déploiement. Elle est d'autant plus importante que le périmètre concerné est large.

5.1.1 Démarche suivie pour optimiser la solution

L'optimisation s'est fait à 3 niveaux :

- Nous avons d'abord optimisé la maintenance de la nouvelle chaîne et spécialement celle des catalogues en les organisant de façon rationnelle.
- Nous avons ensuite amélioré le temps de déploiement en automatisant la mise en œuvre.
- Finalement, nous avons renforcé la robustesse de la nouvelle infrastructure (avec l'outil RMAN) grâce à une meilleure architecture matérielle et logicielle.

Nous pouvons résumer la démarche d'optimisation par le schéma suivant :

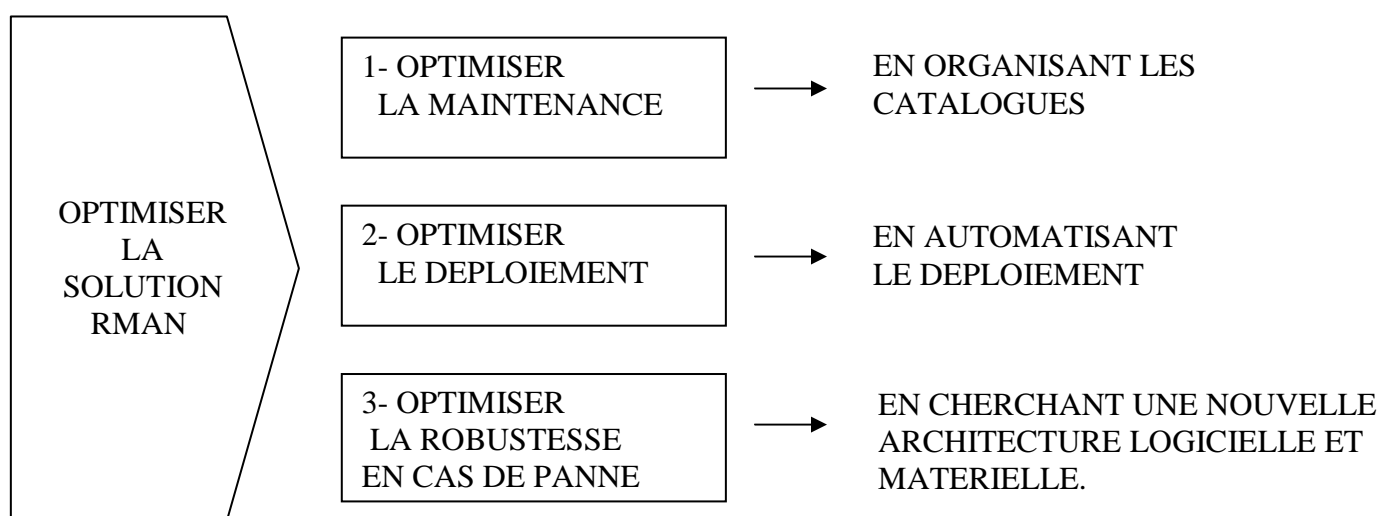


Figure 40: Démarche d'optimisation de la solution RMAN

5.1.2 Etude de l'organisation des catalogues

Pour un nombre limité de bases à sauvegarder, la façon d'organiser les catalogues n'a pas une grande importance ; il est acceptable de créer systématiquement une base dédiée pour l'hébergement du catalogue RMAN. En revanche, pour un nombre d'instances important, la création de ces bases dédiées supplémentaires, et leur administration finit par générer une charge de travail considérable, aussi lourde que celle des bases à sauvegarder elles-mêmes. Il faut donc organiser de façon rationnelle les catalogues et la ou les instances qui vont les héberger.

L'organisation des catalogues a été étudiée conjointement par un architecte, un expert en bases de données et moi-même. La démarche que nous avons suivie est la suivante :

D'abord, nous nous sommes posé trois questions :

- Combien d'instances doivent être consacré pour y installer les catalogues ?
- Comment faut-il répartir et nommer les catalogues dans les schémas ?
- Comment assurer la continuité en cas de panne de(s) l'instance(s) qui héberge les catalogues ?

Nous allons faire la synthèse des réponses, pour tenter de choisir le modèle d'organisation des catalogues RMAN le plus pertinent.

5.1.2.1 Combien d'instances sont nécessaires pour héberger les catalogues RMAN ?

Pour trouver le nombre adéquat d'instances pour héberger les catalogues, 3 choix sont possibles :

- A chaque base à sauvegarder, correspond une instance unique dédiée à l'hébergement de son catalogue RMAN.
- Quelques instances en nombre limité hébergeant plusieurs catalogues chacune.
- Une seule instance pour tous les catalogues.

Les avantages et inconvénients pour chacun de ces choix ont été recensés et résumés dans le tableau qui suit :

<p>Pour chaque base, une instance unique est dédiée à l'hébergement du catalogue RMAN</p>	<p><u>Avantage</u> : chaque catalogue est indépendant</p> <p><u>Inconvénient</u> : nombreuses créations de bases, nombreuses sauvegardes à ordonnancer.</p>
<p>Quelques instances en nombre limité.</p>	<p><u>Avantage</u> : permet de répartir et croiser les sauvegardes RMAN d'instance hébergeant les catalogues.</p> <p><u>Inconvénient</u> : difficile de retrouver un catalogue. Plusieurs sauvegardes et relativement difficile à ordonnancer.</p>
<p>Une seule instance pour tous les catalogues RMAN.</p>	<p><u>Avantage</u> : une seule base à administrer</p> <p><u>Inconvénients</u> : cette base est très critique</p>

Figure 41: Avantages et inconvénients suivant le nombre de catalogues RMAN

Nous l'avons déjà vu, il serait coûteux de créer une instance dédiée pour chaque base à sauvegarder, puisqu'il faut créer autant de bases hébergeant les catalogues que de bases à sauvegarder avec RMAN. Il faut tenir compte de l'exploitation des instances qui vont héberger nos catalogues. Chacune de ces instances doit être administrée et aussi sauvegardée. Plus le nombre de ces bases est important, plus la charge de travail pour la maintenance est importante. C'est pourquoi, il vaut mieux réduire le nombre de ces instances au minimum.

L'hébergement de catalogues dans quelques instances (deux instances ou plus) permet de sauvegarder ces instances avec RMAN en croisant les sauvegardes. Mais cette option a été exclue car elle est difficile à ordonnancer compte tenu des dépendances entre ces instances. En outre, dans le cas de plusieurs instances, il est difficile de savoir où se trouve un catalogue. Cette solution est également plus complexe à administrer que si tous les catalogues étaient logés dans une seule et même instance.

Créer une seule instance pour l'ensemble des catalogues permet de réduire au maximum la charge de travail et d'administration puisqu'il n'y a qu'une seule base à créer, administrer et sauvegarder.

Pour finir, l'éditeur Oracle, recommande que cette instance unique soit dans la version la plus récente possible.

Conclusion : il est judicieux de loger tous les catalogues dans une seule et même instance, dans une version d'oracle la plus récente possible.

5.1.2.2 Comment répartir et comment nommer les catalogues ?

Nous avons une seule instance pour héberger notre ou nos catalogues. Il faut maintenant savoir comment nous allons les organiser à l'intérieur de notre base unique.

Plusieurs alternatives sont possibles :

- créer un unique catalogue, dans un seul schéma Oracle, pour toutes les bases à sauvegarder,
- un schéma par version de base à sauvegarder (Un pour les bases 9i, un autre pour les bases 10, etc.),
- un schéma dédié pour chaque base.

Les avantages et inconvénients ont été recensés et sont résumés dans le tableau qui suit :

Catalogue unique	<p><u>Inconvénients</u> : le catalogue est très volumineux. Export du catalogue réalisé par DP long. Migration délicate qui impacte tous les catalogues simultanément</p> <p><u>Avantage</u> : un seul catalogue à administrer</p>
Un catalogue par version Oracle	<p><u>Inconvénients</u> : catalogue relativement volumineux, export du catalogue réalisé par DP relativement long. Migration délicate car impactant tous les catalogues d'une version donnée.</p> <p><u>Avantage</u> : un nombre réduit de catalogues à administrer</p>
Un catalogue par base	<p><u>Avantage</u>: chaque catalogue est indépendant, facile à administrer et à maintenir, export du catalogue le plus léger possible.</p> <p><u>Inconvénients</u> : de nombreux catalogues, nécessité de règles de nommage pour les catalogues</p>

Figure 42: Avantages et inconvénients selon la répartition des catalogues RMAN

Pour organiser judicieusement nos catalogues, il est nécessaire de cerner deux contraintes importantes. D'abord, le logiciel de sauvegarde DataProtector effectue systématiquement après chaque sauvegarde RMAN, un export du schéma hébergeant le catalogue, qu'il recopie ensuite sur bande. Pour limiter le temps de sauvegarde il faut donc réduire au minimum la taille de cet export, ce qui implique d'avoir le catalogue le plus petit possible. (cf. documentation technique de DataProtector). Ensuite, il faut penser à l'administration des catalogues. Celle-ci peut être délicate, si un catalogue héberge les données de sauvegardes de plusieurs instances, ou en cas de migration de base par exemple.

Pour remplir les deux exigences évoquées ci-dessus, la meilleure solution à notre problème de répartition en base est de créer un catalogue pour chaque base à sauvegarder. Plus exactement, un schéma dédié est créé, dans lequel on y met notre catalogue. (Un catalogue occupe moins de 30 méga-octets de données en générale). Cette solution permet à la fois : une taille de schéma minimal et donc un temps de sauvegarde du catalogue minimal. Chaque catalogue

étant indépendant les uns des autres, l'exploitation est facilitée et chaque catalogue peut être migré ou manipulé facilement.

Les tests ont montré qu'il était difficile de repérer facilement un catalogue, et de s'assurer qu'un autre catalogue du même nom n'avait pas été créé auparavant. C'est pourquoi, le nom des schémas d'accueil a été normé. Cette norme garantit un nom de catalogue unique et permet une identification rapide. Le nom s'écrit de la façon suivante: *cat_<SID>_<SERVEUR>*, où <SID> représente l'identifiant de notre base à sauvegarder avec RMAN et où <SERVEUR> est le nom du serveur. Dans le cas d'un cluster le nom doit être celui du nom du service. Pour les mêmes raisons, les noms des spécifications de sauvegarde ont aussi été standardisés.

Conclusion: chaque instance à sauvegarder disposera d'un catalogue unique, ayant un nom unique (nécessitant des règles de nommage).

5.1.2.3 Comment améliorer la robustesse des catalogues RMAN ?

Pour remplir les exigences du cahier des charges, la robustesse du catalogue a été étudiée puis revue. En cas de défaillance hébergeant les catalogues, il faut prévoir une solution automatique qui permette la continuité de services.

Deux solutions ont été envisagées :

- Une première architecture utilisant deux instances, a été étudiée, et mise en œuvre sur l'environnement de maquette. Dans cette configuration, un exemplaire de chaque catalogue existe sur chacune de ces instances. Il aurait suffi qu'une seule des instances soit en fonctionnement pour réaliser une sauvegarde. En effet, grâce à un paramétrage simple (définition de l'accès en mode « gestion de panne (*failover*) » et « répartition de charge (*load balancing*) » dans le fichier *tnsnames.ora*.) les sauvegardes ont la possibilité de rechercher automatiquement l'une des instances en fonctionnement. Cette solution à deux instances autorise la panne et l'indisponibilité de l'un des nœuds pendant une maintenance de l'un des nœuds par exemple et permet la sauvegarde

mutuelle des deux catalogues. Il faut cependant synchroniser régulièrement les catalogues, ce qui peut se faire automatiquement dans un ordonnanceur: il suffit d'une tâche planifiée et cyclique lançant la commande spécifique RMAN de resynchronisation (« [resync catalog](#) »). Cependant, chaque modification de d'une des bases (agrandissement d'un fichier par exemple), doit être répercuté sur l'autre base. J'ai suggéré à l'équipe projet de mise en place cette solution à deux instances, mais celle-ci a été jugée trop complexe par l'équipe d'administration des sauvegardes et n'a donc pas été mise en œuvre.

- Une deuxième possibilité, a donc été étudiée conjointement par l'architecte, l'expert en bases de données, et moi-même. Pour offrir le niveau de disponibilité attendu dans le cahier des charges, il a été décidé au cours de réunion de travail, d'utiliser un architecture mainte fois éprouvées c'est-à-dire de créer notre instance sur un cluster à deux nœuds. L'instance hébergeant les catalogues, se situe sur l'un des deux nœuds. En cas de panne, l'instance bascule automatiquement sur l'autre nœud. Chaque nœud du cluster est situé sur un site différent. Pour accroître le niveau de sécurité, les disques utilisés pour la base sont sur des baies redondées. Cette architecture a l'avantage d'avoir été étudiée et mise en œuvre au cours d'autres projets. Elle pouvait donc être mise en place rapidement, et à faible coût. Par ailleurs, le niveau de disponibilité des catalogues offrait exactement le niveau de service exigé puisqu'il est calqué sur celui du serveur hébergeant le logiciel DataProtector.

5.1.2.4 Conclusion sur l'organisation des catalogues

L'organisation des catalogues a été étudiée conjointement par un architecte, un expert en bases de données et moi-même.

L'étude a conduit à héberger tous les catalogues dans une seule et même instance, chaque base à sauvegarder disposant de son propre catalogue RMAN stocké dans un schéma dédié. Afin d'identifier facilement un schéma d'un catalogue donné, le nom du catalogue est unique et a été normalisé.

Pour assurer un niveau de disponibilité exigé par le cahier des charges, l'instance hébergeant l'ensemble des catalogues RMAN est située sur un cluster composé de 2 nœuds. En cas de panne sur le serveur correspondant au nœud actif, la base redémarre automatiquement sur l'autre nœud. Les disques sont redondés sur deux baies situées sur deux sites distincts.

5.1.3 Etude de l'automatisation de la mise en œuvre

Avec un nombre important de sauvegardes de bases à migrer vers RMAN, et le nombre d'acteurs qui interviennent dans la mise en place des sauvegardes, il n'est pas question de réaliser des actions qui n'aurait pas été optimisées ou testées. Il n'est pas question non plus de réaliser le déploiement de RMAN manuellement. Le coût du déploiement pourrait être exorbitant. C'est pourquoi, une automatisation du déploiement a été étudiée et réalisée.

5.1.3.1 Automatisation du déploiement dans DataProtector

Lors de la phase de maquette et du pilote PHENIX, la mise en place des sauvegardes dans le logiciel de sauvegarde DataProtector s'était faite de manière entièrement manuelle dans la console d'administration graphique. Il a donc été envisagé de trouver une solution automatisée et fiable utilisant des scripts.

L'idée première était d'exploiter une documentation de l'éditeur du logiciel de sauvegarde qui explique qu'une création d'une spécification de modèle de sauvegarde ou la création de spécification sauvegarde revient à créer des fichiers en mode texte sur le serveur de sauvegarde appelé « *Cell Manager* ». Ces fichiers contiennent toutes les informations et paramètres nécessaires aux sauvegardes. J'ai proposé l'automatisation de la création des spécifications de sauvegarde. Celle-ci aurait consisté à générer ces fichiers par scripts plutôt que de créer les spécifications à la main via le mode graphique.

Cependant, le temps de mise en place de chaque sauvegarde via l'interface, est évalué à quelques minutes seulement, alors que le développement de scripts fiables aurait au moins demandé plusieurs jours, voir plusieurs semaines. Les avantages de cette interface ont été

jugés trop faibles. C'est pourquoi la mise en place des sauvegardes n'a pas été automatisée et est restée complètement manuelle.

5.1.3.2 Automatisation du déploiement dans VTOM

Lors de la mise en place d'une sauvegarde RMAN, les trois quarts environ de la charge de travail sont consommés pendant l'opération d'ordonnancement. J'ai interviewé l'équipe d'administration des sauvegardes pour savoir dans quelle mesure il était possible d'automatiser le processus d'ordonnancement des sauvegardes.

La création des enchaînements de sauvegarde (sans RMAN) se faisait déjà de manière semi-automatique. Elle se déroule de la manière suivante : le demandeur qui souhaite la mise en place de sauvegardes, remplit une demande dans un fichier Excel ; ce dernier fichier est alors complété puis exploité par l'équipe administration des sauvegardes qui, à l'aide de scripts Visual Basic génère un fichier texte ; ce dernier est importé dans VTOM pour générer un enchaînement de tâches comprenant l'ensemble des tâches VTOM et correspondant au scénario de sauvegarde souhaité ; il ne reste alors plus qu'à le finaliser à la main. Cette technique permet de diviser le temps de mise en place par 3 environ.

L'automatisation de mise en place de l'ordonnancement des sauvegardes RMAN n'a donc pas été difficile à mettre en œuvre et l'équipe d'administration des sauvegardes n'a eu qu'à modifier ce fichier Excel afin qu'il prenne en compte les modifications nécessaires.

J'ai suggéré que l'automatisation soit poussée encore plus loin afin qu'il soit possible de générer en une seule étape un ensemble d'enchaînements de sauvegarde pour plusieurs instances. Cependant, ma proposition n'a pas été suivie, car jugée trop risquée et lourde à mettre en œuvre. Il est vrai que le développement nécessaire aurait été complexe à réaliser.

5.1.3.3 Automatisation du déploiement partie RMAN

Pour créer une base de données, l'équipe d'administration des bases de données utilise un programme standardisé qu'elle a développé. Celui-ci réalise de nombreuses vérifications, et

intègre directement la base créée dans le système d'information existant. (Paramétrage des sauvegardes, des fichiers nécessaires à VTOM, mise en place de profils de sécurité, etc.)

Pour gagner du temps, il a été décidé que, lors de chaque création de base, le catalogue RMAN devait être prêt. J'ai donc adapté ce script et désormais lorsqu'une base est créée, ce script génère automatiquement le catalogue RMAN correctement paramétré. Ce script détecte et gère les diverses versions d'Oracle ainsi que les différents systèmes d'exploitation.

Pour les bases déjà existantes, le script de création de base ne peut être utilisé. J'ai donc conçu et réalisé un deuxième script, ainsi que sa procédure de mise en œuvre associée. Celui-ci est capable de créer automatiquement les catalogues pour un ensemble de bases de données (indiquées en paramètre). Il détecte et gère les diverses configurations de bases et de systèmes d'exploitation et peut créer des dizaines de catalogues en seulement quelques minutes, moyennant un minimum de préparation.

5.1.3.4 Automatisation du changement de mot de passe

Le changement dans les bases de données du mot de passe de SYS pour la base à sauvegardée ou celui du propriétaire du catalogue RMAN doivent être répercutés au niveau du logiciel DataProtector car les spécifications de sauvegardes ont besoin de ces mots de passe pour se connecter et exécuter les sauvegardes. Dans les chapitres précédents ces changements se sont fait jusqu'ici à la main. Cette solution n'est acceptable que sur un nombre de bases limité.

La politique de sécurité, impose des changements de mots de passe régulier sur les bases (1 à 2 par an suivant le type d'environnement) ce qui pourrait constituer à terme une problématique d'envergure pour un périmètre de plusieurs centaines de bases, d'autant plus qu'elle implique deux équipes (administration des bases et administration des sauvegardes). Dès le début du projet j'ai alerté sur cet aspect. J'ai aussi réalisé un ensemble de test, et préparé des scripts visant à modifier en masse ces mots de passe. Cependant, ces aspects de sécurité ont été jugés hors périmètre pour l'instant puisque non prioritaire.

5.1.3.5 Conclusion concernant l'automatisation

Pour optimiser le déploiement de la solution RMAN sur la totalité du parc, les tâches les plus lourdes ont été automatisées.

La création des catalogues a été complètement automatisée grâce à deux scripts (Shell UNIX), l'un pour les nouvelles bases, un autre pour les bases à migrer vers RMAN.

L'automatisation des créations de spécification de sauvegarde n'a pas été jugée suffisamment profitable. Et se réalise donc à la main dans la console d'administration.

Enfin, la création d'enchaînement VTOM pour RMAN a pu être automatisée à 70%. Demander la mise en place d'une sauvegarde revient à remplir un fichier Excel. Ce dernier fichier est complété puis exploité par l'équipe administration des sauvegardes. Ce fichier est doté d'un script Visual Basic qui permet de générer d'un clic de bouton un fichier texte; ce dernier est alors importé dans VTOM pour générer un enchaînement de tâches comprenant l'ensemble des tâches VTOM, correspondant au scénario de sauvegarde souhaité ; il ne reste alors plus qu'à le parfaire à la main

5.2 VALIDATION SUR L'ENSEMBLE DU PERIMETRE

La validation d'une solution, quelle qu'elle soit, est un exercice difficile lorsqu'elle doit être réalisée sur un périmètre large et hétérogène.

A ce stade de l'étude, n'avons pas évalué l'impact de RMAN sur l'ensemble des instances. Nous avons mis en œuvre seulement deux combinaisons de versions Oracle / UNIX : Oracle 10g avec AIX 5.3 pour la maquette, et Oracle 10g avec HP-UX 11iv3 pour le pilote PHENIX. De plus, les sauvegardes mises en place n'avaient pas à cohabiter avec d'autres sauvegardes (sans RMAN).

Or, le périmètre défini par le cahier des charges comprend de nombreuses bases, dans de nombreuses versions et variantes, d'Oracle, de systèmes d'exploitation, et de types de serveurs. Se pose alors la question suivante : comment valider la solution RMAN pour tous les cas possibles? Cette question est d'autant plus cruciale, qu'elle se pose pour les bases concernées par les réglementations financières ou pharmaceutiques imposant une qualification rigoureuse de tout changement.

5.2.1 Démarche pour valider la solution sur l'ensemble du périmètre

Pour valider la solution sur l'ensemble du périmètre, j'ai suivi la démarche suivante :

J'ai commencé par réaliser un recensement des diverses combinaisons de versions de bases et de système d'exploitation les plus utilisées : 6 combinaisons représentent plus de 95% de l'ensemble total. J'ai ensuite mené les tests, et j'ai validé la solution avec 6 combinaisons de versions. Toutes les procédures et scripts ont été revus pour ces versions, lors de tests sur l'environnement de marquette.

Les autres combinaisons de versions n'ont pas été formellement validées car le risque à la mise en œuvre de la solution est considéré comme faible au regard de l'effort à réaliser pour tout valider: toutes les versions sont supportées par l'éditeur de base de données (Oracle) et celui du logiciel DataProtector (HP). Elles seront également très proches de celles qui seront été testées et validées.

Pour finir, j'ai mené une étude, et effectué des tests sur des serveurs mutualisés, pour vérifier que les sauvegardes RMAN pouvaient cohabiter avec l'ancien mode de sauvegarde. Ces tests ont été confortés grâce à la mise en œuvre d'un pilote.

5.2.2 Extension aux autres versions Oracle et OS

Faire des tests sur toutes les versions d'oracle et pour tous les types d'OS (Operating System en anglais, système d'exploitation en français) constitue un travail considérable car plusieurs centaines de combinaisons de versions de bases et de version d'OS sont nécessaires. C'est donc, bien sûr exclu, puisqu'il faut à la fois disposer du matériel nécessaire, dans chacune des versions, et avoir du temps à consacrer aux tests. C'est pourquoi une démarche plus simple et moins coûteuse en temps a été entreprise.

Un recensement des bases Oracle a mis en évidence qu'UNIX est de loin le système d'exploitation le plus utilisé. On trouve majoritairement des versions HP-UX (11iv1 à 11iv3), AIX (4.3, 5.2, 5.3 et 6). HPUX et AIX représente 95 % du parc. Nous avons aussi quelques serveurs SUN et LINUX. Quelques serveurs WINDOWS complètent le parc. J'ai donc choisi de concentrer nos efforts sur ces seuls OS et de ne retenir que les versions HP-UX 11.31 et AIX 5.3.

En ce qui concerne les versions de bases, nous trouvons des bases Oracle en 8.0.x, 8.1.x, 9.2.x, 10.1.x, 10.2.x, et 11.1.x. La version 8 n'étant plus supportée, il a été décidé, en comité de projet, de se concentrer sur les versions 9.2.0.6, 10.2.0.4 et 11.1.0.7, car elles sont plus stables et recommandées par Oracle. Nous ferons l'hypothèse que leur validation entraînera la celle de toutes versions 9,10 et 11. En effet, du point de vue de l'utilitaire RMAN les versions 9 sont proches entre elles, et idem pour les versions 10 et 11.

Avant de démarrer mes tests, j'ai pris la précaution de vérifier l'ensemble des combinaisons de bases et de systèmes étaient validées en consultant les différentes matrices de compatibilité fournies par les éditeurs. Sur le site Internet de support « *Metalink* » d'Oracle, on trouve une matrice de compatibilité qui nous renseigne sur les combinaisons possibles de moteur et pour l'outil RMAN. Sur le site d'Hewlett-Packard, on obtient un document similaire pour les combinaisons DataProtector / RMAN.

Conclusion : au final nous avons délibérément réduits nos tests aux seules versions 9.2.0.6, 10.2.0.4, 11.1.0.7 et aux seuls OS AIX 5.3 et HP-UX 11iv3 ce qui nous fait au total 6 combinaisons de versions à tester. Les autres combinaisons n'ont pas été testées mais en sont très proches. De surcroît, elles sont supportées par les éditeurs HP et Oracle.

5.2.3 Extension du périmètre aux serveurs mutualisés

Avant de déployer la solution RMAN sur un parc de serveurs existant, il faut préalablement s'assurer de ne pas perturber les sauvegardes des serveurs hébergeant déjà des bases.

Tout d'abord, il faut vérifier s'il est possible de faire cohabiter sur un même serveur l'ancien mode de sauvegarde avec les nouvelles sauvegardes RMAN. Les sauvegardes RMAN et non RMAN doivent fonctionner sans se gêner: les anciennes purges des archives sur bande ne doivent interférer avec celles gérées par l'outil RMAN. A l'inverse, RMAN générerait des erreurs et détecterait qu'il lui manque des archives.

Le processus de sauvegarde a dû être disséqué. Des tests ont été réalisés sur l'environnement de maquette. Mais rapidement la conclusion s'est avérée très simple: il ne peut y avoir d'interaction entre les 2 modes de sauvegarde. En effet, l'ancien mode de sauvegarde utilise un fichier de paramétrage qui contient la liste des instances à sauvegarder. Pour migrer une sauvegarde vers RMAN, il suffit d'enlever toute référence à la base dans ce fichier. Les tests en environnement de maquette se sont concluants. (Et sera confirmé au cours du pilote SHIVA à venir.)

Conclusion : plusieurs sauvegardes avec et sans RMAN peuvent parfaitement cohabiter sur un même serveur de base de données Oracle.

5.2.4 Tâches réalisées pour étendre le périmètre

Pour généraliser les procédures sur un périmètre aussi large que possible, plus d'un mois d'effort et de persévérance a été nécessaire.

J'ai préparé une plateforme spécifique, comprenant des bases de tests à sauvegarder. J'ai commencé par l'installation des divers noyaux Oracle (8.1.7.4, 9.2.0.8, et 10.2.0.4 et 11.1.0.6), sur deux machines, l'une en UNIX AIX 5.3 et une autre en version HPUX 11iv3. Quatre noyaux oracle et 4 bases (8, 9i, 10g, 11g) oracle ont donc été installés, sur AIX et sur HPUX. L'installation a été faite en respectant les normes et procédures de l'entreprise ce qui a permis de réduire les risques, et de détecter rapidement les problèmes liés uniquement à l'environnement

Pour chaque version de système d'exploitation, et d'Oracle, deux groupes de tests ont été nécessaires pour valider les procédures et les scripts mis en œuvre :

- des tests qui ne concernent que les nouvelles bases
- d'autres qui ne concernent que les bases déjà sauvegardées sans RMAN

Les tests en version 8 ont été rapidement abandonnés et exclus du projet car cette version n'est plus supportée par l'éditeur Oracle.

Un script standardisé pour la création des bases, a été modifié : désormais à chaque création d'une nouvelle base, RMAN est pris en compte automatiquement. Un catalogue RMAN est créé d'office, dans la bonne instance, avec le bon paramétrage. Le script gère les spécificités liées aux versions de système d'exploitation. (Certaines variables d'environnement peuvent changer, et les noms des répertoires et des bibliothèques RMAN peuvent varier.) D'autre part, les paramètres d'un catalogue RMAN, ou les scripts peuvent changer d'une version d'Oracle à une autre.

Pour créer le catalogue RMAN d'une base déjà existante, un deuxième programme spécifique a été développé. Celui-ci écrit en langage Shell script UNIX a été conçu pour créer en une

seule fois tous les catalogues de toutes les bases d'un serveur donné, ce qui permet d'optimiser le temps de création des catalogues.

Les modèles de sauvegardes RMAN dans DataProtector, ont dû être légèrement retouchés pour prendre en compte l'ensemble des versions Oracle. Par exemple la configuration des canaux de communication RMAN (paramètre « Allocation Channel ») a été revue.

Toutes les procédures ont dû être rejouées et validées une à une pour toutes les versions (Création et paramétrage des catalogues, création des spécifications de sauvegarde, ordonnancement, test de sauvegardes et restauration). Elles ont été améliorées et complétées si nécessaire.

Enfin, une dernière série de tests a été réalisée en mélangeant d'anciennes sauvegardes et les nouvelles (avec RMAN). Ils ont rapidement été concluants et n'ont pas posé de problèmes particuliers.

Conclusion : cette phase importante de tests a été assez longue et difficile, mais elle a permis de valider toute la chaîne de sauvegarde (VTOM, DataProtector et RMAN), pour 6 combinaisons de versions différentes, soit plus de 95% du parc de base de données. Elle a également prouvé que les anciennes sauvegardes (sans RMAN) pouvaient très bien cohabiter avec nouvelles sauvegardes RMAN.

5.3 PREPARATION ET FORMATION

La phase qui précède le déploiement est cruciale. Elle consiste à préparer et former les divers acteurs impliqués, à mettre en place l'infrastructure nécessaire.

5.3.1 Formation et assistance aux changements

L'introduction du composant RMAN induit des changements et une plus grande complexité dans les processus de mise en place des sauvegardes. Il convient donc de piloter l'innovation afin qu'elle soit facilitée et surtout acceptée par les différents acteurs.

Pour accompagner le changement, quatre actions ont été mises en place :

- **L'information des parties:** tous les acteurs de la sauvegardes (une trentaine) ont participé à une séance d'information intitulée « RMAN, ce qui va changer... ». Les avantages de la solution, les changements et les contraintes ont été communiqués.
- **Formation technique du personnel :** un enseignement technique a été dispensé à trois équipes. Chacune de ces équipes (Equipe en charge de l'administration des sauvegardes (ADS), équipe en charge de l'ordonnancement, et les administrateurs de bases de données) ont participé à des séances de travaux pratiques personnalisées en fonction de leur rôle dans le projet. (cf. fiche de cours en annexe).
- **Assistance au démarrage :** un tuteur pour chaque équipe, a été nommé pour suivre pas à pas les mises en place des premières sauvegardes.
- **Support en cas d'incident :** une fois la phase d'assistance achevée, les divers acteurs continuent à bénéficier d'un contact privilégié qui assure le support de haut niveau, éventuellement avec l'aide des divers fournisseurs (HP, Oracle Storagetek ...)

5.3.2 Préparation de l'infrastructure

La préparation de l'infrastructure a consisté à mettre en place le cluster qui héberge les catalogues. Elle a été réalisée juste avant le pilote SHIVA. (cf. plus loin). Un cluster UNIX HP-UX 11i3 comprenant 2 nœuds a été installé par l'équipe d'administration système. L'équipe d'administration des bases de données a ensuite pris le relais en installant un moteur

Oracle en version 11gR2, puis en créant l'instance qui contiendra tous les futurs catalogues RMAN.

Pour valider la nouvelle infrastructure, une série des tests a été menée.

- Quelques catalogues ont été créés.
- Des essais de bascules du clusteur, avec la base ont permis de valider une très faible indisponibilité en cas de panne. (quelques minutes le temps de basculer vers l'autre nœud). Les sauvegardes avec RMAN ont été réalisées sur chaque nœud du clusteur.
- Les procédures de sauvegarde et de restauration, de l'instance hébergeant les catalogues, ont été développées, testées, puis publiées.
- L'ensemble bien sûr a été documenté pour répondre aux exigences documentaires.

5.4 PILOTE DE MIGRATION (SHIVA)

Pour préparer les migrations des bases déjà existantes et sauvegardées avec l'ancienne méthode, un pilote de migration a été réalisé. Ce pilote nommé SHIVA se distingue du précédent (PHENIX) car les bases concernées sont déjà sauvegardées mais sans l'outil RMAN (alors que le pilote PHENIX ne concernait que de nouvelles bases). Il se distingue aussi, car, il a utilisé une centralisation des catalogues RMAN dans une seule base dédiée. (Alors que pour le pilote PHENIX, chaque catalogue était logé dans une base qui leur était réservée).

5.4.1 Objectifs du pilote SHIVA

Ce deuxième pilote consiste à migrer des sauvegardes existantes vers des sauvegardes RMAN sur trois environnements distincts: développement, pré production et production, avec deux bases par environnement. Les bases sont plutôt volumineuses, avec 600 Go de données chacune. Chacune est située sur des serveurs mutualisés hébergeant une vingtaine de bases.

Seules les sauvegardes de bases SHIVA seront migrées, les autres sauvegardes de bases (sans RMAN) resteront inchangées.

Le pilote SHIVA devait donc permettre de :

- valider les procédures de migration des sauvegardes vers RMAN
- valider qu'il est possible de faire cohabiter l'ancien mode et le nouveau mode de sauvegarde RMAN.
- évaluer le gain en temps et en volumétrie des sauvegardes
- Evaluer les problèmes de cohabitation entre ancien et nouveau mode de sauvegarde
- Evaluer les impacts d'une centralisation de tous les catalogues dans une base unique

5.4.2 Tâches réalisées lors du pilote SHIVA

Le pilote SHIVA a commencé par le déploiement des programmes (agents) permettant à VTOM de communiquer avec RMAN. (cf. maquette). Cette partie est simple mais a nécessité l'arrêt de toutes les bases sur le serveur y compris celle non concernée par la migration vers RMAN.

Ensuite la base chargée d'accueillir et centraliser tous les catalogues a été créée. La sauvegarde de celle-ci a été ordonnancée, mais sans utiliser RMAN.

Les catalogues ont été créés et chargés à partir du script déjà utilisé pour la maquette et le pilote PHENIX. Ce script revu pendant la phase d'extension du périmètre, a permis de créer tous les catalogues nécessaires à nos sauvegardes, en une seule fois, pour l'ensemble des bases d'un serveur. Comme prévu, ce script s'est connecté sur la base des catalogues. Il a créé les catalogues en chargeant automatiquement les paramètres RMAN en fonction des versions du système d'exploitation et de la base de données.

Le travail réalisé dans DataProtector a été quasi identique à celui réalisé pour le pilote PHENIX : Les modèles de sauvegarde DataProtector n'ont pas été touchés. Les spécifications de sauvegardes ont été créées en respectant les procédures.

La part la plus importante du travail a été la gestion de la planification sous VTOM. Dans un premier temps l'ancienne sauvegarde n'a pas été touchée, et un nouvel enchaînement a été créé pour les sauvegardes de type RMAN. Une fois la nouvelle sauvegarde RMAN testée et validée, l'ancienne planification a été supprimée dans VTOM.

Les tests de validation ont consisté à restaurer l'ensemble des bases sur un serveur vierge. Tout y a été remonté, le système d'exploitation, la vingtaine de bases et les applications. Cette opération de restauration des systèmes a été faite pour chaque environnement : développement, recette, production.

5.4.3 Retour d'expérience du pilote SHIVA

La procédure de migration des sauvegardes a pu être validée pratiquement sans être retouchée. La cohabitation des 2 modes de sauvegarde (sans RMAN et avec RMAN) a pu être démontrée.

Une synchronisation est nécessaire entre les diverses équipes. L'autorisation d'arrêt des bases n'a pas été longue à obtenir puisqu'elle a demandé l'accord d'un seul acteur.

La première mise en place de sauvegarde comportait des erreurs. Les informations transmises par l'équipe d'administration des bases à destination de l'équipe d'administration des sauvegardes ont été mal interprétées. (Information de connexion aux bases). Ensuite, alors que la sauvegarde ne pouvait avoir lieu qu'à froid (pas de mode archivage), elle a été mise en œuvre à chaud, entraînant des erreurs lors des sauvegardes. Pour finir, le déclenchement de la sauvegarde, a dû être décalé (dans VTOM) suite à des contentions de lecteurs. Les deux équipes ont su collaborer pour résoudre les problèmes qui ont perduré deux jours. (Chaque sauvegarde et chaque restauration dure plusieurs heures à chaque fois.)

Les tests de restauration des bases ont été longs. D'une part, l'équipe chargée de l'administration des sauvegardes, et donc aussi de la gestion des bandes a dû s'y reprendre à plusieurs fois pour récupérer les bandes nécessaires. En effet, les sauvegardes incrémentales augmentent la complexité de la restauration car il faut récupérer les bandes des sauvegardes

de niveau 0, puis les incrémentales, et enfin les bandes pour les fichiers d'archivages. Pour faciliter la récupération des médias, une assistance a été nécessaire auprès de l'équipe chargée de rapatrier les médias.

La charge de travail a pu être évaluée pour chaque équipe. Comme pour le pilote PHENIX, la charge de travail se concentre pour l'essentiel dans la mise en place de l'ordonnancement des sauvegardes. En particulier puisqu'il faut d'abord dans un premier ajout la nouvelle sauvegarde RMAN, et désactivée l'ancienne (sans RMAN). Dans un deuxième temps, et après validation, il faut supprimer l'ancienne sauvegarde. Au vu du temps passé lors de ce pilote, il sera difficile pour une personne de migrer sous VTOM plus de 5 bases à la fois en un seul jour.

Les temps et les volumes de sauvegardes ont été considérablement diminués (d'un facteur 6) grâce aux sauvegardes incrémentales faites en semaine. Le week-end, on note une diminution de 35% avec des sauvegardes complètes. Le pilote démontre donc l'intérêt de la solution.

5.5 DEPLOIEMENT

5.5.1 Périmètre

Nous l'avons déjà vu, l'essentiel du parc, soit plus de 90% de bases de données est composé des versions 9, 10, 11 sur des plateformes HPUX et AIX. Au final c'est plus de 500 bases qui sont éligibles immédiatement. Les bases ne correspondant pas à ces critères ne sont pas concernées par le déploiement actuel, mais feront l'objet d'un lot ultérieur.

5.5.2 Lotissement

Avant de généraliser la solution il est important de planifier de manière rigoureuse le déploiement pour éviter les tâches inutiles. Pour gagner du temps lors du déploiement, un

lotissement a été réalisé. Les bases du périmètre ont été regroupées selon un ensemble de règles, définies lors de réunions.

- Pour des raisons de sécurité et pour respecter les procédures, il vaut mieux travailler projet par projet, en migrant dans l'ordre, un à un, les environnements de développement, recette, qualification, puis ceux de production.
- Afin de bénéficier rapidement des avantages offerts par RMAN, les bases volumineuses ont été jugées prioritaires.
- Pour éviter les va-et-vient sur les serveurs, il est préférable de travailler, autant que possible, serveur par serveur. En conséquence, dès l'outil RMAN est mis en œuvre sur un serveur, on cherchera autant que possible à basculer vers RMAN toutes les sauvegardes de bases qui s'y trouvent.

En application de ces règles, sept lots ont été définis.

5.5.3 Déploiement et retour d'expérience

Le déploiement lui-même ne pose pas de problème particulier. Pour l'instant, nous n'avons pas détecté de problèmes liés aux versions de bases ou de systèmes. Les procédures se sont révélées fiables. Chaque équipe n'a eu qu'à les appliquer.

J'ai assuré l'essentiel du support aux diverses équipes sur des points mineurs et j'ai même été sollicité pour des problèmes de sauvegardes qui n'avaient au final rien à voir avec le déploiement de l'outil RMAN.

Les nouvelles bases créées sont désormais systématiquement sauvegardées avec RMAN, sans problème, à un rythme régulier et normal. (Une à deux créations par semaine environ). Une dizaine de sauvegardes RMAN ont ainsi été mises en place depuis la deuxième moitié du mois d'octobre.

En revanche, les anciennes sauvegardes ne sont migrées vers RMAN qu'à un rythme très lent, si bien que seulement une vingtaine de sauvegardes ont été portées vers RMAN depuis le mois d'octobre. De nombreux retards sont à signaler. Le déploiement a commencé en septembre 2010, alors que le département de production informatique est en cours d'externalisation de ses activités. De nombreuses ressources ont été réaffectées pour réaliser les opérations de transitions, alors qu'elles auraient été nécessaires au bon déroulement du projet. Nous constatons aussi qu'il est difficile d'obtenir un feu vert lorsqu'il est nécessaire de stopper des bases, pour interfacer DataProtector avec RMAN. (cf. chapitre 4). Il en résulte des retards sur le planning initial.

5.5.4 Retour d'expérience de l'utilisation de RMAN

RMAN a amélioré notablement les performances : les sauvegardes sont plus rapides dès que la volumétrie n'est pas trop faible. Les sauvegardes sont aussi moins volumineuses et consomment donc moins de bandes. Lors de sauvegarde différentielle, les gains atteignent fréquemment les 80%. Les sauvegardes complètes sont aussi plus rapides et moins volumineuses mais dépendent énormément du taux remplissage de la base. Une base est sauvegardée d'autant plus rapidement qu'elle est peu remplie.

5.6 RESUME DU CHAPITRE 5

Ce chapitre a montré comment la solution RMAN a pu être généralisée sur un périmètre aussi large que possible, puis comment elle a été déployée.

L'organisation des catalogues a d'abord été revue pour faciliter la maintenance. Ils sont désormais regroupés dans une seule et même base de données. Chaque catalogue dispose de son propre schéma ce qui les rend indépendant les uns des autres et facilite leur administration (suppression ou migration de catalogue par exemple). Pour répondre aux exigences du cahier des charges, une deuxième base de qualification a été nécessaire. Un cluster composé de deux nœuds héberge nos catalogues pour accroître leur disponibilité.

Plus d'un mois de tests et d'efforts ont permis valider la solution sur les bases en version 9,10 et 11, pour les systèmes d'exploitation AIX et HPUX ce qui représente plus de 95% du parc total. Si les autres versions n'ont pas été formellement validées, elles sont cependant supportées par l'éditeur de base de données (Oracle) ainsi que celui du logiciel DataProtector (HP).

Pour accélérer le déploiement, l'industrialisation des diverses tâches a été recherchée. La création des catalogues a été complètement industrialisée grâce à deux scripts (Shell UNIX), l'un pour les nouvelles bases, un autre pour les bases à migrer vers RMAN. En revanche, l'automatisation des créations de spécification de sauvegarde n'a été développée, car jugée peu profitable. Enfin, la création d'enchaînement VTOM pour RMAN a pu être automatisée à 70%.

Le cluster hébergeant l'instance pour les catalogues a été mis en place puis testé selon les normes de qualité et la réglementation pharmaceutique (GxP) et financière (SoX). Le déploiement a été précédé d'un accompagnement au changement, sous forme notamment de séances d'information, et de formations techniques.

Finalement, le déploiement a démarré au courant du mois d'octobre, mais subit de nombreux retards. A ce jour une quarantaine de bases sont sauvegardées à l'aide de notre nouvelle solution RMAN.

Chapitre 6 : BILAN, RESULTATS & PERSPECTIVES

Le projet RMAN a commencé en septembre 2009 et est toujours en cours fin février 2011. Ce chapitre tire un bilan.

6.1 RESULTATS DU PROJET

Au moment où je finis ce rapport (février 2011), une quarantaine de bases sont sauvegardées avec RMAN. Cependant le passage vers RMAN continue, mais semble heureusement s'accélérer.

L'introduction du composant RMAN dans la chaîne existante a pu être réalisée sans perturbation et avec un coût raisonnable. Désormais, la sauvegarde RMAN fait partie du catalogue de service offert par la production informatique.

RMAN a amélioré notablement les performances : les sauvegardes sont plus rapides (pour peu que la volumétrie ne soit pas trop faible). Les sauvegardes sont moins volumineuses et consomment donc moins de bandes. Pour les bases assez importantes le gain est fréquemment de 50% lorsque la base est sauvegardée intégralement (dépend du remplissage de la base), et lors de sauvegarde différentielle, les gains atteignent fréquemment les 80%.

RMAN a également résolu certains problèmes d'indisponibilité de bases, ainsi que des problèmes de contention de lecteur, spécialement pour les bases volumineuses.

On peut déjà considérer que le projet est un succès même s'il s'étale sur une période de plusieurs mois et alors qu'il reste encore un grand nombre de bases à passer sous RMAN.

6.2 BILAN PERSONNEL

6.2.1 Ma contribution au projet

J'ai activement participé à la coordination du projet, en faisant constamment le lien entre les experts et architectes et les équipes de production, ce qui a permis d'accélérer les mises en œuvre et surtout d'économiser de l'énergie et un temps précieux.

J'ai été un contributeur majeur lors de la rédaction de l'expression des besoins. J'ai assisté le chef de projet, et aidé les équipes chargées de l'étude de la solution à passer de manière automatisé et fiable du stade de maquette à celui de la production à une large échelle.

J'ai été le principal acteur de l'industrialisation des processus (sinon le seul). Mon travail a permis de migrer massivement, en une seule opération les catalogues RMAN en un temps réduit au strict minimum. J'ai aussi travaillé sur l'optimisation de l'organisation des catalogues. J'ai réalisé seul tous les tests de version Oracle et de systèmes d'exploitation ce qui a permis d'étendre de manière contrôlée le périmètre de la solution. A vrai dire personne ne souhaitait prendre à sa charge ce travail laborieux.

J'ai pu mettre à profit mes compétences de formateur, compétences que j'ai acquises auparavant lorsque j'étais sous les drapeaux. J'ai organisé des séances d'information et dispensé un enseignement technique, et pratique, au cours d'ateliers techniques. J'ai pris à ma charge l'information des diverses équipes lors de séances intitulées « RMAN, ce qui va changer », incluant la rédaction d'un support (PowerPoint), l'organisation et l'information elle-même. C'est au cours de ces sessions d'information que j'ai pu « vendre » la solution RMAN en mettant en avant les avantages. J'ai aussi animé des ateliers pratiques destinés aux diverses équipes techniques. En particulier lieu pour l'équipe d'administration des bases de données.

6.2.2 Bénéfices personnels tirés du projet

Tout au long du projet, j'ai bénéficié d'une large autonomie J'ai pu prendre de nombreuses initiatives, comme par exemple celle d'étudier une architecture plus robuste pour les catalogues.

Le projet RMAN m'a permis d'approfondir ma connaissance sur diverses technologies. J'ai eu la chance d'avoir à ma disposition un environnement de maquette comprenant deux serveurs de test UNIX (AIX et HPUX) et un environnement de maquette DataProtector connecté à une robotique de sauvegarde. J'ai pu affiner ma compréhension du logiciel DataProtector et comprendre comment l'interfacer avec l'Ordonnanceur VTOM. J'ai grandement perfectionné mon expertise sur les techniques de sauvegarde et restauration des bases Oracle avec l'outil RMAN.

Au-delà de l'aspect technique, j'ai surtout énormément appris sur la gestion d'un projet et particulièrement sur l'aspect humain du management. J'ai du apprendre à gérer mon impatience lorsque le projet piétinait. Cette impatience était d'autant plus forte que le projet RMAN était le sujet de mon mémoire d'ingénieur. J'ai appris à composer avec les divers intérêts en jeu, pour aider de mon mieux le projet.

6.2.3 Limites du projet

De nombreux points de blocages ont retardé le projet de manière considérable. J'ai par exemple dû attendre près de 3 mois avant que l'on mette à disposition les machines destinées à héberger les catalogues.

Certains freins ou blocages étaient la conséquence de mésententes entre certains acteurs du projet. L'un d'entre eux a même cessé définitivement toute collaboration, en claquant la porte en pleine réunion. La profonde réorganisation, conséquence de l'externalisation de la production informatique, a aussi très largement pesée.

Enfin, une certaine résistance aux changements a contraint le chef de projet et moi-même à mieux communiquer et à vendre « la solution » auprès des divers acteurs lors des réunions de

projet. Avec du recul je pense que dès le début du projet, il aurait fallu consacrer du temps à communiquer et à faire la promotion de la solution RMAN. On constate cependant, au fil du temps, une meilleure adhésion à la solution de la part des divers acteurs. Le travail de communication pour mettre en valeur l'outil RMAN, a donc fini par payer.

Je n'ai pu traiter certains sujets de manière aussi approfondie que je l'aurais souhaité.

Par exemple, le changement des mots de passes nécessaire aux sauvegardes aurait pu (du ?) être largement automatisé. En effet, pour respecter la politique de sécurité, les mots de passes doivent être régulièrement modifiés. On imagine la difficulté de changer des mots de passes en masse (plusieurs centaines de bases) s'il faut coordonner plusieurs équipes, ce qui ne manquera pas de constituer à terme une problématique d'envergure.

J'aurais souhaité que l'équipe projet s'implique d'avantage dans la partie automatisation des procédures. J'ai passé beaucoup de temps à mettre au point des procédures et des scripts pour créer en masse et paramétrer les catalogues de manière automatique. J'ai aussi pris l'initiative de réaliser en maquette une série de tests pour les composants DataProtector et VTOM. Mais certains résultats n'ont pas été exploités pour l'instant. A mon avis, la partie automatisation de VTOM et DataProtector, aurait pu être approfondie. Cependant, plusieurs de mes développements ont tout de même permis de fiabiliser et d'accélérer la mise en œuvre.

6.3 PERSPECTIVES

Au moment où s'achève la rédaction du mémoire, environ une quarantaine de bases sont sauvegardées avec l'outil RMAN. Le projet a tout pour devenir un succès, mais il reste environ 500 bases à migrer dans le périmètre défini par le cahier des charges.

Rappelons que les bases en version 8 d'Oracle ont été exclues de l'étude et du projet. En effet, elles sont peu nombreuses, et introduisent inutilement de la complexité. En outre, ces bases sont destinées à disparaître puisqu'elles ne sont plus supportées par l'éditeur Oracle.

Plusieurs versions du système d'exploitation Windows d'une part et UNIX d'autre part (Sun et Linux) n'ont pas encore été étudiés car il n'en existe qu'une poignée (Environ une vingtaine). Le projet n'a pas pu prendre en compte les bases sauvegardées avec un ordonnanceur autre que VTOM ou un logiciel de sauvegarde différent de DataProtector. Un dernier lot est déjà prévu pour migrer toutes ces sauvegardes vers RMAN.

Alors que les bases utilisées par le progiciel SAP avaient été exclues du projet, une étude de sauvegarde des bases SAP avec RMAN a été lancée. Nul doute, qu'elle devrait profiter pleinement de toute l'expérience acquise durant notre projet.

BIBLIOGRAPHIE

Références bibliographiques :

(Par ordre alphabétique des noms d'auteurs)

[1] **Jean-François CARPENTIER.** *La sécurité informatique dans la petite entreprise : état de l'art et bonnes pratiques.* ENI éditions, 276 pages avril 2009 –ISBN 978-2-7460-4820-1

[2] **Preston DE GUISE.** *Entreprise systems : backup and recovery, a corporate insurance policy .*CRC Press, Auerbach publications, Taylor and Francis group – 306 pages 2009 – ISBN 978-1-4200-7639-4

Manuels des éditeurs de logiciels consultés :

[3] **Hewlett -Packard development Company L.P.** *HP OpenView Storage Data Protector: Integration guide for Oracle and SAP release A.06.00.* Manual, édition July 2006 – 396 pages - Manufacturing Part Number B6960-96008

[4] **Hewlett -Packard development Company L.P.** *HP Data Protector A.06.10, Troubleshooting guide.* Manual first edition November 2008 – 168 pages - Manufacturing Part Number B6960-96037

[5] **Hewlett -Packard development Company L.P.** *HP Data Protector A.05.50, Administrator's Guide.* Manual édition October 2005 – 946 pages - Manufacturing Part Number B6960-90106

[6] **Hewlett -Packard development Company L.P.** *HP Data Protector A.06.00, Concept Guide.* Manual édition July 2006 – 460 pages - Manufacturing Part Number B6960-96001

[7] **Oracle Company.** *Oracle 10g sauvegarde et restauration avancées, manuel du stagiaire - en 3 volumes -* Edition 1.0 Novembre 2006 (Part Number : D22057FR10 - D46891)

[8] **Oracle Company.** *Best Practices for Oracle Databases - Oracle White Paper -16 pages -* September 2005 (Part number: D22057FR10 - D46891)

[9] **Oracle Company.** *Oracle database, Backup and Recovery Basics - 220 pages -* Novembre 2005 (Part number: B14192-03)

[10] **Oracle Company.** *Oracle database, Backup and Recovery Reference 11g - 452 pages* – August 2008 (Part number B28273-03)

[11] **Oracle Company.** *Oracle database, Backup and Recovery User's Guide - 456 pages* – Août 2008 (Part number: B14191-03)

Notes techniques consultées (de l'éditeur de logiciel Oracle) :

Notes publiées sur le site internet oracle accessible à l'adresse: metalink.oracle.com

Ce site nécessite un compte fourni par l'éditeur Oracle.

[12] *Recovery Manager - Frequently Asked Questions* [doc. n° 76736.1]

[13] *RMAN: Backup Shell Script Example* [doc. n° 137181.1]

[14] *How to Query the RMAN Recovery Catalog* [doc. n°98342.1]

[15] *How to Backup Different Release Databases with the Same Catalog Database*
[doc. n°123210.1]

[16] *Recovery catalog for RMAN backup* [doc. n° 452529.1]

[17] *RMAN: Quick debugging guide* [doc. n° 132941.1]

Références Internet:

[18] **HEWLETT PACKARD**

Linking Oracle with Omniback / Data Protector on UNIX systems

http://support.openview.hp.com/dataprotector_docs.jsp

(Dernière consultation le 15 janvier 2011)

[19] **GREGORY BROISSARD**

Présentation de l'outil RMAN D'Oracle

<http://oracle.developpez.com/guide/sauvegarde/rman/>

(Dernière consultation le 15 janvier 2011)

[20] **AUTEUR INCONNU**

Oracle database Backup and Recovery Backup FAQ

http://www.orafaq.com/wiki/Oracle_database_Backup_and_Recovery_FAQ

(Dernière consultation le 15 janvier 2011)

[21] **SONIA HAMILTON**

Oracle how to purge old RMAN Backups

<http://soniahilton.wordpress.com/2008/12/18/oracle-how-to-purge-old-rman-backups/>

(Dernière consultation le 15 janvier 2011)

LISTE DES FIGURES

Figure 1: Lecteur DAT et cartouches DDS-4.....	17
Figure 2: Cartouche DLT	17
Figure 3: Bras du robot prenant une bande de sauvegarde.....	18
Figure 4: Exemple de robot de sauvegarde (Storagetek)	19
Figure 5: Exemple de petit robot de sauvegarde (HP Surestore)	19
Figure 6: Répartition du marché des logiciels de sauvegarde (2005, source GARTNER).....	22
Figure 7: Schématisation d'une infrastructure de sauvegarde	24
Figure 8: Architecture simplifiée d'une base Oracle.....	33
Figure 9: Exemple de lancement de RMAN à l'aide de l'invite de commande UNIX.....	38
Figure 10: Exemple d'utilisation de sauvegardes incrémentales de type différentiel.....	40
Figure 11: Exemple d'utilisation de sauvegardes incrémentales de type cumulatif	41
Figure 12: Organisation fonctionnelle de Sanofi-Aventis.....	45
Figure 13: L'équipe DBA chez Sanofi-Aventis	46
Figure 14: Extérieur robot de sauvegarde 9310	47
Figure 15: Robot prenant une bande de sauvegarde.....	47
Figure 16: Politique de rétention	49
Figure 17: Phases du projet, en deux parties.....	54
Figure 18: Description simplifiée de la chaîne de sauvegarde (sans RMAN)	62
Figure 19: Exemple d'enchaînement dans une fenêtre graphique de VTOM.....	64
Figure 20: Ordonnancement de 2 tâches A et B avec B dépend de A	64
Figure 21: Ordonnancement type pour les sauvegardes à froid	65
Figure 22: Ordonnancement type pour les sauvegardes à froid des fichiers de base.....	66
Figure 23: Ordonnancement type pour les sauvegardes des fichiers d'archives.....	68
Figure 24: Script de lancement des sauvegardes « shob2sauv ».....	70
Figure 25: Interface graphique du logiciel de sauvegarde DataProtector	73
Figure 26: Architecture simplifiée DataProtector	74
Figure 27: Création d'un modèle de sauvegarde.....	75
Figure 28: Composants principaux de l'outil RMAN.....	80
Figure 29: Description simplifiée de la chaîne de sauvegarde cible (avec RMAN)	82
Figure 30: Déploiement de l'agent RMAN pour DP avec la console d'administration.....	85
Figure 31: Illustration du déploiement de l'agent RMAN pour DataProtector (suite)	85
Figure 32: Création d'un modèle de sauvegarde.....	87
Figure 33: Pseudocode RMAN du modèle	88
Figure 34: Création d'une spécification de sauvegarde avec la console d'administration	89
Figure 35: Informations de connexion nécessaires à la spécification de sauvegarde	89
Figure 36: Lignes de code à ajouter dans le script UNIX qui lance les sauvegardes DP.....	91
Figure 37: Scénario de sauvegarde à chaud RMAN	93
Figure 38: Scénario de sauvegarde des archives avec RMAN.....	94
Figure 39: Scénario de sauvegarde à froid RMAN	95
Figure 40: Démarche d'optimisation de la solution RMAN	105
Figure 41: Avantages et inconvénients suivant le nombre de catalogues RMAN.....	107
Figure 42: Avantages et inconvénients selon la répartition des catalogues RMAN	109