



Un chiffre d'affaire qui frise régulièrement l'insolence malgré la crise, des outils couvrant l'ensemble des problématiques d'intégration de données (couvrant les problématiques d'un ETL, de l'EAI, des contrôles des données de référence, de la qualité de données ou encore du suivi d'une donnée depuis la base de données jusqu'à sa restitution, etc.), un panel d'outils qui se veulent très intuitifs, un positionnement de leader dans l'intégration et le référentiel d'entreprise, les modules d'Informatica ne cessent de faire parler d'eux.

Malgré une concurrence féroce avec un grand nombre d'outils d'intégration et un tarif relativement élevé, d'après les différents baromètres et études, son prix n'est pas un frein à son utilisation. Bien au contraire, le GARTNER le place en tête des outils d'intégration de données, et de nombreux clients le considèrent comme leur outil préféré.

A travers vos différentes questions, on s'aperçoit que vous êtes nombreux à vous intéresser à cet outil, que vous développez sous des ETL open sources ou concurrents.

Nous sommes un groupe d'expert Informatica de 5 personnes [dont une personne est certifiée et 2 sont en cours de certification] ayant pour objectif de faire connaître Informatica, notamment à travers notre expérience terrain. Pour cela, nous vous mettons à disposition des fiches. Celles-ci n'ont pas la prétention d'être la référence ou de vouloir donner des directives de développement aux consultants décisionnels. Cependant, nous espérons qu'il s'agira d'un bon retour d'expérience du terrain, et qu'il sera bénéfique à chacun d'entre vous.

Un espace d'échange est en cours de construction, mais vous pouvez d'ores et déjà nous joindre à l'adresse suivante : expert_informatica@unovia.fr

Tout le travail accompli par ce groupe est fait de façon bénévole, n'hésitez donc pas à les remercier et à les encourager.

Stéphane THIA





SOMMAIRE

<i>ORGANISATION POWERCENTER</i>	3
A. Présentation générale	3
I. Principes de fonctionnement.....	3
Le moteur de traitement.....	3
Les outils clients serveur de Powercenter.....	3
II. Schéma d'architecture logicielle	4
III. Schéma d'architecture technique	4
B. Le référentiel Informatica	5
C. La gestion des utilisateurs	5
I. Groupes et utilisateurs.....	5
II. Détails des privilèges et permissions définis dans le référentiel.....	6
<i>ORGANISATION LOGIQUE SUR LE SERVEUR INFORMATICA</i>	7
<i>REGLES POWERCENTER</i>	8
A. Gestion du référentiel de sécurité Power Center	8
B. Organisation du développement.....	8
<i>GLOSSAIRE</i>	9



ORGANISATION POWERCENTER

A. PRESENTATION GENERALE

PowerCenter est un ETL (Extraction, Transformation des données et Loading), multi sources (SGBDR, fichiers plats ASCII ou EBCDIC) et multi cibles (SGBDR, fichiers plats). Il assure l'alimentation de l'ensemble des bases de données décisionnelles (ODS, Datawarehouse, Datamarts) ou non (systèmes opérationnels).

I. PRINCIPES DE FONCTIONNEMENT

C'est un outil de développement sécurisé qui offre toutes les fonctionnalités nécessaires aux transformations, contrôles de qualité et d'intégrité et de répartition des informations dans les bases de données. A l'aide des outils clients serveurs de Powercenter, le développeur définit les données sources et cibles, les transformations et confie le traitement au moteur Powercenter.

LE MOTEUR DE TRAITEMENT

Le moteur de traitement est généralement situé sur un serveur Unix. Il permet d'extraire, de transformer et de charger des données. Les métadonnées sont stockées dans un référentiel communément appelé repository. Ce dernier est hébergé par une instance de base de données. On y trouvera l'ensemble des informations nécessaires au bon déroulement des traitements :

- ★ Les informations relatives à la sécurité applicative (informations liées à la gestion des utilisateurs et des profils : noms, mots de passe, accès aux données, aux ressources, ...)
- ★ Les informations relatives aux objets de développement (méta-données, objets, transformations,)

Un moteur Powercenter est connecté à un serveur de référentiel ; il accède alors à la base contenant le référentiel. Il accède aux bases cibles ou sources via IPC (sur les bases se trouvant sur le même serveur) ou SQL*Net (sur les bases distantes).

LES OUTILS CLIENTS SERVEUR DE POWERCENTER

La suite d'outils clients/serveurs de Powercenter, installée sur les postes développeurs, permet d'administrer et de définir les traitements en mettant à jour les informations du référentiel PowerCenter, situé sur le serveur. Elle communique avec l'instance de base de données contenant le référentiel via le serveur de référentiel (TCP/IP).

Elle se compose d'un ensemble d'outils décrit au chapitre 4 « Composants de l'offre Powercenter d'Informatica »



Le Designer : Application permettant le développement des mappings afin d'extraire, transformer et charger des données.



Le Workflow Manager : Application permettant l'ordonnancement et la configuration des mappings.



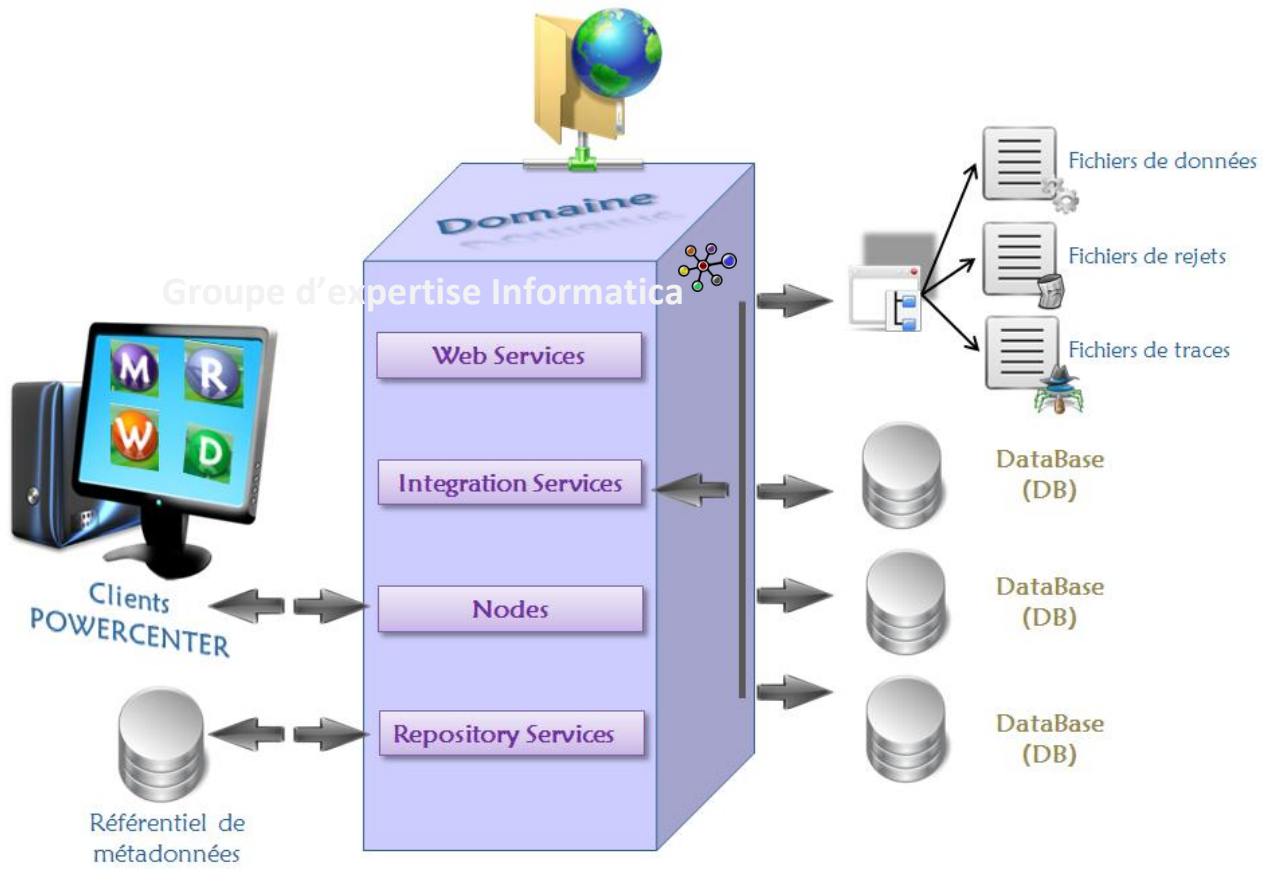
Le Workflow Monitor : Application permettant le suivi de l'exécution des workflows.



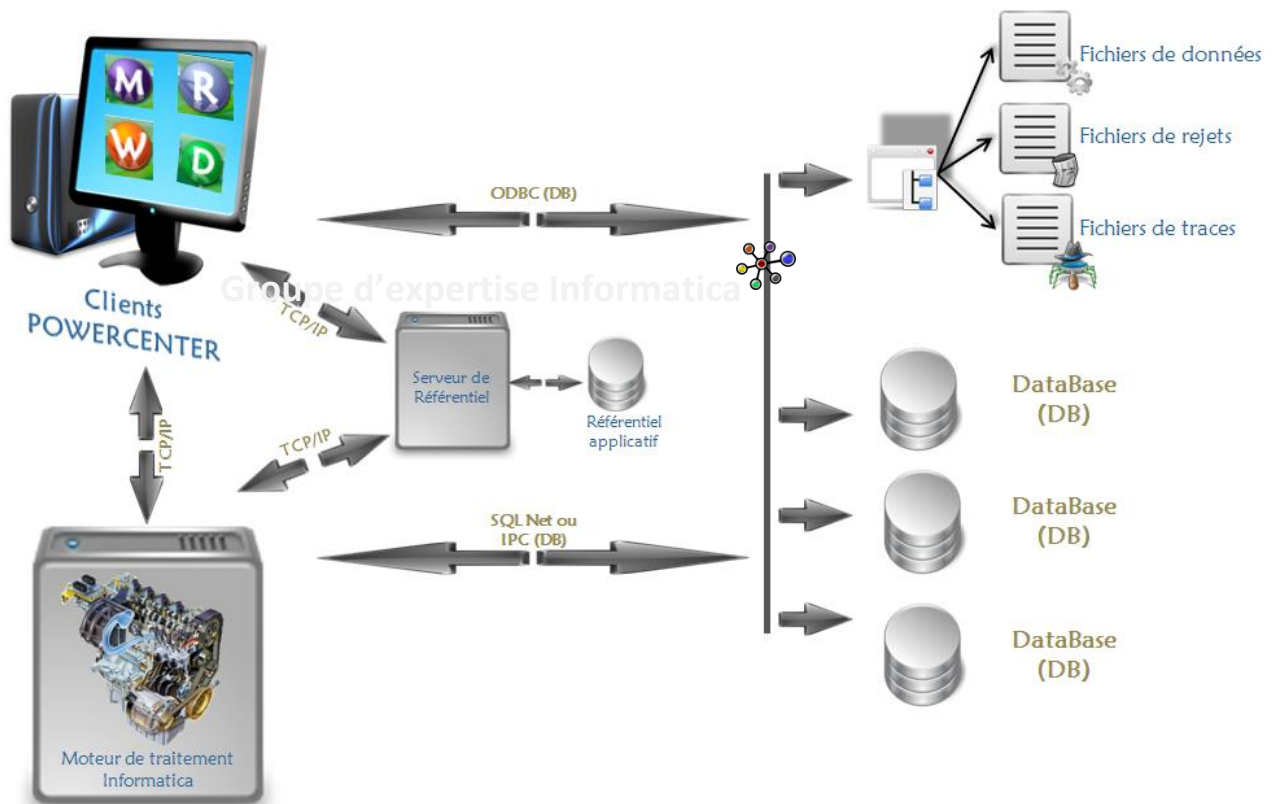
Le Repository Manager : Application permettant la gestion des référentiels, des répertoires propres aux référentiels, des utilisateurs, des droits...



II. SCHEMA D'ARCHITECTURE LOGICIELLE



III. SCHEMA D'ARCHITECTURE TECHNIQUE





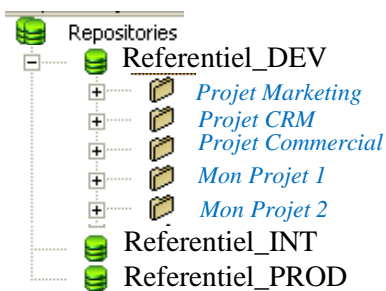
B. LE REFERENTIEL INFORMATICA

Le Référentiel est un dictionnaire de métadonnées sous PowerCenter dans lequel est défini et stocké l'ensemble des opérations de transformation nécessaires à une alimentation.

Un référentiel est composé de répertoires, nommés « folders », servant à séparer les alimentations propres à chaque application, ou destinés à un ensemble d'applications dans le cas de folders partagés. Les consultations et les modifications de ce référentiel sont opérées via un serveur de référentiel PowerCenter.

Les métadonnées d'un référentiel sont stockées dans une base de données hébergée par une instance SGBDR.

Un moteur PowerCenter est connecté à un et un seul référentiel. Ce dernier est alors considéré comme actif. Les autres référentiels sont appelés dictionnaires « dormants » (dictionnaire de sauvegarde, de migration...).



C. LA GESTION DES UTILISATEURS

I. GROUPES ET UTILISATEURS

Tout utilisateur défini dans le référentiel est identifié par un login et un mot de passe. L'utilisateur peut changer son mot de passe via le module « Repository Manager ».

Pour toute application, nous préconisons de créer un groupe affecté à chaque composante de l'architecture : ODS, Datawarehouse, Datamarts. Ces groupes sont alors dotés de privilèges, droits globaux sur l'ensemble du référentiel.

A leur création, les utilisateurs héritent des privilèges accordés au(x) groupe(s) auquel ils appartiennent. Chaque utilisateur appartient à un groupe.

Il existe au moins un folder par composante de l'architecture (ODS, Datawarehouse, Datamarts). Ainsi, un groupe est affecté à chaque folder. L'un des membres du groupe sera désigné « propriétaire » du folder : c'est en général un utilisateur générique créé par l'administrateur. Ce login est constant et rend flexible l'éventuel changement de personnel dans l'équipe projet.

Il est possible de définir des droits, ou « permissions » au niveau du folder (R, W, X) pour trois entités distinctes : le propriétaire du folder, les membres de son groupe, et les autres utilisateurs définis dans le référentiel.



II. DETAILS DES PRIVILEGES ET PERMISSIONS DEFINIS DANS LE REFERENTIEL

Les privilèges accordés aux développeurs sont les suivants :

SUJET	PRIVILEGE	DESCRIPTION
Folder	Copy	Copier des folders à l'intérieur d'un repository ou vers un autre repository.
	Manage Version	Dans un repository versionné, changer le statut des folders.
Runtime Objects	Create, Edit, and Delete	Créer, éditer et supprimer des sessions, des objets de configuration, des workflows, des tasks et des worklets.
	Manage Version	Dans un repository versionné, changer le statut, recover et purger les versions des runtime objects. Annuler les checkouts faits par les autres utilisateurs.
	Monitor	Suivre les workflows et les tasks dans le workflow monitor
	Execute	Star, cold start, recover sur tasks et workflows.
Global Objects	Create Connections	Créer des connexions.
	Manage Deployment Groups	Dans un repository versionné, créer des deployment groups.
	Create Labels	Dans un repository versionné, créer des labels.
	Create Queries	Créer des objets Queries
Tools	Access Designer	Se connecter au repository en utilisant le Designer.
	Access Repository Manager	Se connecter au repository en utilisant le Repository Manager. Lancer des commandes pmrep.
	Access Workflow Manager	Se connecter au repository en utilisant le Workflow Manager.
	Access Workflow Monitor	Se connecter au repository et aux Integration Services en utilisant le Workflow Monitor.
Design Objects	Create, Edit, and Delete	Créer, éditer, et supprimer des business components, paramètres et variables de mapping, mappings, mapplets, transformations et user-defined functions.
	Manage Versions	Dans un repository versionné, changer le statut, recover et purger les versions des design objects. Annuler les checkouts faits par les autres utilisateurs.
Sources and Targets	Create, Edit, and Delete	Créer, éditer, et supprimer des cubes, dimensions, définitions de sources, et définitions de targets.
	Manage Versions	Dans un repository versionné, changer le statut, recover, et purger les versions des objets sources et cibles. Annuler les checkouts faits par les autres utilisateurs.



ORGANISATION LOGIQUE SUR LE SERVEUR INFORMATICA

Sous Unix, l'utilisation de PowerCenter permet l'utilisation de variables, notamment pour :

- ✓ la définition des sources : `$PMSourceFileDir`,
- ✓ des logs : `$PMSessionLogDir` et `$PMWorkflowsLogDir`,
- ✓ des fichiers de Cache : `$PMCacheDir`,
- ✓ des fichiers temporaires : `$PMTempDir`,
- ✓ des fichiers de rejets : `$PMBadFileDir`,
- ✓ et des fichiers d'état des traitements : `$PMStorageDir`.

Ces variables évitent aux développeurs de définir le chemin des fichiers utilisés lors des alimentations de façon statique.

★ **Répertoire de stockage des fichiers de compte rendu de sessions : `$PMSessionLogDir`**

Ce répertoire est destiné au stockage des fichiers de log générés par les sessions. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers de compte rendu de workflow : `$PMWorkflowLogDir`**

Ce répertoire est destiné au stockage des fichiers de log générés par les workflows. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des scripts shell : `$PMExtProcDir`**

Ce répertoire est destiné au stockage des scripts UNIX utilisés dans Powercenter (pré et post session, commandes). C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers de Rejet : `$PMBadFileDir`**

Ce répertoire est destiné au stockage des fichiers de rejets (.bad) générés par Powercenter au moment de la procédure d'alimentation. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des caches : `$PMCacheDir`**

Ce répertoire est destiné au stockage des fichiers de cache générés par Powercenter au moment de la procédure d'alimentation. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers temporaires : `$PMTempDir`**

Ce répertoire est destiné au stockage des fichiers temporaires générés par les tris Powercenter (objet Sort) au moment de la procédure d'alimentation. Ces fichiers sont détruits à la fin de chaque alimentation. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers plats cibles : `$PMTargetFileDir`**

Ce répertoire est destiné au stockage des fichiers plats cibles générés par Powercenter lors d'une alimentation. Ces fichiers sont utilisés pour des procédures de contrôle, de déclenchement en réponse à des besoins spécifiques ou pour des échanges de données entre environnements par exemple. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers de paramètres : `$PMRootDir`**

Ce répertoire est destiné au stockage des fichiers de paramètres utilisés dans les sessions et les workflows Powercenter. Les fichiers de paramètre sont optionnels. C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers sources : `$PMSourceFileDir`**

Ce répertoire est destiné au stockage des fichiers sources utilisés dans les sessions et les workflows Powercenter (en dehors des fichiers liste et des fichiers source issus de la procédure du contrôle arrivée). C'est une variable Powercenter obligatoire et unique.

★ **Répertoire de stockage des fichiers de lookup : `$PMLookupFileDir`**

Ce répertoire est destiné au stockage des fichiers de lookup utilisés dans les sessions et les workflows Powercenter. C'est une variable Powercenter obligatoire et unique.



REGLES POWERCENTER

A. GESTION DU REFERENTIEL DE SECURITE POWER CENTER

- ★ Un groupe ne peut voir que les Folders qui lui sont attribués, ainsi que les folders communs.
- ★ Les membres d'un groupe affecté à un Folder peuvent Lire, Ecrire et Exécuter tous les objets inclus dans ce folder.

B. ORGANISATION DU DEVELOPPEMENT

- ★ **Le tracing level** : Les objets PowerCenter doivent avoir un Tracing Level normal

Le Tracing Level correspond au niveau de détail de la Log pour cet objet. L'utilisation d'une Log détaillée est justifiable uniquement dans l'environnement de développement. La taille des Logs générées pour un niveau plus fin de détail est trop volumineuse. Le Tracing Level normal inclut toutes les informations nécessaires à la résolution d'erreurs. Afin de faciliter le contrôle et la mise en œuvre de ce Tracing Level, il est demandé de surcharger systématiquement la valeur du tracing Level à « normal » au niveau des sessions.

- ★ **Les objets Aggregator, Joiner, Rank, Sort et Lookup**

Le fichier de stockage pour le cache relatif à ces objets (Cache Directory ou le fichier de stockage temporaire), qui se trouve dans le Temp Directory, doit être positionné sur un répertoire spécifique au projet. Un « File system » spécifique pourra être monté sur ce répertoire pour des besoins spécifiques.

- ★ **Les répertoires spécifiques**

L'ensemble des fichiers Log (session et workflow), les fichiers de rejet, les scripts (SQL, shell, ...) appelés dans Powercenter, les fichiers de paramètres et les fichiers cibles d'une alimentation doivent être positionnés dans les répertoires spécifiques au projet.

- ★ **Erreurs de sessions**

Les sessions doivent être configurées pour s'arrêter après avoir rencontré un nombre d'erreurs fixes. Ce nombre d'erreurs doit être compris entre 1 et 1000.

- ★ **Best practices de développement**

Les sessions et workflows doivent pouvoir être exécutés plusieurs fois à la suite sans générer d'erreur, y compris s'ils sont interrompus en cours de traitement. Les opérations d'update else insert doivent être privilégiées lorsque la volumétrie le permet. Lorsque ce n'est pas le cas, un router doit être utilisé avec lookup préalable sur la cible afin de décider si un enregistrement doit être inséré, mis à jour ou ignoré. Ceci n'est qu'un exemple. Nous vous proposerons dans les fiches suivantes, une fiche dédiées aux Best practices de développement.



GLOSSAIRE

Référentiel POWERCENTER (ou Repository)	Hébergé par une instance de base SGBDR, le référentiel est le lieu de stockage et de centralisation de l'ensemble des informations nécessaires au bon déroulement des traitements (Sécurité, développements, ordonnancement des traitements d'alimentation).
Dictionnaire	Référentiel non actif : aucun moteur de traitement Powercenter n'y est référencé, il n'est donc pas possible de lancer des traitements d'alimentation directement à partir de ce référentiel.
Folder (ou dossier)	Sous-ensemble du Référentiel constitué d'un seul niveau d'arborescence, le folder est dédié à un projet (Datamart) ou un Sujet (Reprise). Il stocke la définition des sources, des cibles, des transformations et des mappings développés par l'application.
Mapping	Brique unitaire de transformation qui traduit les règles de gestion reliant les sources aux cibles.
Mapplet	Sous entité de la notion de mapping réutilisable dans plusieurs mappings
Session	Programme unitaire associé à un et un seul mapping. La session exécute la procédure d'alimentation définit au sein de ce même mapping.
Workflow ou Batch	Suite d'instructions décrivant un processus d'alimentation comprenant le lancement d'une session, l'envoi d'un mail, ou l'exécution d'un script shell.
Worklet	Sous-entité de la notion de workflow réutilisable dans plusieurs workflows
Privilège	Les privilèges affectés à un groupe ou à un utilisateur sont définis sur l'ensemble du référentiel. Ils permettent d'effectuer des actions sur le référentiel et de démarrer et/ou de stopper le serveur Powercenter. Les privilèges donnés au niveau du référentiel ne permettent pas aux utilisateurs d'accéder à des dossiers (folders) particuliers sans permissions supplémentaires.
Permission	Les permissions permettent d'effectuer certaines tâches au sein d'un dossier donné. Il existe trois niveaux de permission sur un Dossier : permissions pour le Propriétaire du Dossier, permissions pour le Groupe d'utilisateurs auquel appartient le Propriétaire et permissions pour le reste des utilisateurs du Référentiel. Ces permissions donnent trois possibilités d'accès au folder : lecture, écriture, exécution.
Owner d'un folder	Propriétaire du folder. Chaque folder est attribué à un groupe, qui désigne un responsable, et le nomme propriétaire du groupe. Celui-ci peut avoir des permissions particulières sur le folder.
Variable Powercenter	Variable utilisée au sein de la session, définie globalement pour l'ensemble des projets exploitant le Référentiel. Elle pointe donc sur un environnement unique, et partageable. Il existe notamment des variables distinctes pour la définition du chemin des fichiers logs, des fichiers de rejets, des fichiers sources, des procédures Shell de pré et post session...
Point de montage	En environnement UNIX, correspond à l'emplacement d'un système de fichiers