

Le potentiel et les défis du Big Data

UIMM

Mardi 2 et mercredi 3 juillet 2013
56 avenue de Wagram 75017 PARIS

EXABUILDER

**Conserver les Big Data,
source de valeur pour demain**

Définir les Big Data

Les Big Data à travers les 4 « V »

Variété

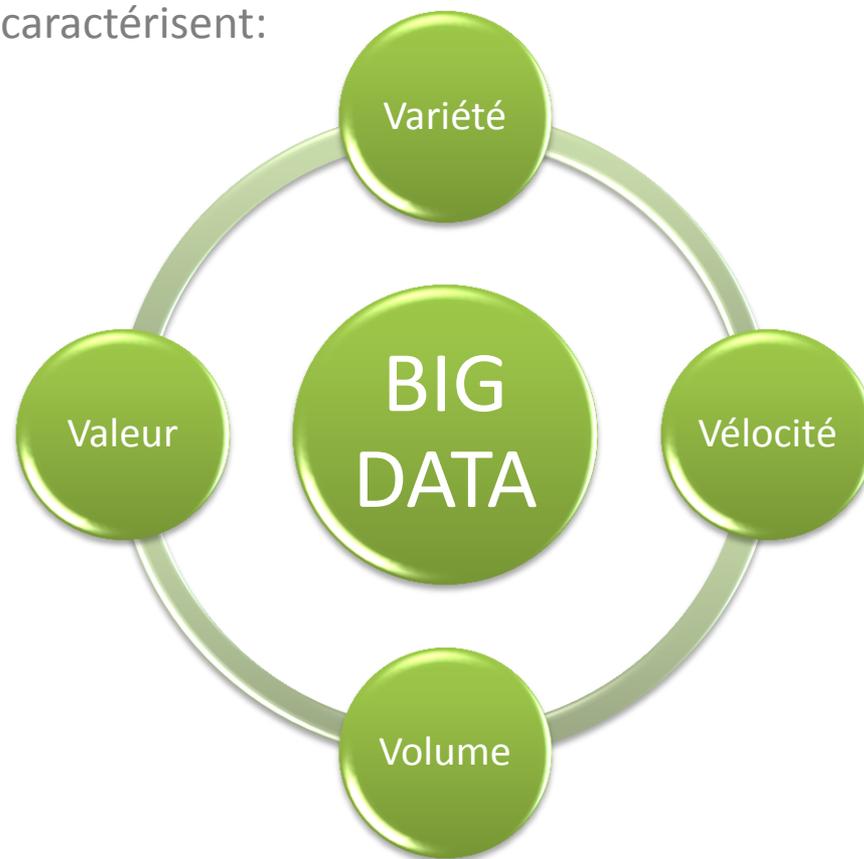
Vélocité

Volume

Valeur

Les Big Data à travers « 4 V's »

Les **Big Data** vont au-delà des grandes quantités d'information, elles représentent de nouveaux ordres de grandeur **de capture, de stockage, de recherche, de partage, d'analyse, de visualisation...** qui doivent s'adapter à 4 grandes dimensions qui les caractérisent:



Les Big Data à travers « 4 V's » : Variété

Variété: les données ne sont plus homogènes, elles sont souvent très hétérogènes et provenant de nombreuses sources, pour être utilisées dans des domaines très variés :

- données structurées, semi structurée, non structurées
- données simples ou complexes,
- de provenance variées,
- générées par des applications ou par l'Homme
- dans des formats hétérogènes
- à densité d'information variée
- ...

Ce qui les rends **difficilement exploitables avec les outils traditionnels.**

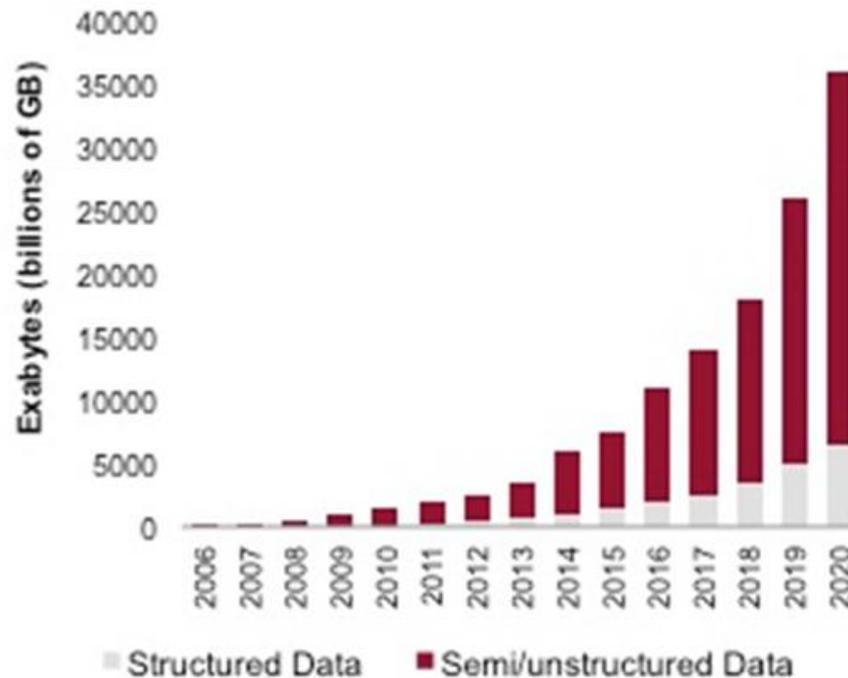
Les Big Data à travers « 4 V's » : Vélocité

Vélocité: le Big Data nécessite donc également une notion temporelle forte associée à de gros volumes. C'est à dire, réagir quasiment en temps réel face à un évènement ou pouvoir le revisiter ultérieurement avec un autre angle de vue:

- identifier une fraude c'est bien! 10 minutes après qu'elle arrive... c'est tard
- décoder le génome prenait 10 ans, il peut désormais l'être en 7 minutes
- la mise en ligne de tous les contenus sur Michael Jackson fut instantanée en temps réel le jour se mort ...

Les Big Data à travers « 4 V's » : Volume

Volume : les volumes de données sont en pleine expansion: de 2,8 Zettaoctets en 2012, ils devraient représenter 35 Zettaoctets en 2020, pourtant la croissance des ressources IT ne devrait **pas dépasser 40%** sur la période. (IDC)



Les Big Data à travers « 4 V's » : Valeur

Valeur : les analyses et la conservation des Big Data doit s'inscrire dans une volonté sous-jacente d'en retirer de la valeur, aujourd'hui ou demain:

- In the developed economies of Europe, government administrators could save more than €100 billion in operational efficiency improvements alone by using Big Data, not including using big data to reduce fraud and errors and boost the collection of tax revenues (*McKinsey&Company*)
- As recently as 2009 there were only a handful of big data projects and total industry revenues were under \$100 million. By the end of 2012 more than 90 percent of the Fortune 500 will likely have at least some big data initiatives under way (*SmartPlanet*)

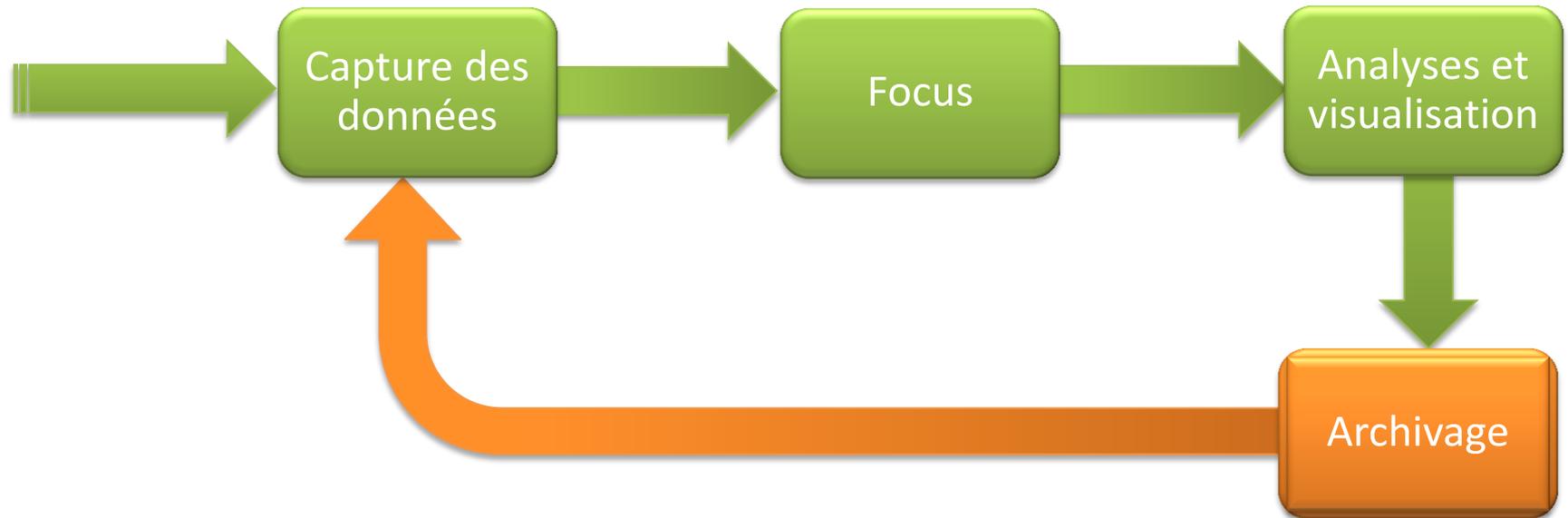
Cycle de vie des Big Data

Big Data Analytics

Conserver, source de valeur... et de coût

Cycle de vie des Big Data

L'analyse des Big Data est souvent vue comme de traitements de flux générés par des machines, comme des logs, des images satellites pour des traitements immédiats...



... mais va aussi servir à des analyses sur d'autres types de données, ou des données plus anciennes qui auront été conservées sur le long terme

94% of Hadoop users perform analytics on large volumes of data not possible before;
88% analyze data in greater detail (*Ventana Research*)

Conserver, sources de valeur...

Lorsque les **principales applications opérationnelles** ont traité les Big Data, les données ont **toujours une valeur** pour des raisons de :

- data mining,
- compliance
- e-discovery
- diagnostics,
- sécurité,
- corrélation,
- applications ou analyses secondaires...

Qui amène la notion de « **vouloir garder** » au lieu de « **devoir garder** », pour de plus en plus d'acteurs dans les organisations.

D'ailleurs **l'archivage** est sans doute la **première brique** de l'infrastructure qui atteindra la statut de « **Big Data** » aussi bien en termes de **volumes** atteignant les PétaOctets, qu'en terme de **variété** d'information hétérogènes qui s'y trouvent.

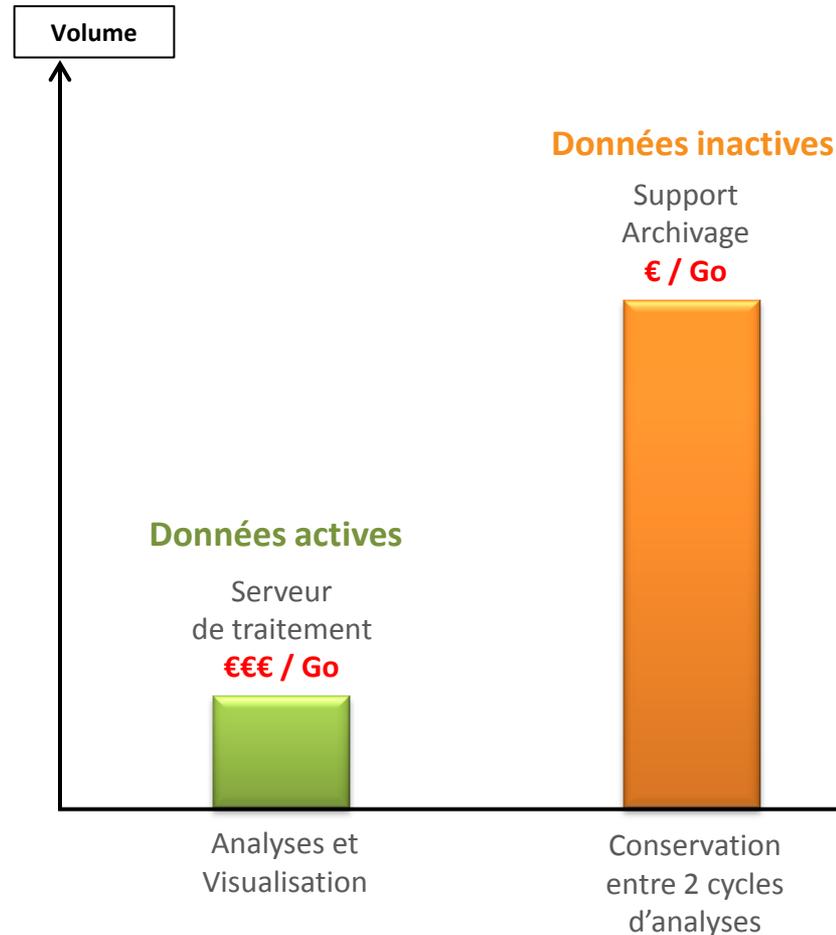
L'archivage va au-delà du stockage

L'archivage des Big Data doit néanmoins s'inscrire dans **une contrainte principale**:

- Une **gestion du stockage** qui permet **d'optimiser les coûts** de la conservation des données inactives en pleine expansion

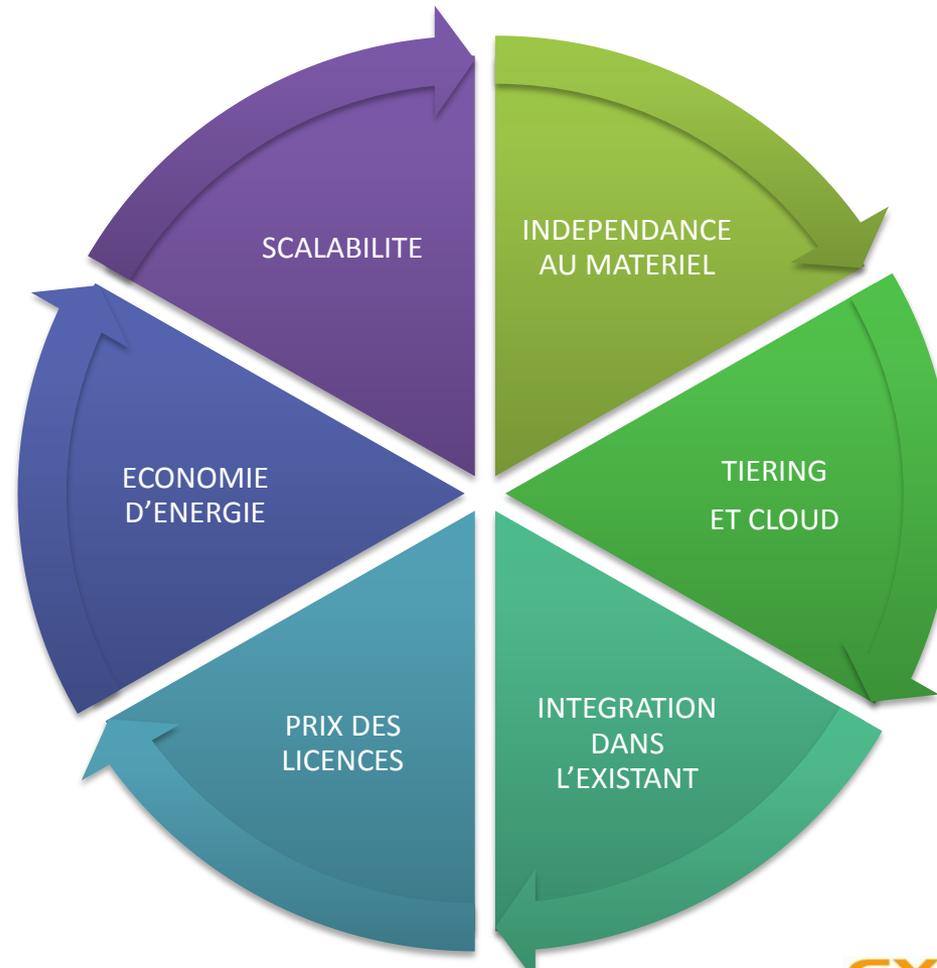
Qui ne doit **pas en faire oublier** une **seconde** primordiale **pour garantir la valeur** :

- Des mécanismes de **garantie d'intégrité et de pérennité des données** conservées



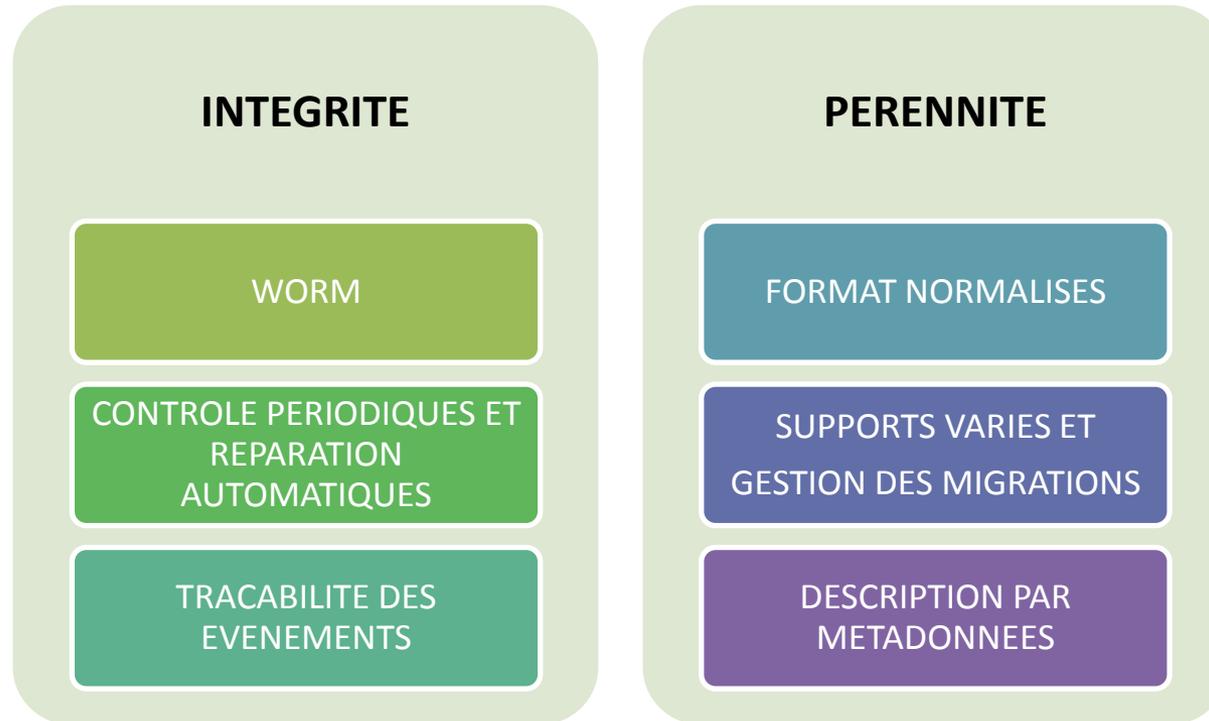
Les besoins induits : stockage économique

Une solution de stockage qui réponde aux nouveaux défis de l'archivage de Big Data, aussi bien en termes techniques qu'économiques, se doit d'offrir :



Les besoins induits : intégrité et pérennité

Mais pour conserver la valeur à la donnée, il faut s'assurer de sa fiabilité, et l'archivage est là pour cela:



Bad data or poor data quality costs US businesses \$600 billion annually (*Fathom*)

Le système Exabuilder

Technologie

Gestion du stockage optimisé

Maintien de la valeur

Technologie

UN MOTEUR D'ARCHIVAGE...



Postes de travail



Scripts



Applications

Guichet Exabuilder

*Valorisation
du contenu*

*Gestion du
cycle de vie*

*Gestion
de la sécurité*

Entrepôts Exabuilder



RAID, NAS, SAN



Bande



Cloud



Autres...

...COMPLETEMENT INTEGRABLE

Postes de travail



Application métier

Moteur d'archivage Exabuilder

Stockage



RAID, NAS, SAN



Bande



Cloud



Autres...

Gestion du Stockage optimisée

Scalabilité

- Virtualisation des supports
- Redimensionnement des ressources transparent et sans interruption
- Forte scalabilité : de 10 To à >> Po et gestion milliards objets
- Gestion automatique des migrations technologiques

Performance

- > 100 Mo /sec par disque et par lecteur, sauf limitation externe
- Performance sur petits fichiers et grands fichiers
- Segmentation des très grandes archives
- Placement optimisé des données
- Extraction au niveau fichier, depuis le support le plus performant

Maîtrise des coûts

- Faibles couts de Licence
- Choix du média le plus adapté
- Réduction de la facture énergétique

Maintien de la valeur

Intégrité des archives

- Archives WORM
- Intégrité garantie par « checksums » (SHA-512) au niveau archive et fichier
- Identification automatique de corruption et réparation automatique

Pérennité des archives

- Format TAR (PAX IEEE 1003.2 POSIX.2) sans limitation, non compressé
- Réversibilité et relogeabilité
- Utilisation de supports pérennes (30 ans)
- Gestion automatique des copies multiples
- Développements sur technologies standards :
- Traçabilité complète des évènements



Quelques exemples d'utilisation



FRANCE GENOMIQUE

- Service mutualisé au Très Grand Centre de Calcul (TGCC) de Bruyères-Le-Chatel pour les 20 principales plateformes françaises de génomique

BULL – SERVIWARE

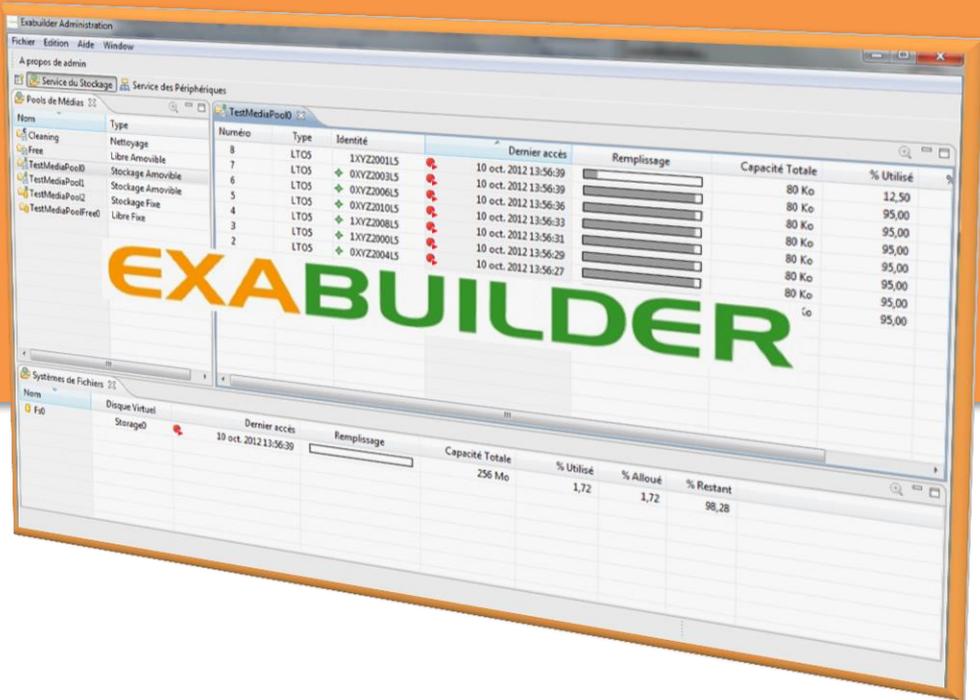
- Option d'archivage intégrée dans eXtreme Factory, l'offre de Cloud HPC de Bull
- Volonté de proposer une solution d'archivage de données pour qu'elles restent au plus prêt du calcul



Merci de votre attention

Annexes

Démonstration produit



Complet pour l'administrateur

Périphériques virtualisés en espaces de stockage

Entrepôts et stratégies d'archivage

Gestion des utilisateurs

Suivi de l'application

Périphériques et espaces de stockage

Les **périphériques** : RAID, NAS, SAN, bandes magnétiques...

... sont **virtualisés** en **espaces de stockage**:

- Pools de médias
- Systèmes de fichiers
- Clouds...

Numéro	Type	Identité	Dernier accès	Remplissage	Capacité Totale	% Utilisé
8	LT05	1XYZ2001L5	10 oct. 2012 13:56:39	<div style="width: 12.50%;"></div>	80 Ko	12.50
7	LT05	0XYZ2003L5	10 oct. 2012 13:56:39	<div style="width: 95.00%;"></div>	80 Ko	95.00
6	LT05	0XYZ2006L5	10 oct. 2012 13:56:36	<div style="width: 95.00%;"></div>	80 Ko	95.00
5	LT05	0XYZ2010L5	10 oct. 2012 13:56:33	<div style="width: 95.00%;"></div>	80 Ko	95.00
4	LT05	1XYZ2008L5	10 oct. 2012 13:56:31	<div style="width: 95.00%;"></div>	80 Ko	95.00
3	LT05	1XYZ2000L5	10 oct. 2012 13:56:29	<div style="width: 95.00%;"></div>	80 Ko	95.00
2	LT05	0XYZ2004L5	10 oct. 2012 13:56:27	<div style="width: 95.00%;"></div>	80 Ko	95.00
1	LT05	1XYZ2005L5	10 oct. 2012 13:56:26	<div style="width: 95.00%;"></div>	80 Ko	95.00

Nom	Disque Virtuel	Dernier accès	Remplissage	Capacité Totale	% Utilisé	% Alloué	% Restant
Fs0	Storage0	10 oct. 2012 13:56:39	<div style="width: 1.72%;"></div>	256 Mo	1,72	1,72	98,28

Entrepôts et stratégies d'archivage

Sur les **espaces de stockage** sont définis des **entrepôts**, au niveau desquels sont définies les **stratégies d'archivage**

The screenshot displays the 'Edition de l'Entrepôt' (Edit Warehouse) dialog box in Exabuilder Administration. The dialog is titled 'Edition de l'Entrepôt' and is for 'Espace de Stockage & Langue d'Indexation'. It features several sections:

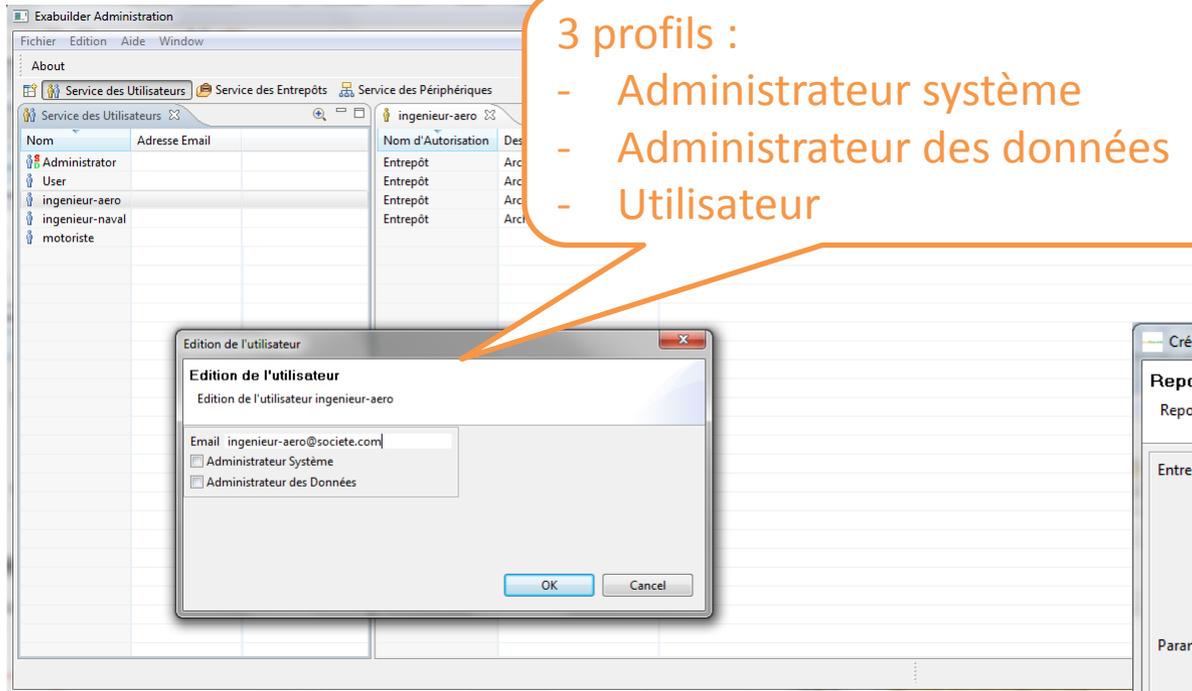
- Choix d'Espace de Stockage:** A list of storage spaces with checkboxes. 'Fs0' is checked. 'TestMediaPool0', 'TestMediaPool1', and 'TestMediaPool2' are also listed.
- Langue d'Indexation:** A dropdown menu set to 'Français'.
- Indexation du contenu des fichiers:** A checked checkbox.
- Taille maximale indexée (Mo):** A numeric input field set to '8'.

Four callout boxes highlight specific settings:

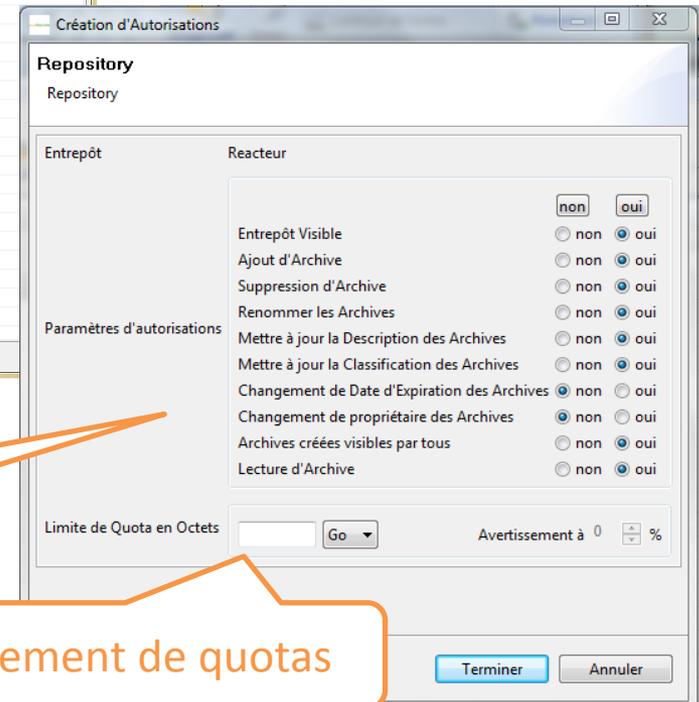
- Durée de rétention:** Points to the 'Durée de rétention' field in the 'Changer la Date d'Expiration des Archives' dialog, which is set to 6 months.
- Nombre de copies:** Points to the 'Durée de rétention' field in the same dialog.
- Type de médias:** Points to the 'Choix d'Espace de Stockage' list in the 'Edition de l'Entrepôt' dialog.
- Paramètres d'indexation :** Points to the 'Langue d'Indexation' and 'Indexation du contenu des fichiers' options in the 'Edition de l'Entrepôt' dialog.

The 'Changer la Date d'Expiration des Archives' dialog also shows a calendar for November 2012 and a 'Délai de grâce' (grace period) of 15 days.

Gestion des utilisateurs



Gestion fine des droits



Positionnement de quotas

Suivi complet

Exabuilder Administration

A propos de admin

Service des Travaux Service du Stockage Service des Périphériques

Travaux Travaux Soumis Travaux en Cours Travaux Terminés

Liste des Travaux	Utilisateur	Description	Status des Travaux	Priorité	Heure de Soumissi
8	ingenieur-aero	Finalisation de l'Archive	Terminé	Elevée	10 oct. 2012 13:56
278	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
279	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
280	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
281	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
284	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
285	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
286	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
287	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
288	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
289	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
270	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
271	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
272	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
273	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
274	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
275	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
0	ingenieur-aero	Déplacement de Médium "0XYZ2003L5" vers Emplacement Libre de Librairie "03584L32 [L0]"	Terminé	Maximale	10 oct. 2012 13:56
276	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
277	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
257	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
258	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
259	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
260	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56
261	ingenieur-aero	Copie de Fichier vers Espace de Stockage "F0"	Terminé	Elevée	10 oct. 2012 13:56
262	ingenieur-aero	Copie de Fichier vers Espace de Stockage "TestMediaPool0"	Terminé	Elevée	10 oct. 2012 13:56

Service des Travaux Paramètres

Traçabilité des évènements sur les archives

Suivi du système

Service des Périphériques Service des Utilisateurs Service des Entrepôts

Action	Date	Utilisateur
Sélection par recherche	7 mars 2013 17:34:45	ingenieur-naval
Lecture des metadonnées	7 mars 2013 17:34:48	ingenieur-naval
Extraction des données	7 mars 2013 17:35:00	ingenieur-naval
Sélection par recherche	7 mars 2013 17:35:26	Administrator
Lecture des metadonnées	7 mars 2013 17:35:41	Administrator

Archives	Taille
1	190 Ko
Cheminee	0 Octet
Cockpit	101,5 Ko
Diesel	418 Ko
Propulsion	2,291 Mo
Proxe	102,5 Ko
Reacteur	166 Ko
TestRepository	0 Octet
Volet	0 Octet

Entrepôts Propriétés

Contenu de l'Archive Historique Propriétés

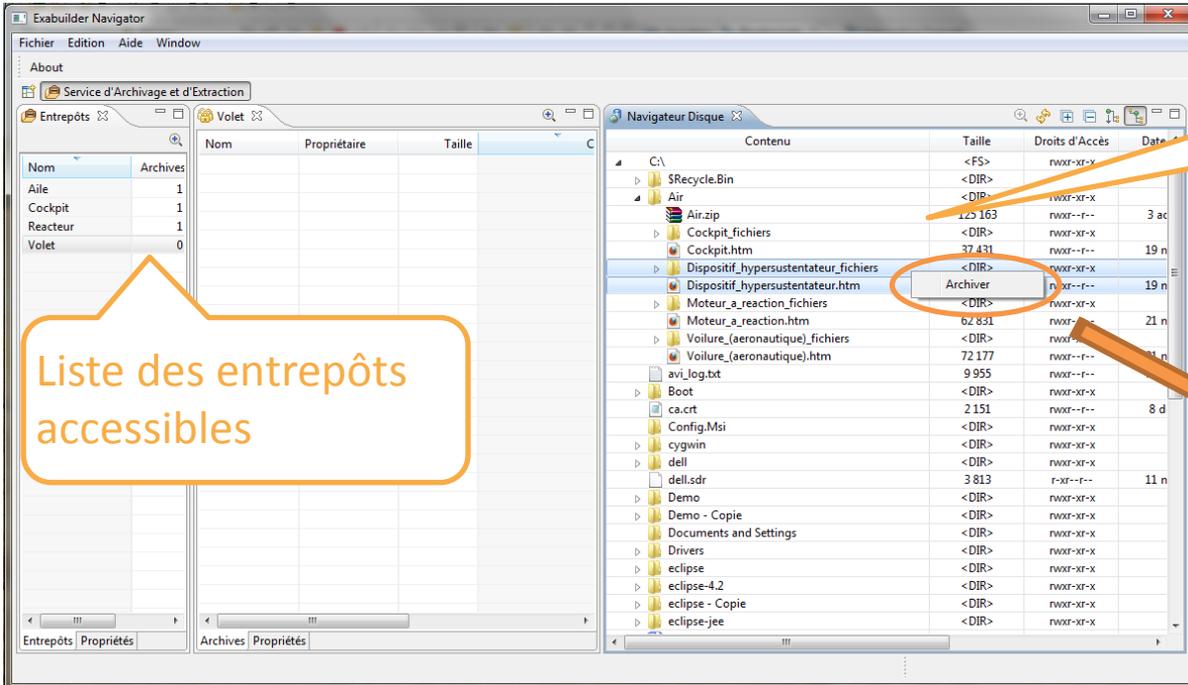
Simple pour l'utilisateur

Archiver

Extraire

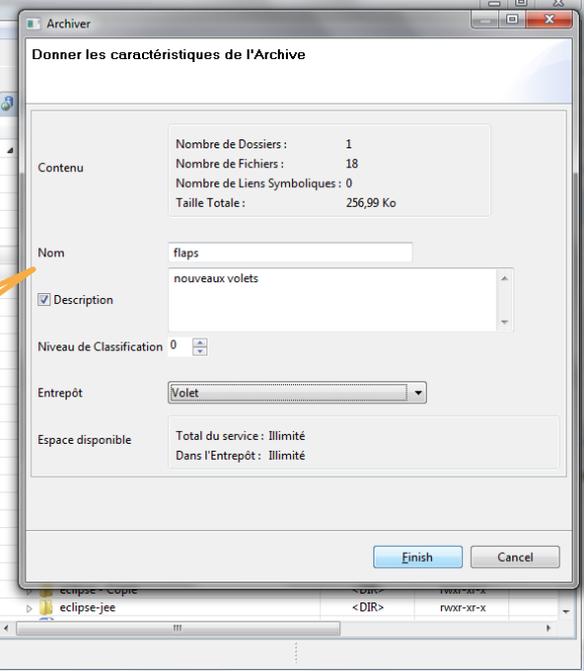
Rechercher

Archiver



Liste des entrepôts accessibles

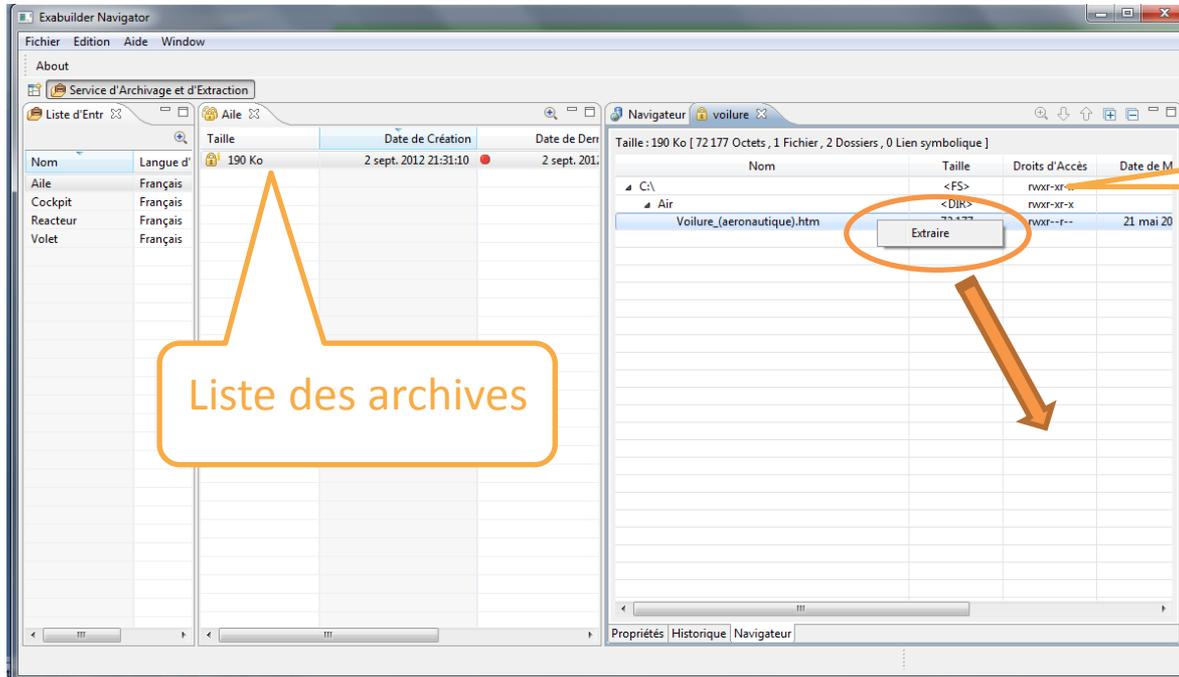
Navigateur disque local ou réseau



Définition des caractéristiques de l'archive :

- nom
- description
- entrepôt de destination

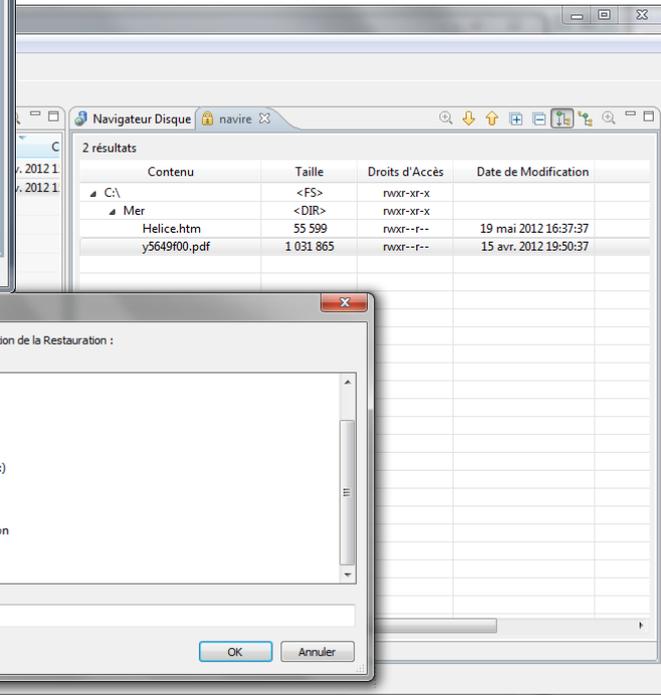
Extraire



Liste des archives

Contenu d'archives

Choix de la destination d'extraction



Rechercher

The image shows a composite screenshot of the Exabuilder Navigator application. It features three main windows and three callout boxes:

- Top Left Window:** A search engine window titled "Moteur De Recherche" with a text input field containing "Hélice" and "OK" and "Cancel" buttons. A callout box above it says "Recherche par mots clés".
- Top Right Window:** A file explorer window showing a table of search results. A callout box above it says "Fichiers contenant les mots recherchés dans métadonnées ou contenu".
- Bottom Left Window:** A file explorer window showing a table of search results. A callout box below it says "Archives contenant les mots recherchés".

The bottom right window shows a detailed view of search results with the following table:

Contenu	Taille	Droits d'Accès	Date de Modification
C:\	<FS>	rwxx-xx-x	
Mer	<DIR>	rwxx-xx-x	
Helice.htm	55 599	rwxx--f--	19 mai 2012 16:37:37
y5649f00.pdf	1 031 865	rwxx--f--	15 avr. 2012 19:50:37