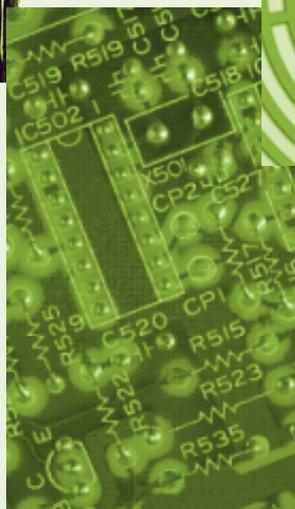
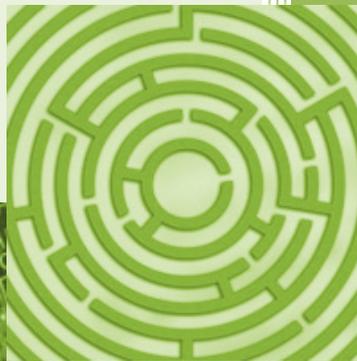


Optimisation et
rationalisation
de l'infrastructure
de stockage ...

Green Storage : Enjeux et Facteurs clés de succès

... pour un
développement
plus durable



Frédéric Laura

Mastère MSIT Executive 2009

HEC - Ecole Supérieure
des Mines de Paris :

« Management des Systèmes
d'Information et des Technologies »

Green Storage, Enjeux et Facteurs Clés de Succès :

De l'optimisation à la
rationalisation de l'infrastructure
de stockage ...

... pour un développement
plus durable



*Cette thèse professionnelle a été réalisée à la Direction des Opérations Infrastructures de la société **Bull**, dans le cadre du Mastère executive «**Management des Systèmes d'Information et des Technologies**», conduit conjointement par l'École des Hautes Etudes Commerciales (**HEC**) et l'École Nationale Supérieure des Mines de Paris (**ENSMP**).*

Tuteur Académique : **Fabien Coelho**
Enseignant-chercheur à l'École Nationale Supérieure des Mines de Paris

Tuteur Professionnel : **Alain Le Billan**
Directeur technique avant-vente stockage Bull

Rédacteur : **Frédéric Laura**
Consultant avant-vente stockage Bull
Frederic.laura@bull.net
Fredericlauramsit@yahoo.fr

Thèse soutenue le 15 avril 2009 à l'ENSMP

REMERCIEMENTS

Je tiens en premier lieu à remercier **Marie-Hélène Delmond**, professeur à HEC et **Robert Mahl** directeur du Centre de Recherche Informatique de l'École des Mines, pour m'avoir accepté au sein du Mastère et prodigué de nombreux conseils riches d'enseignements tout au long du cycle.

Je tiens aussi à remercier vivement **Fabien Coelho**, mon tuteur académique, qui m'a guidé dans l'élaboration de cette thèse professionnelle. Ses savoirs scientifiques et méthodologiques m'ont été très précieux. Sa disponibilité, sa patience, et toutes ses critiques constructives m'ont motivé et soutenu tout au long de la rédaction.

Je remercie particulièrement **Alain Le Billan**, mon tuteur professionnel, pour sa compréhension et le soutien dont il a fait preuve tout au long du Mastère. Je le remercie de m'avoir consacré du temps pour faire le point sur les travaux accomplis et pour tous nos échanges concernant les problématiques techniques et managériales.

Je tiens à saluer tout le personnel HEC Executive Education pour son efficacité et notamment **Angela Combres** pour son support actif et **Nancy Piacentini** pour son important soutien.

J'adresse mes plus vifs remerciements aux professionnels sollicités dans le cadre de l'étude. Leur expérience et disponibilité ont été remarquables, notamment :

- Sophie Bouteiller, chargée de mission au CIGREF
- Jean-Marc Rietsch, président de l'association FedISA
- Bruno Brocheton, DSI EuroDisney
- Philippe Sersot, Directeur de Production CALYON et président du CRIP

Je remercie aussi chaleureusement tous mes collègues pour le temps accordé et la qualité de leurs aides lors de mes nombreuses sollicitations et plus particulièrement :

- Benjamin Bergeron
- Francis Bianchin
- Pascal Langreau
- Ahcene Latreche
- Philippe Reynier

Je tiens aussi à remercier chaleureusement mes collègues et managers qui m'ont permis de réaliser le Mastère :

- Juliette Arnould et Véronique Beaumont
- Alain Le Billan et Christian Du Pontavice
- André Hélie Joly et Joël Pons

Enfin, je remercie de tout mon coeur, mes plus précieuses motivations : Estelle et Rémi, pour leur soutien et inestimable présence, ainsi que mes parents et mon frère, pour la valeur de leur affection.

RESUME EXECUTIF

Green IT, effet de mode ou réalité ?

La crise environnementale s'accompagne aujourd'hui d'une crise économique qui nous force à prendre conscience des conséquences du développement effréné des systèmes informatiques, des bénéfices potentiels liés à l'amélioration de leur efficacité, et de leur capacité à faire progresser de nombreux domaines, pour un développement plus durable.

Les data centers rejettent 170 millions de tonnes de CO₂ par an, c'est plus que les rejets de l'Argentine. Leur rendement est médiocre : en moyenne, seulement 1 watt sur 3 est utilisé par les systèmes d'information, le reste étant consommé par l'infrastructure énergétique.

Toute amélioration du data center doit aussi prendre en compte l'infrastructure de stockage qui porte une part importante de responsabilité :

- les données ne cessent de croître : plus de 30 % par an. Elles sont vitales pour l'entreprise et leur protection est un enjeu majeur quotidien ;
- s'adaptant à l'évolution actuelle de l'activité, les ressources de stockage fonctionnent toutes quasiment en continu, même lorsque leurs données ne sont pas accédées ;
- comme les serveurs, le stockage est souvent utilisé à 20 à 40 % de sa capacité ;
- plus de la moitié des données sont sur un espace de stockage inapproprié à leur valeur.

Pour rendre le stockage véritablement « éco-responsable », les facteurs de succès sont multiples : matériels, logiciels, architectures, auxquels il faut associer une réelle gouvernance des données, visant à canaliser en amont leur prolifération.

L'objectif de la thèse est d'identifier les systèmes, fonctionnalités, architectures, recommandations et meilleures pratiques, optimisant et rationalisant l'infrastructure de stockage pour un développement plus durable.

Quatre parties principales la composent. L'introduction souligne l'impact environnemental des T.I.C.¹ et des systèmes de stockage, le développement des valeurs « vertes » et des réglementations, et apporte une définition du Green IT et du Green Storage.

L'étape suivante est l'inventaire des problématiques clés : l'augmentation de la consommation énergétique, la densité des systèmes, et la saturation du datacenter ; les pollutions ; les systèmes isolés, dispersés et leur sous utilisation ; l'explosion des volumétries, des données redondantes et dormantes.

L'analyse des problématiques discerne ensuite les recommandations et meilleures pratiques : implémentation dans le data center ; composants et systèmes plus économes ; architectures : consolidation, virtualisation et mutualisation ; fonctionnalités optimisant les volumétries ; rationalisation des espaces de stockage, usage de la sauvegarde et de l'archivage ; gouvernance des données ; catalogue de services de stockage ; limites des solutions proposées.

La dernière partie expose l'avenir du stockage « vert », les autres voies de recherche à approfondir ainsi que les développements possibles issus du travail de la thèse, comme de nouveaux contenus à intégrer aux réponses à appels d'offres, et la création d'un outil d'évaluation de l'état de la gouvernance de l'infrastructure de stockage, à destination de la DSI : un « baromètre Green Storage ».

¹ Technologies de l'Information et de la Communication

TABLE DES MATIÈRES

REMERCIEMENTS.....	VII
RESUME EXECUTIF.....	11
CONTEXTE PROFESSIONNEL.....	19
<hr/>	
ENJEUX ECOLOGIQUES DES TECHNOLOGIES DE L'INFORMATION ET DES SYSTEMES DE STOCKAGE DE DONNEES.....	21
<hr/>	
1.1 INTRODUCTION.....	22
1.1.1 Enjeux écologiques.....	24
1.1.2 Enjeux économiques.....	25
1.2 POLLUTIONS DES T.I.C.....	26
1.2.1 Typologie des pollutions.....	26
1.2.2 Analyse du cycle de vie.....	28
1.3 LES ACTEURS.....	28
1.3.1 Typologie des systèmes de stockage des données.....	28
1.3.2 L'évolution vers des valeurs plus « vertes » des principaux acteurs.....	29
1.4 REGLEMENTATIONS « VERTES » DES T.I.C.....	31
1.4.1 Réglementations environnementales.....	31
1.4.2 Les normes « vertes ».....	32
1.4.3 Le programme « Code of Conduct » européen.....	33
1.4.4 Développement durable et « Corporate Social Responsibility ».....	33
1.4.5 Le développement du Green IT en Europe et en France.....	34
1.5 DEFINITIONS DU « GREEN IT » ET DU « GREEN STORAGE ».....	35
1.5.1 « Green IT ».....	35
1.5.2 « Green Storage ».....	37
<hr/>	
IDENTIFICATION DES PROBLEMATIQUES ET DES FACTEURS CLES.....	39
<hr/>	
2.1 L'EVOLUTION DES SYSTEMES D'INFORMATIONS ET DU DATA CENTER.....	41
2.2 L'EVOLUTION DES SYSTEMES DE STOCKAGES DE DONNEES.....	41
2.2.1 Le m ² du data center toujours plus cher et consommateur d'énergie.....	43
2.2.2 Évolution de la consommation énergétique des systèmes de stockage.....	44
2.2.3 Analyser la consommation énergétique du data center.....	44
2.2.4 Mesure énergétique des infrastructures de stockage.....	46
2.2.5 Outil de reporting et management énergétique.....	47
2.2.6 Difficulté d'approvisionnement énergétique.....	47
2.3 ENCOMBREMENT SPATIAL.....	48
2.4 POLLUTIONS.....	49
2.5 ANALYSE DU CYCLE DE VIE (ECOBILAN).....	50
2.6 ARCHITECTURES DE STOCKAGE.....	51
2.6.1 Des infrastructures dispersées.....	51
2.6.2 Sous utilisation.....	52
2.7 EXPLOSION DU VOLUME DES DONNEES.....	53
2.7.1 Diversité des données.....	54
2.7.2 Données redondantes.....	55

2.7.3	Données dormantes ou persistantes	57
2.7.4	Restitution d'enquête sur les problématiques de gouvernance des données	58
2.8	ÉVOLUTION DU ROLE DE LA DSI.....	59
2.9	DIFFERENCIER LES PERFORMANCES GREEN.....	60
2.10	LA RESPONSABILITE SOCIETALE D'ENTREPRISE.....	62
2.11	METHODOLOGIE DE COLLECTE ET D'IDENTIFICATION DES PROBLEMATIQUES ET FACTEURS CLES.....	63

**FACTEURS CLES DE SUCCES « GREEN STORAGE»,
RECOMMANDATIONS ET MEILLEURES PRATIQUES.....65**

3.1	IMPLEMENTATIONS, TECHNOLOGIES ET COMPOSANTS	67
3.1.1	Implémentation des systèmes dans le data center	67
3.1.2	Technologies réduisant l'empreinte spatiale dans le data center	68
3.1.3	Technologies et composants optimisant l'efficacité énergétique.....	70
3.1.4	Évaluer la consommation énergétique.....	73
3.2	ARCHITECTURES ET FONCTIONNALITES	75
3.2.1	Consolidation.....	75
3.2.2	Les fonctionnalités logicielles optimisant les volumes de données.....	75
3.2.3	Les systèmes disques.....	80
3.2.4	Virtualisation	81
3.2.5	Réseaux de stockage.....	82
3.2.6	Architecture organisée par « Tiers » et « HSM énergétique ».....	83
3.2.7	Contrôle du cycle de vie des données.....	86
3.2.8	Rationalisation des investissements selon la valeur des données	87
3.2.9	Mutualisation	88
3.3	CANALISER LA CROISSANCE DES DONNEES	89
3.3.1	La Gouvernance des données.....	90
3.3.2	Une « Green Data Governance »	94
3.3.3	Optimiser les besoins des directions métiers	95
3.3.4	Un catalogue de solutions de stockage orienté « Niveaux de Service ».....	96
3.4	AUDITS, OUTILS ET PROGRAMME D'AMELIORATION	99
3.4.1	Audit de l'infrastructure électrique et de la climatisation	99
3.4.2	Audit spécifique et approfondi de l'infrastructure de stockage	100
3.4.3	Outils de gestion de l'efficacité du data center.....	100
3.4.4	Programme d'amélioration basé sur le volontariat.....	101
3.5	EXEMPLE DE BALANCED SCORECARD "GREEN" (BSC)	102
3.6	CRÉATION D'UNE BALANCED SCORECARD "GREEN STORAGE"	104
3.6.1	Balanced Scorecard Green Storage : initiative « Data Governance".....	106
3.7	L'INITIATIVE « GREEN STORAGE » DU SNIA	106
3.8	RESPONSABILITE SOCIETALE D'ENTREPRISE, DEVELOPPEMENT DURABLE ET GREEN IT	107
3.9	RECOMMANDATIONS D'ASSOCIATIONS DES T.I.C.	108
3.9.1	Le CIGREF.....	108
3.9.2	Le CRIP	108
3.9.3	La FedISA.....	109
3.10	SYNTHESE DES STRATEGIES GREEN STORAGE	110
3.10.1	Maximiser le taux d'utilisation des infrastructures de stockage.....	110
3.10.2	Déplacer les données vers l'espace de stockage adapté	111
3.10.3	Accroître la densité utile du stockage.....	111

3.10.4	Evoluer vers une infrastructure de stockage orientée service	111
3.11	SYNTHESE ILLUSTREE DES STRATEGIES “GREEN STORAGE”	112
3.12	SYNTHESE ET CLASSEMENT DES FACTEURS CLES DE SUCCES	113
3.13	LIMITE DES SOLUTIONS ET RECOMMANDATIONS PROPOSEES	115

CONCLUSION.....	119
------------------------	------------

4.1	UN MARCHÉ EN PLEIN DEVELOPPEMENT	121
4.2	LE PARADOXE DES T.I.C. ET DE L’ÉCOLOGIE	122
4.3	SOLUTIONS : DU VERT CLAIR AU VERT FONCE	123
4.4	LES PRINCIPALES VOIES DE DEVELOPPEMENT A APPROFONDIR	124
4.5	SUGGESTIONS ET DEVELOPPEMENTS POSSIBLES ISSUS DE LA THESE	125
4.5.1	Apport de nouveaux contenus aux réponses à appels d’offres	125
4.5.2	Elaboration d’un baromètre d’évaluation Green Storage	126
4.6	EN CONCLUSION.....	128
BIBLIOGRAPHIE – SOURCES		131
ANNEXES		135
ABREVIATIONS		138
TABLE DES ILLUSTRATIONS		140

CONTEXTE PROFESSIONNEL

Bull avec un chiffre d'affaires de 1 132 millions d'euros en 2008, détient une position de leader européen, maîtrisant tous les éléments de la chaîne de valeur informatique et aidant les entreprises et les administrations à bâtir, optimiser, exploiter et rentabiliser leurs systèmes d'information.

Bull Storeway, intégrateur de solutions de stockage

Avec plus de 1000 références client en France, des activités de recherche, de conseil, de service et de maintenance soutenues par 200 collaborateurs, les objectifs de Bull sont d'apporter des solutions de stockage permettant de :

- **faire plus** : contrôler les coûts, optimiser le patrimoine, accroître la flexibilité, gagner en efficacité énergétique ;
- **faire mieux** : gérer le risque, délivrer les niveaux de service requis, démultiplier la valeur de l'information, respecter l'environnement.

Ces objectifs, stimulés par les crises économiques et écologiques, sont les priorités essentielles des clients de Bull, de leurs systèmes d'information et de leurs data centers.

Bull Storeway a la mission d'aider ses clients à réduire leurs coûts, tout en respectant leurs contraintes de productivité et en limitant leur impact sur l'environnement.

Pour répondre à ces enjeux, toute son activité de conseil, de l'ingénieur technico-commercial au consultant et jusqu'à l'architecte, doit appréhender de nombreuses problématiques et connaître les solutions et meilleures pratiques pour optimiser et rationaliser les infrastructures de stockage.

Cette thèse porte donc l'espoir, d'être un outil contribuant à la réalisation de nos objectifs et de ceux de nos clients.

ENJEUX ECOLOGIQUES DES TECHNOLOGIES DE L'INFORMATION ET DES SYSTEMES DE STOCKAGE DE DONNEES

Ce chapitre introduit la problématique écologique actuelle et les impacts sur l'environnement liés aux formidables développements des Technologies de l'Information et de la Communication, et notamment celui du stockage des données.

Seront abordés...

- les enjeux écologiques, énergétiques et économiques ;
- la pression « verte » due au changement des attentes du marché et aux réglementations sans cesse croissantes ;
- la définition de l'informatique verte : « green IT » ;
- et de « l'éco-stockage », plus souvent appelé « green storage ».

"Les T.I.C. sont responsables de 2 % des émissions mondiales de CO₂, autant que l'aviation" Gartner²

1.1 Introduction

281 exaoctets de données numériques en 2008 et près de 1 800 en 2011...

« How Much Information ? »³ est un ambitieux programme de l'université américaine de Californie ayant pour objet de comptabiliser le développement de l'univers numérique.

Pour mieux comprendre la signification de certaines valeurs de stockage exprimées, en voici un bref rappel imagé avec du sable (source EMC), si un Bit⁴ d'information est égal à un grain de sable⁵ alors :

Kilo-octet (Ko) : 1 000 octets ou 10³ octets
15 Ko : une page de traitement de texte 100 Ko : une photographie basse résolution
Méga-octet (Mo) : 1 000 000 octets ou 10⁶ octets
1 Mo : une cuillère de sable 1 Mo : un petit roman 10 Mo : une minute de son haute-fidélité
Giga-octet (Go) : 1 000 000 000 octets ou 10⁹ octets
1 Go : un tas de sable de 30 cm de côté 100 Go : un étage d'une bibliothèque rempli de revues académiques
Tera-octet (To) : 1 000 000 000 000 octets ou 10¹² octets
1 To : un bac à sable de 50 m ² et de 30 cm de profondeur 1 To : 50 000 arbres transformés en papier et imprimés 400 To : La base de données du National Climatic Data Center
Peta-octet (Po) : 1 000 Tera-octets or 10¹⁵ octets
1 Po : une plage de 1,6 km de long, 30 m de large et 30 cm de profond. 2 Po : toutes les bibliothèques de recherche académique américaines

IDC a publié en mars 2008 un livre blanc significatif sur l'importance de la croissance des données issues de nos univers numériques : "The Diverse and Exploding Digital Universe"⁶ qui estime le total des données numériques à 281 exabytes⁷ soit 281 milliards de gigabits. D'ici 2011, l'univers numérique devrait être égal à 10 fois la taille de celui de 2006, soit 1 800 exabytes pour une croissance annuelle de 60 %.

² <http://www.gartner.com> (source 2007 tenant compte de l'ensemble du cycle de vie des produits)

³ http://hmi.ucsd.edu/howmuchinfo_work.php

⁴ Bit : contraction de binary et digit, chiffre binaire 0 ou 1, également l'unité de mesure informatique désignant la quantité élémentaire d'information représentée par un chiffre du système binaire (source Wikipédia)

⁵ Source EMC Forum 2008

⁶ http://education.emc.com/content/common/docs/ism/ISM_diverse-exploding-digital-universe.pdf

⁷ Soit 2,81 x 10²¹ bits

Pour avoir un ordre de grandeur :

- 1 kilo-octet (Ko) = 2^{10} octets = 1 024 octets, soit 2 à la puissance 10.

Une page de texte issue d'un traitement de texte « pèse » en moyenne de 15 Ko.

- 1 exa-octet (Eo) = 2^{60} octets = 1 024 Po (pétaoctet) soit 1 152 921 504 606 846 976 octets.

<i>Exa-octet (Eo) : 1 000 000 000 000 000 000 octets ou 10^{18} octets</i>
1 Eo : une plage de 1600 km de long, 30 m de large et 30 cm de profondeur 2 Eo : Tout le volume d'information généré en 1999
<i>Zetta-octet (Zo) : 1 000 Exa-octets ou 10^{21} octets</i>
1 Zo : une plage couvrant toute la côte des États-Unis sur 30 cm de profondeur et 1,2 km de large
<i>Yotta-octet (Yo) : 1 000 Zeta-octets ou 10^{24} octets</i>
1 Yo : assez de sable pour enfouir toute la surface des États-Unis sous 100m

Le développement accéléré de l'usage des caméras (ex : vidéosurveillance) et des chaînes de télévisions numériques sont parmi les facteurs les plus contributeurs de cette formidable croissance. Selon IDC, si la masse numérique créée en 2006 était imprimée, elle serait égale à 72 piles de livres hautes de 150 millions de kilomètres (distance de la terre au soleil).

IDC mentionne d'autres accélérateurs comme nos besoins personnels en protection de données et en mobilité : téléphone portable, ordinateur portable, PDA et le développement de technologies comme la RFID⁸ et le VoIP⁹...

Pour nous rendre compte, simplement de la part de notre contribution « personnelle » au développement de cet univers, testez le programme « *Personal Digital Footprint Calculator* »¹⁰ développé en 2007 par IDC. Ce programme évaluera vos activités numériques quotidiennes et vous fournira en mégabits par seconde, le résultat de vos volumétries de données générées par vos photographies, appels téléphoniques, téléchargements de musique ou vidéo et courriers électroniques. Le résultat est instructif et l'on peut obtenir raisonnablement un total quotidien de 2 084 méga-octets !

Il convient d'ajouter à cet impact personnel, notre « ombre numérique » générée par nos apparitions dans les enregistrements de vidéosurveillance, nos opérations bancaires et téléphoniques, nos relevés de consommations énergétiques, nos données médicales (bilans, scanners), la liste est ici loin d'être exhaustive...

➔ *En 2007, le Gartner nous alertait : « les T.I.C. polluent autant que les avions »¹¹ et représentent 2 % des rejets en dioxyde de carbone (CO₂) liés à l'activité humaine - soit la même pollution que celle provoquée par toute l'aviation. Greenpeace pourrait bientôt déployer ses Rainbows Warriors devant les data centers...*

⁸ RFID : Radio Frequency Identification

⁹ VoIP : voix sur réseau IP

¹⁰ www.emc.com/digital_universe

¹¹ Le Gartner Group tient compte du cycle de vie complet des PC, serveurs, imprimantes et téléphones portables ainsi les infrastructures techniques nécessaires à leur utilisation : backbone internet, réseau fibre, centres d'hébergement et d'appels...

1.1.1 Enjeux écologiques

Les activités de l'homme impactent gravement l'environnement : d'après le groupe d'expert intergouvernemental sur l'évolution du climat (GIEC), les niveaux d'émissions de CO₂ augmentent ainsi que les températures de l'air et des océans. En conséquence, **nous sommes responsables** de la hausse du niveau de la mer, de la fonte des calottes glaciaires et des conditions météorologiques plus extrêmes.

Le rapport du GIEC

Créé en 1988 à l'initiative du G7¹², le GIEC¹³ a été conjointement mis en place par l'Organisation Météorologique Mondiale (OMM) et le PNUE : Programme des Nations-Unies pour l'environnement. Le GIEC communique régulièrement un rapport¹⁴ sur l'évolution du climat, auquel participe 2500 scientifiques internationaux.

Le gouvernement français comme 191 autres états participants restitue au public des synthèses¹⁵ mettant en évidence **les réalités du réchauffement climatique** dont le tableau suivant, qui tient compte d'une possible élévation de température entre 1 et 5 degrés :

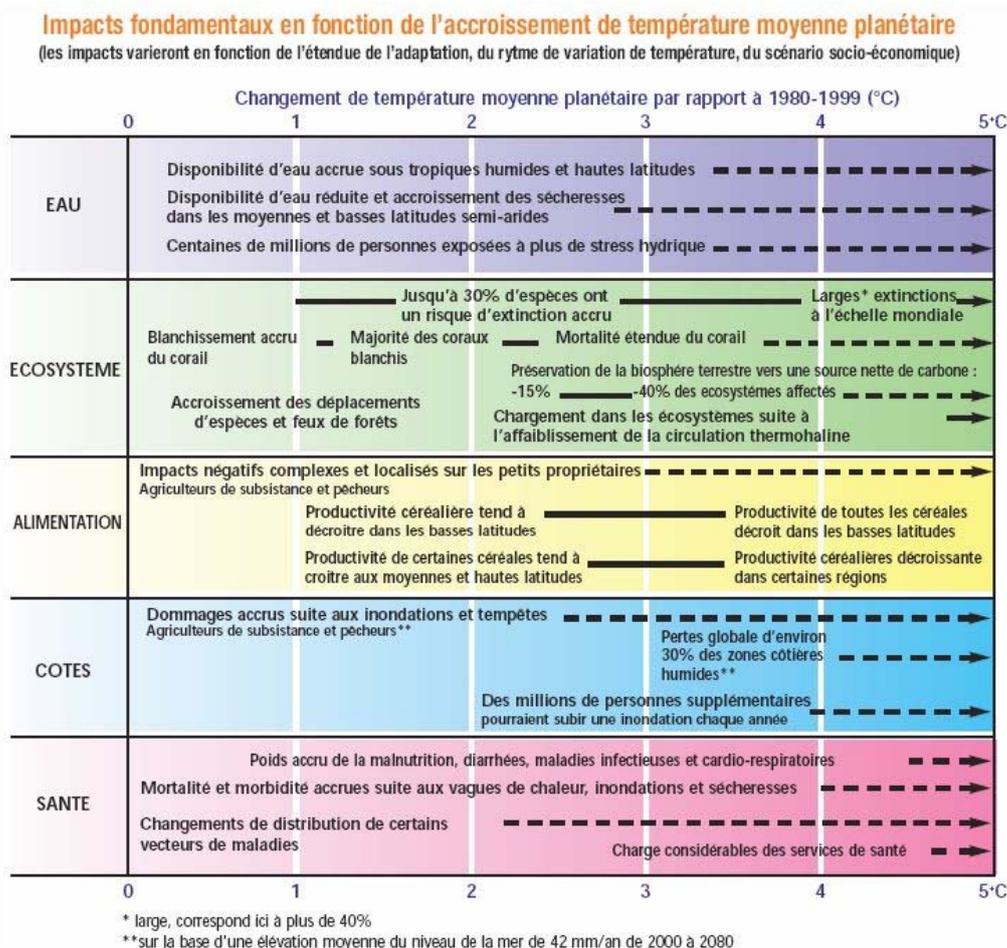


Figure 1. Extrait du 4e rapport d'évaluation du climat, GIEC, 2007

¹² G7 : Groupe des 7 pays les plus riches : USA, Japon, Allemagne, France, Grande-Bretagne, Canada, Italie

¹³ En anglais, le GIEC devient IPCC : Intergovernmental Panel on Climate Change

¹⁴ http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_spm.pdf

¹⁵ http://www.effet-de-serre.gouv.fr/images/documents/DOC_20GIEC_20INTERNET%5B1%5D.pdf

Entre 1,1 et 2,9 °C de hausse de la température : c'est le scénario le plus optimiste d'ici 2100 ans contre 2,4 à 6,4°C pour le plus pessimiste.

Quelles sont les caractéristiques du phénomène de changement climatique ?

« L'atmosphère de la Terre peut être comparée à une couverture qui retient, en partie, grâce à la présence de gaz à effet de serre, la chaleur rayonnée par la planète. Certains de ces gaz à effet de serre sont émis en proportion croissante par les activités humaines, dont pour 77 % le gaz carbonique venant de la combustion d'énergie fossile et de la déforestation, et intensifient donc ce phénomène de « couverture » depuis deux siècles. Il en résulte un réchauffement de la surface de la Terre. »

Le rapport détaille les effets du changement climatique observés : onze des douze dernières années sont parmi les plus chaudes depuis 1850 ; au cours du XXe siècle, le niveau de la mer s'est élevé de 17 cm et de 3 mm par an entre 1993 et 2003 : le double de la moyenne constatée au XXe siècle ; en 100 ans, le climat s'est réchauffé de 0,74 °C.

Quels constats pour les émissions de gaz à effet de serre ?

Les émissions sont constituées de dioxyde de carbone (CO₂), de méthane (CH₄), de protoxyde d'azote (N₂O) et de différents gaz frigorigènes (HFC, PFC et SF₆). Leurs émissions ont augmenté de 70 % entre 1970 et 2004. Les gaz carboniques ont atteint une concentration de 379 ppm¹⁶ en 2005 contre 280 ppm avant l'ère industrielle, le plus haut niveau constaté depuis 650 000 ans.

➔ *Les répercussions des activités humaines sur la planète sont indéniables et particulièrement depuis l'ère industrielle. Les effets des émissions de CO₂ sont les plus dommageables pour l'environnement. Continuer à dégrader notre environnement mettra en péril notre propre survie liée à celle de l'écosystème.*

1.1.2 Enjeux économiques

Des prévisions toujours en forte croissance pour les systèmes de stockage

IDC a publié en juin 2008 l'étude « *The Real Costs to Power and Cool All the World's External Storage* »¹⁷ dont certaines informations sont révélatrices : « Sur les 5 prochaines années (2008-2012), l'industrie du disque dur fournira 8 fois ce qu'elle avait fourni dans les 11 dernières années » et « Avec un coût moyen mondial de 0,07 \$ par kilowatt/heure en 2007, le coût global pour alimenter et climatiser tous les disques durs dépasse le milliard de dollars ».

« La capacité totale des disques durs livrés atteint 1 642 Petabytes¹⁸, augmentant de 51,8 % d'une année sur l'autre ».

Les technologies Green IT sont renforcées par la crise économique

Selon Frank Gens, vice-président senior du cabinet d'étude IDC, en 2009, la transformation des systèmes d'informations vers des modèles plus « verts » ne sera pas seulement une source de coût, mais contribuera à permettre la réduction des dépenses.

¹⁶ Ppm : partie par million soit 1 ppm correspond à une molécule par million

¹⁷ Source : document n° 212714, www.IDC.com

¹⁸ Petabyte : 10¹⁵ octets = 1000 To = 1 000 000 000 000 000 octets

→ En effet selon IDC, en 2009, une majorité d'entreprises connaîtra des restrictions budgétaires favorisant des investissements dans des projets leur permettant de réaliser des économies comme le Green IT.

1.2 Pollutions des T.I.C.

Les T.I.C. polluent dès leur fabrication, pendant leur utilisation (énergie, chaleur) et jusqu'à leur fin de vie avec des déchets très toxiques et coûteux à traiter.

1.2.1 Typologie des pollutions

Les systèmes informatiques contiennent des substances toxiques que la nature ne peut éliminer d'elle-même : mercure, plomb, cadmium, polybromodiphényléthers, polybromobiphényles (PBB)...

Tant lors de la fabrication de produits informatiques, que pendant leur utilisation (consommation électrique), des NOx sont émis : ces gaz irritants pénètrent dans les voies respiratoires provoquant des problèmes de santé, mais aussi les NOx favorisent la formation d'ozone dans les basses couches de l'atmosphère. Pour finir, ces gaz contribuent aux retombées acides et à l'eutrophisation.

Les NOx¹⁹ proviennent essentiellement du transport routier (en baisse grâce au pot catalytique), du secteur de l'industrie et de la production d'énergie : combustions fossiles (gaz naturel, pétrole, charbon) comme certaines centrales thermiques, installations de chauffage et usines d'incinérations.

Du fait de la nature de l'électricité française : le nucléaire, nos data centers sont peu responsables des émissions atmosphériques polluantes comme les NOx²⁰.

Les déchets électroniques

Le Laboratoire Fédéral Suisse d'essai des matériaux et de recherche estime dans son étude « Global Perspectives on E-Waste »²¹ déjà publiée depuis 2005, que 500 millions de PC en fin de service entre 1994 et 2003 contiennent : 2 872 000 tonnes de plastiques, 718 000 tonnes de plomb, 1 363 tonnes de cadmium, 863 tonnes de chrome et 287 tonnes de mercure.

Les pollutions liées à la production de l'énergie électrique

L'illustration suivante est issue du panorama de l'électricité mondiale par type d'énergie de l'OCDE :

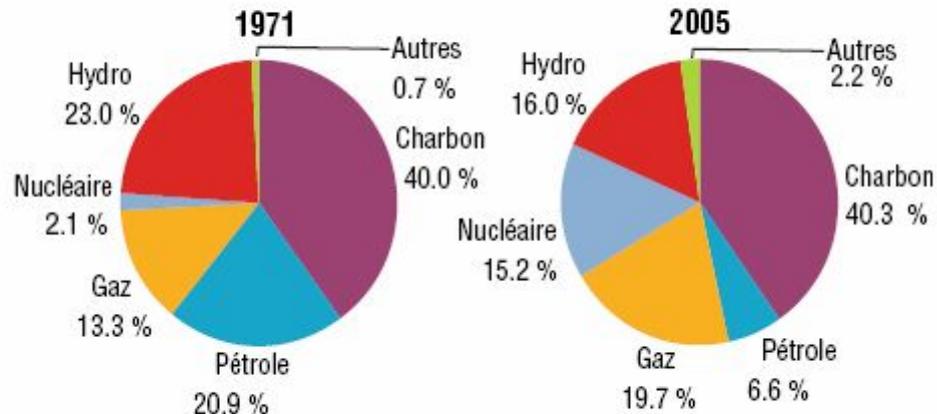
¹⁹ NOx : monoxyde d'azote (NO) et dioxyde d'azote (NO2)

²⁰ Le programme national de réduction des émissions de polluants atmosphériques (SO2, NOx, COV, NH3) en application de la directive 2001/81/CE du 23/10/2001 est en ligne sur www.ecologie.gouv.fr

²¹ http://www.empa.ch/plugin/template/empa/*/51486/---/l=2

Production d'électricité mondiale par type d'énergie

En pourcentage de la production d'électricité mondiale



StatLink <http://dx.doi.org/10.1787/387076311823>

Figure 2. Production d'électricité mondiale par type d'énergie

Le charbon est toujours en 2005, la première matière utilisée pour la production d'électricité suivi de près par le gaz et le pétrole. C'est donc majoritairement une énergie fossile, responsable en grande partie des émissions de CO₂, qui est utilisée pour créer de l'électricité.

Les principaux pays producteurs d'électricité n'ont pas les mêmes responsabilités et enjeux environnementaux selon la nature de l'énergie utilisée²² :

	Charbon	Hydrocarbure	Électronucléaire
États unis	53 %	26 %	22 %
Europe	33 %	21 %	35 %
Chine	78 %	4 %	2 %
Inde	81 %	12 %	2 %
Russie	22 %	62 %	11 %

En France 77 % de l'électricité est d'origine électronucléaire.

Les enjeux écologiques et économiques sont ainsi très différents selon la nature de l'électricité employée et le pays où elle est consommée. Climatiser un data center en Afrique du Nord est plus difficile en raison de la chaleur et plus impactant pour l'environnement du fait que l'énergie électrique sera produite dans des centrales utilisant du charbon ou du pétrole.

²² Source : Agence Internationale de l'énergie, World energy outlook 2006 <http://www.iea.org>

1.2.2 Analyse du cycle de vie

L'analyse du cycle de vie des produits : « du berceau à la tombe »

Analyser le cycle de vie, c'est évaluer l'impact total d'un produit sur l'environnement. Des méthodes élaborées dans les années 90 appelées « Analyses de Cycle de Vie »²³ (ACV) donnent un état environnemental global du produit « from cradle to grave », se traduisant littéralement « du berceau jusqu'à la tombe » : extraction des matières premières, fabrication et transport, utilisation, mise en décharge ou recyclage.

Chaque produit peut avoir ainsi un impact plus ou moins important sur l'effet de serre, l'acidification de l'atmosphère, l'épuisement des ressources naturelles, la dégradation des cours d'eau (eutrophisation). L'ACV et son écobilan nous révèlent les produits plus ou moins « indolores » à l'environnement.

L'exemple suivant de l'impact sur l'environnement d'un ordinateur²⁴ est significatif :

- construction de l'unité centrale : équivalent énergétique de 373 litres de pétrole ;
- la fabrication nécessite 2800 Kg de matières premières ;
- déchet de fabrication : 140 Kg plus 24 Kg de déchets toxiques ;
- rejet de CO₂ pour tout son cycle d'utilisation : 650 Kg ;
- acidification d'atmosphère : 5Kg d'équivalent SO₂ ;
- et la consommation de l'électricité pendant tout le cycle d'utilisation.

En 2003, Eric Williams, professeur à l'Université des Nations Unies (UNU), coauteur de « Computers and the Environment : Understanding and Managing Their Impacts »²⁵ estime que la construction d'un ordinateur avec écran 17 pouces (tube cathodique) a nécessité 260 Kg d'énergie fossile (11 fois le poids de l'équipement), 22 Kg de produits chimiques et 1,5 tonne d'eau et affirme aussi que la baisse de l'électricité consommée par les processeurs est malheureusement compensée par leur multiplication récente dans l'ordinateur.

→ *Les produits technologiques polluent dès leur naissance, les progrès rapides font diminuer leur durée de vie et leurs déchets toxiques ne cessent donc de croître.*

1.3 Les acteurs

Les systèmes de stockages voient leur rôle s'étendre, leurs technologies se spécialiser pour répondre aux nouveaux enjeux des systèmes d'informations. Tous les acteurs sont concernés face aux impacts environnementaux des T.I.C. : des particuliers aux entreprises, des constructeurs aux sociétés de services, des associations jusqu'aux institutions.

1.3.1 Typologie des systèmes de stockage des données

Les différentes technologies de systèmes de stockage concernées par le Green Storage

Les baies de stockages et leurs disques durs représentent une part importante de la consommation énergétique dans le data center. Mais il convient d'élargir le Green Storage à d'autres infrastructures hébergeant ou manipulant les données.

²³ Article Sciences et Vie « Construire un monde durable » hors série Juin 2008

²⁴ Article Sciences et Vie « Construire un monde durable » hors série Juin 2008

²⁵ http://update.unu.edu/archive/issue31_5.htm

Afin d'être plus complets sur le périmètre des systèmes de stockage concernés, les infrastructures suivantes stockent, sauvegardent, archivent, exploitent des données dans le data center :

- les systèmes de stockages « sécurisés », dédiés à la production et aux activités de l'entreprise jugées comme critiques ;
- les systèmes de stockage « massifs », très capacitifs ;
- les systèmes de sauvegarde et d'archivage sur disques durs, sur bandes magnétiques ou disques magnéto-optiques ;
- tous les réseaux et interfaces dédiés aux échanges de données.

1.3.2 L'évolution vers des valeurs plus « vertes » des principaux acteurs

Le marché domestique

Les particuliers voient augmenter fortement leur consommation électrique liée aux technologies numériques et ils cherchent les moyens de la réduire. Néanmoins, ils souhaitent montrer aussi un comportement plus vertueux et respectueux de l'environnement, et ont par exemple pour la plupart déjà privilégié l'achat d'équipements électroménagers « économes » de classe « A » ou « B ».

Mais à part pour de nouvelles générations de micro-ordinateur, les rayons informatiques ne mentionnent que très rarement les performances énergétiques des composants de stockage et le consommateur peut rarement en faire un critère de choix « vert ».

Le marché des entreprises

Les T.I.C. devenant l'épine dorsale de leurs activités, les entreprises connaissent aussi une consommation énergétique sans cesse croissante, dont les coûts et l'approvisionnement deviennent même parfois prohibitifs (selon les pays). Les data centers sont souvent âgés et saturés par le développement continu des T.I.C.

Plus sensibilisées aux enjeux environnementaux et sociétaux, les entreprises doivent répondre aux nouvelles exigences de développement durable demandées par les institutions, les investisseurs, leurs employés et partenaires. Elles ont conscience que les « pressions vertes » deviennent un enjeu de réputation et d'image d'une entreprise plus citoyenne et responsable.

Les fournisseurs des T.I.C.

Comme toute entreprise, ils sont bien entendu confrontés aux mêmes problématiques et enjeux « verts ». Les constructeurs, intégrateurs et sociétés de services tentent pour la plupart de capter ces nouveaux marchés et de répondre aux nouvelles attentes afin de se distinguer de la concurrence.

Les acteurs politiques et réglementaires

Les accords mondiaux environnementaux semblent extrêmement complexes à réaliser. En Europe et en France, les institutions ont commencé depuis peu à encourager les initiatives vertes dans les T.I.C.. Parallèlement, elles poussent les autres secteurs à se servir des nouvelles technologies pour réduire leur impact et améliorer leur efficacité.

Les programmes, normes et standards (RoHS²⁶, WEEE, ISO, EMAS, Code Of Conduct...) se développent pour apporter des certifications valorisantes et distinctives.

²⁶ Restriction of the use of certain Hazardous Substances in electrical and electronic equipment

Les associations sans but lucratif et les clubs utilisateurs

De récentes associations « vertes » spécialistes des T.I.C. comme le Green Grid, Climate Saver ou une experte du stockage comme la SNIA, contribuent à sensibiliser les différents acteurs sur les problématiques « vertes » et à faire des recommandations pratiques.

Les associations d'utilisateurs et de DSI comme le CIGREF, le CRIP ou le CESIT rassemblent leurs compétences et forment toutes des groupes de travail « Green IT ».

Le tableau ci-dessous publié par Gartner²⁷ en 2007, met en évidence les changements qui s'opèrent dans les attentes des différents acteurs :

	The present		The future
Green regulations	Voluntary	→	Mandatory
Green consumers	Minority	→	Majority
Investor focus	Growing	→	Intense
Executive view	Leading-edge	→	Mainstream
Business initiatives	Niche projects	→	Core to business
Carbon tax	None	→	Multiple
Green IT	Interesting	→	Required

Figure 3. Gartner "A profound Economic Change Is Gathering Pace"

Une pression verte, naissante en 2007, est amenée rapidement et dans le futur proche à s'affirmer et se développer de différentes manières selon le schéma :

- les efforts écologiques d'abord fondés sur le volontariat sont de plus en plus « exigés » par l'état et deviennent progressivement « obligatoires » ;
- les consommateurs « verts » qui étaient une minorité se développent, et seront bientôt une majorité. Ils auront donc un pouvoir important sur les marchés ;
- le critère « vert », naissant pour les investisseurs sera bientôt exigé et pressant ;
- les directions tiendront de plus en plus compte du critère vert ;
- les marchés « verts », souvent des marchés de niche aujourd'hui, verront leurs parts devenir importantes voire dominantes dans de nombreux secteurs ;
- les taxes et impôts liés à l'environnement apparaissent et se développeront sous de multiples formes.

➔ *Des technologies de l'information plus « vertes » sont non seulement considérées et appréciées, mais deviennent exigées, voire obligatoires, pour répondre aux demandes croissantes des différents acteurs, des marchés et de l'urgence environnementale.*

²⁷ Forrester Research "Topic overview : Green IT" par Christopher Mines, Euan Davis
<http://www.forrester.com/Research/Document/0,7211,43494,00.html>

1.4 Réglementations « vertes » des T.I.C.

L'Europe rattrape son retard, notamment par de nouveaux labels, normes et code de conduite en plein développement. L'Europe applique de nouvelles directives plus contraignantes lors de la fabrication et le traitement des déchets et s'évertue à mettre en œuvre des projets d'efficiences énergétiques.

1.4.1 Réglementations environnementales



Directive européenne RoHS : restriction de l'utilisation de certaines substances dangereuses dans les équipements électriques et électroniques

Le 27 janvier 2003, le parlement européen a adopté la directive 2002/95/CE, afin de limiter l'emploi de substances dangereuses dans les équipements électroniques et électriques.

Cette directive abrégée RoHS dénombre 6 substances dont l'emploi ne devra pas dépasser des seuils (0,1 % par unité de poids de matériau homogène sauf pour le cadmium : 0,01 %) : le plomb (Pb), le mercure (Hg), le cadmium (Cd), le chrome hexavalent (Cr+6), les Polybromodiphényléthers (PBDE), les polybromobiphényles (PBB)

Directive européenne WEEE/RAEE

Parallèlement, le parlement Européen a aussi adopté la directive 2002/96/CE ou WEEE²⁸/RAEE, concernant les déchets d'appareils électriques et électroniques, leur collecte, traitement et recyclage. L'objectif est de responsabiliser le producteur, et d'offrir une seconde vie à certains matériels, par recyclage, afin de réduire encore l'empreinte écologique.

Le programme européen REACH²⁹

Ce nouveau règlement rentré en vigueur en juin 2007, renforce l'enregistrement, l'évaluation, l'autorisation et les restrictions des produits chimiques. Fin 2008, près de 4 400 entreprises françaises sur environ 5 500 ont préenregistré leurs substances. Le non-enregistrement ou le non-respect des mesures d'interdiction pourront être sanctionnés pénalement de 2 ans d'emprisonnement et de 75 000 euros d'amende.

Les écolabels « globaux »

Blue Angel³⁰, créée en 1977 par le ministère de l'écologie allemand, est la plus ancienne certification dans le monde applicable au matériel informatique. Ses critères sont nombreux : de la conception au recyclage des produits, diminution des pollutions, réduction des consommations énergétiques, des émissions chimiques... Plus de 20 pays l'utilisent ainsi que des entreprises comme Dell, Fujitsu, HP, Samsung...

L'EPEAT³¹, issue du Green Electronic Council américain (GEC) en 1992, signifie « outil d'évaluation environnementale des produits électroniques ». C'est la mise en œuvre du

²⁸ WEEE : Waste Electronic and Electrical Equipment ou DEEE : déchet d'équipement électrique ou électronique

²⁹ REACH : Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals

³⁰ <http://www.blauer-engel.de/en/index.php>

³¹ <http://www.epeat.net>

standard IEEE 1680-2006 (standard for Environmental Assessment of Personal Computer products) couvrant la conception, la fabrication et l'utilisation. Seuls les États-Unis et l'Europe utilisent cet écolabel, mais de nombreuses entreprises l'ont adopté comme Apple, Dell, Fujitsu, HP, Sony, Toshiba...

TCO, à l'origine conçu en Suède, pour reconnaître les écrans écologiquement responsables, cet écolabel s'est étendu aux PC et serveurs. Son champ est large : émission de champs électromagnétiques, consommation d'énergie, respect de la directive RoHS, recyclabilité. De nombreux pays et entreprises ont adopté cet écolabel, depuis sa création en 1980.



L'Ecolabel Européen : couvrant la conception, la fabrication, l'utilisation et le recyclage, ce label naissant souhaite contrôler la réduction et la consommation d'énergie, l'utilisation de mode veille, la limitation d'usage de substances dangereuses. Pour le moment, il est trop tôt pour juger de l'adoption et du développement de ce label.

Les écolabels incitant à la réduction de consommation d'énergie



Energy Star³² : créé en 1992 par l'agence américaine pour la protection de l'environnement (EPA), ce label a pour objectif de mesurer l'efficacité énergétique des matériels informatiques. En 2001, la Commission européenne a signé un accord avec les États-Unis pour aider son adoption en Europe. Ce label est utilisé dans le monde entier par de nombreuses entreprises

80plus : programme destiné à améliorer l'efficacité énergétique des alimentations électriques des systèmes informatiques. Sa certification exige au moins 80 % d'efficacité pour une alimentation électrique.

D'autres programmes environnementaux adaptés à l'informatique sont en développement actuellement : « Choix environnemental » (Canada), « Green Guard » (USA, orienté CO₂), « PC Green Label » (Japon), « CECPC » (Chine), « Ecologo »³³ (Canada).

→ En Europe, c'est en 2010, que la directive EuP (European Energy using Product) devrait prendre en compte au-delà de l'efficacité énergétique, l'ensemble du cycle de vie du produit et recouvrir les directives DEEE et RoHS.

1.4.2 Les normes « vertes »

La norme ISO 14001 : Environnement et Systèmes de management environnemental.

Réalisée par l'ISO³⁴, son objectif est de réduire au minimum l'impact des activités sur l'environnement. Cela ne signifie pas que l'entreprise ne pollue plus, mais qu'elle met en œuvre des moyens pour respecter la réglementation existante, faire preuve d'amélioration continue et s'engager à prévenir la pollution.

Si l'entreprise certifiée ISO 14001 publie une déclaration environnementale conforme à l'EMAS³⁵, elle peut bénéficier alors de cette « écocertification » créée en 1995 par la

³² www.energystar.com

³³ <http://www.ecologo.org>

³⁴ ISO : International Organisation for Standardization

³⁵ EMAS : Eco Management and Audit Scheme

commission européenne. L'EMAS³⁶ prend en compte les impacts environnementaux des achats de l'entreprise, du travail des sous-traitants et fournisseurs, des transports et de l'écobilan des produits ou services.

Les normes orientées « développement durable » et complémentaire aux normes environnementales sont ISO 26000 -responsabilité sociale des organisations- et DS21000 -responsabilité sociétale et développement durable.

1.4.3 Le programme « Code of Conduct » européen

Un code de bonne conduite des data centers a été lancé en octobre 2008 par la commission européenne et doit aussi permettre d'atteindre l'objectif de réduction des émissions de CO₂ de 20 % d'ici 2020. Un guide des meilleures pratiques expose les critères et recommandations de choix des équipements, de climatisation, des architectures serveurs-stockages, de gestion des données, etc.

Basé sur l'initiative volontaire, le C.O.D. a pour objectif de sensibiliser les opérateurs et propriétaires (« participants ») à l'impact environnemental de leur centre de données, mais aussi toutes les autres parties prenantes -en soutien- (« endorsers ») comme les constructeurs, les sociétés de conseil, les fournisseurs d'énergies, les standards et associations normatives, pour les inviter à développer, améliorer et intégrer les meilleures pratiques possible.

Les objectifs poursuivis par ce programme sont nombreux et parmi les plus attendus :

- le développement et la promotion de métriques pour évaluer et mesurer l'efficacité énergétique ;
- la mise à disposition d'un guide clair et pédagogique sur les meilleures pratiques ;
- un forum ouvert de discussion en Europe de toutes les parties prenantes ;
- une coordination des pratiques avec celles des autres initiatives internationales ;
- accroître une prise de conscience des dirigeants, des fournisseurs et des investisseurs de l'impact positif de bonnes pratiques sur le centre de données, sur la dépense énergétique et surtout sur l'environnement, pour un développement durable.

1.4.4 Développement durable et « Corporate Social Responsibility »

La plupart des entreprises françaises intègrent leur stratégie Green IT dans un projet global de Responsabilité Sociétale de l'Entreprise (RSE)³⁷ consistant principalement à montrer qu'elles agissent de façon éthique et en contribuant au développement économique durable.

La mondialisation et une concurrence de plus en plus forte imposent de devoir se distinguer dans son offre et ses activités par des critères moraux plus respectueux de l'homme et de son environnement.

Le degré d'alignement de l'entreprise avec les valeurs de ses employés et clients à sa stratégie corporate est un critère concurrentiel grandissant. Dès 2005, KPMG³⁸ publie une enquête internationale consacrée à la « Corporate Social Responsibility »³⁹ ou

³⁶ http://ec.europa.eu/environment/emas/index_en.htm

³⁷ RSE : Responsabilité Sociétale de l'Entreprise CSR : Corporate Social Responsibility

³⁸ <http://kpmg.fr>

³⁹ « International Survey of Corporate Responsibility Reporting 2005 », http://www.kpmg.com.au/Portals/0/KPMG_Survey_2005_3.pdf

Responsabilité Sociétale d'Entreprise, dont voici quelques critères essentiels s'appliquant à la stratégie d'entreprise :

- réputation, transparence, engagement sur le développement durable ;
- impact social : respect des lois et de l'individu, profitabilité et société civile, équité sociale ;
- achat, approvisionnement et logistique éthiques ;
- respect de l'environnement.

Ainsi, l'entreprise est jugée globalement sur ses activités et leur moralité : elle doit en conséquence veiller et encourager au mieux, tous les dispositifs humains et matériels qui pourraient avoir un impact positif sur son rapport annuel « Environmental Social Report ».

Les systèmes d'informations sont donc au cœur de ces nouveaux enjeux, car ils sont à la fois l'épine dorsale de l'activité de l'entreprise et ont des impacts plus ou moins forts sur l'environnement. Ils sont évidemment aussi attendus comme un levier de performance et d'optimisation pour toutes les activités (ex. : logistique, communication, marketing,...).

➔ *Les T.I.C., incluant les systèmes de stockage, sont au cœur d'enjeux essentiels et de responsabilités sociétales critiques, car elles détiennent les données vitales de l'entreprise, des partenaires, des clients et des employés. Aussi, leur développement exponentiel les expose à être pris logiquement pour responsable d'une importante part de l'impact environnemental du data center.*

1.4.5 Le développement du Green IT en Europe et en France

En Europe, un slogan : « 20-20-20 »

En mars 2007, la commission européenne a adopté une série de mesures ayant pour objectif de réduire de 20 % les émissions de gaz à effet de serre (sur la base de 1990) et de passer à 20 % la part des énergies renouvelables d'ici 2020.

Grenelle de l'environnement

En septembre 2008, le gouvernement a décidé de repousser toute extension du concept de bonus-malus aux T.I.C. afin qu'il n'y ait aucun nouvel impact négatif sur la consommation des ménages.

France Numérique 2012

Le plan de développement de l'économie numérique française a été publié en octobre 2008⁴⁰, et aborde le domaine Green IT.

Sur les 154 actions proposées, seuls deux groupes sont liés à « l'informatique durable » : l'incitation au télétravail et le développement du numérique comme levier de la transformation.

Ainsi, les actions 113 à 117 favorisent et stimulent le télétravail. Le chapitre 3.13 « Recourir au numérique pour accélérer la mutation environnementale de la société » comprend les actions 134 et 135 traitant respectivement :

- Action 134 « Améliorer les matériels et les processus de production » : un programme de recherche sur l'efficacité énergétique dans les centres de données en fait partie (...) prenant

⁴⁰ <http://francenumerique2012.fr>

des engagements de réduction de consommation d'énergie et de développement d'éco-conception et de recyclage des produits et services.

- Action 135 : création de prix « Green » : « deux distinctions spécifiques, l'une pour le secteur producteur de TIC, l'autre pour le secteur utilisateur permettront de valoriser les entreprises utilisant les TIC pour réduire la consommation d'énergie et les émissions de CO₂ (...) »

Le plan de développement reconnaît que les T.I.C. « constituent d'abord un gisement considérable de réduction de nos émissions de CO₂. Grâce aux T.I.C., nous disposons des moyens techniques nécessaires pour construire un autre modèle de croissance, plus durable. (...) Il nous faut donc anticiper l'ère de la rareté et penser dès aujourd'hui à une nouvelle génération d'appareils, plus autonomes et moins gourmands en énergie. L'enjeu est également comportemental, des gestes simples doivent devenir de vrais réflexes notamment dans les entreprises. Enfin, le Grenelle de l'Environnement a montré que nous avons encore d'importants efforts à fournir en matière de recyclage des déchets électroniques et des métaux lourds contenus dans ces appareils. »

→ En 2007 et 2008, des initiatives européennes et françaises se concrétisent, 2009 devrait apporter au mois de mai les conclusions de la nouvelle commission française Green IT. Elle aura pour objectif de développer les technologies de l'information « vertes » ainsi que leurs capacités à être un levier d'économies énergétiques pour d'autres secteurs.

1.5 Définitions du « Green IT » et du « Green Storage »

Même si l'informatique a toujours eu pour vocation d'améliorer notre productivité et la performance de nos activités, on a surtout exigé d'elle qu'elle le soit en fonctionnant le plus correctement possible. Aujourd'hui sa responsabilité est plus forte, de nouvelles composantes « vertes » apparaissent et se développent. Elles ont des caractéristiques et des objectifs distinctifs, y compris pour les systèmes de stockage de données.

1.5.1 « Green IT »

Origine du Green IT

On peut rattacher l'origine des premiers composants informatiques plus « verts », plus efficaces énergétiquement, et récompensés par un label, au programme américain EnergyStar créé en 1992, chargé de promouvoir les économies d'énergie.

Motivations principales stimulant le Green IT

IDC, interrogeant 459 DSI Européennes en 2008, a mis en évidence dans son enquête que le développement durable, la protection de l'environnement et la responsabilité sociétale des entreprises (RSE) sont désormais au cœur des stratégies de développement.

Les entreprises intègrent le développement durable et une attitude plus « verte » dans leurs activités en adoptant en premier lieu ces initiatives :

- programmes de recyclage ;
- audit énergétique des bâtiments et des salles informatiques ;
- encouragements au télétravail, à l'utilisation d'outils de téléconférences et d'applications « collaboratives » réduisant le besoin de réunion et de transport ;
- investissements dans les technologies de virtualisation de serveurs.

Plusieurs problématiques fortes sont en faveur du développement du Green IT :

- la hausse générale de l'énergie : les sources permettant la fabrication de l'électricité sont de plus en plus rares et coûteuses, certains data centers ont des difficultés d'approvisionnements énergétiques ;
- la densité des composants (serveur à lame, stockage massif...) entraîne aussi une forte hausse des émissions de chaleur et donc des besoins en climatisation ;
- les mesures (lois, directives) prises pour contrôler les émissions de CO₂ et développer le recyclage des produits ;
- se différencier des offres de celles des pays émergents très concurrentiels sur le rapport qualité/prix par des produits plus respectueux de l'environnement dans leur fabrication, leur utilisation, et leur recyclage ;
- la progression des réglementations « vertes » vers les data centers qui deviennent donc très souvent non conformes à la directive RoHS, par exemple ;
- la forte augmentation des coûts de l'immobilier et des exigences de continuité de service plus contraignantes, rendent complexe et difficile la création ou l'extension du data center ;
- l'infrastructure électrique du data center souvent obsolète et/ou utilisée à son maximum due à l'augmentation de la densité des équipements ;
- le besoin général de mieux gérer les technologies du data center pour apporter plus de disponibilité, de productivité et pouvoir faire baisser les coûts opérationnels.

➔ *Une stratégie Green IT peut donc consister pour l'entreprise à avoir une stratégie de développement durable en réduisant l'impact sur l'environnement du data center par la transformation et l'utilisation plus efficace de ses ressources, avec l'objectif de respecter ou d'améliorer les performances de toutes les activités.*

Selon IDC⁴¹, l'informatique verte est composée de nombreuses facettes pouvant :

- être un composant dans un produit, un produit complet ou alors la solution globale et comprendre : du matériel, du logiciel et des prestations de services ;
- impliquer toutes les technologies de l'information : PC, serveurs, stockages ainsi que les équipements de communications et réseaux ;
- concerner les stratégies d'entreprise, la responsabilité sociétale, les réglementations, les meilleures pratiques associées ainsi que les initiatives de bon sens ;
- viser les émissions de CO₂ et toutes pollutions en général, la consommation énergétique, le refroidissement et la disponibilité de l'énergie ;
- être liées à l'utilisation, la performance et le niveau de disponibilité des infrastructures ;
- concerner tous les acteurs de la chaîne IT : assembleurs-intégrateurs, constructeurs, SSII, acheteurs, utilisateurs, ect.

➔ *Une définition large du Green IT, se traduisant par informatique « verte » ou « éco-responsable », peut consister en un produit, une solution, un service ou une pratique, pouvant améliorer l'efficacité énergétique, diminuer l'impact sur l'environnement, et contribuer au développement durable.*

⁴¹ Conférence IDC, Karim Bahloul, le 23/10/08 "L'informatique verte : phénomène de mode ou tendance de fond ?"

1.5.2 « Green Storage »

L'évolution de la consommation énergétique des systèmes de stockage

Dès août 2007, l'agence de protection de l'environnement américaine (EPA), a émis au congrès un rapport comparant 4 composants principaux du data center et leur évolution en consommation énergétique. Cette étude démontre que les systèmes de stockage ont le taux de croissance en consommation énergétique le plus fort : 191 %, et la plus haute consommation : 3,2 milliards de kWh.

Ci-dessous, l'extrait du rapport de l'EPA :

Consommation électrique (milliard kWh)

	2000	% total	2006	% total	Croissance
Systèmes de stockage	1,1	18%	3,2	32,30%	191%
Infras Réseaux	1,4	23%	3	30,30%	114%
Serveurs midrange	2,5	41%	2,2	22,20%	-12%
Serveurs Highrange	1,1	18%	1,5	15,20%	36%
Total	6,1	100%	9,9	100%	62%

Les infrastructures de stockage ont aussi de nombreux progrès à faire : pour le respect de l'environnement, par une plus grande efficacité énergétique, la réduction de leur empreinte carbone et le retraitement de leur déchet ; dans l'optimisation de leur productivité ; dans la rationalisation de leur usage et des comportements liés.

Voici quelques exemples significatifs rapportés des conférences EMC forum et IDC Green IT Day données fin 2008 :

- plus de la moitié des données est sur un espace de stockage inapproprié à leur valeur ;
- le taux d'utilisation des systèmes de stockage se situe entre 20 et 40 % ;
- le manque de classification des données fait exploser les coûts de recherche de l'information et contribue à rendre les données persistantes.

Comme pour le Green IT, le Green Storage peut consister en un composant de stockage, une solution, un service ou une pratique, pouvant améliorer l'efficacité énergétique, diminuer l'impact sur l'environnement, et contribuer au développement durable.

➔ Une définition plus large du Green Storage, ajoute à l'amélioration de l'efficacité énergétique des infrastructures de stockage : des technologies et mécanismes d'optimisations de leurs capacités ; des fonctionnalités et pratiques pour rationaliser leurs usages ; des dispositifs de gouvernance des données, pour agir en amont sur leur forte croissance.

IDENTIFICATION DES PROBLEMATIQUES ET DES FACTEURS CLES

Depuis 30 ans, les systèmes d'informations n'ont cessé d'évoluer et de s'épandre. En 2009, de la « salle machine » de quelques serveurs, au « data center » de 46 000 M² de Las Vegas ou de 67 000 M² en Chine, de nouvelles problématiques cruciales apparaissent :

- manque d'espace, très forte densité énergétique, climatisation inadaptée ou saturée ;
- hausse de prix : M², électricité (selon pays), refroidissement ;
- sous utilisation des systèmes, inadaptation de la charge à l'activité ;
- systèmes vieillissants : composants toxiques, recyclage et traitement des déchets ;
- contraintes légales : renforcement des réglementations et normes environnementales.

Et notamment, s'ajoutent pour les systèmes de stockages, ces problématiques :

- multiplicité et spécialisation des systèmes, pour toujours plus de disponibilité, de performance et de protection des données ;
- croissance constante et non maîtrisée des données ;
- surévaluation des solutions proposées aux directions métiers ;
- surallocation de moyens lié aux pressions des nouvelles réglementations.

“Les mêmes causes qui nous procurent des biens peuvent également nous causer des maux, tout en nous offrant le moyen de les éviter.”

Démocrite, extrait de [Fragments]

2.1 L'évolution des Systèmes d'Informations et du data center

Pour comprendre l'importance croissante des systèmes de stockage dans les systèmes d'information, un bref historique de l'évolution de leur architecture au sein du data center est nécessaire. L'illustration ci-dessous résume les 3 principales étapes de cette évolution :

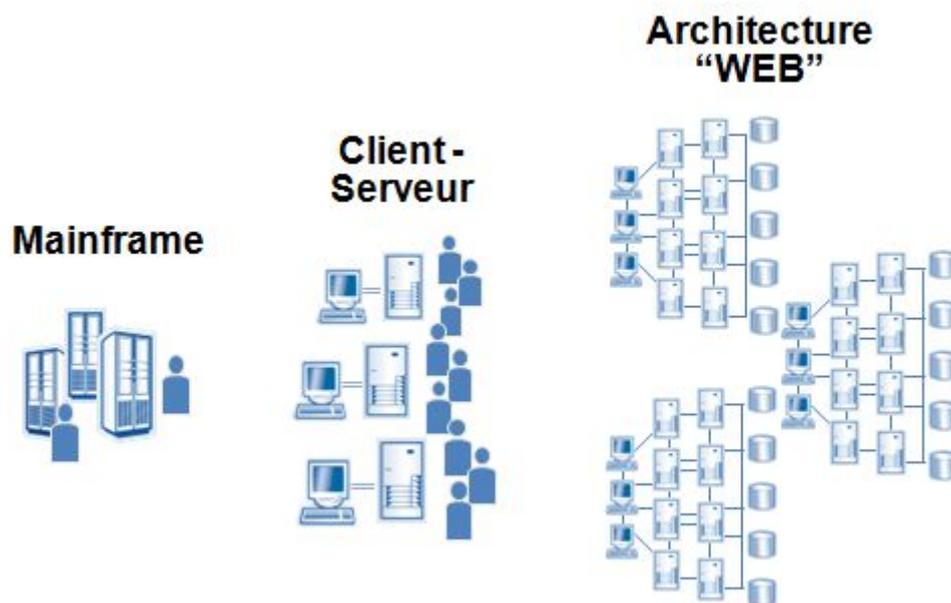


Figure 4. Evolution du data center

En passant de l'architecture de l'ère « Mainframe » à celle du « Client-Serveur », les serveurs commencent à se développer dans le data center. La troisième architecture est liée au développement du « WEB » et intègre beaucoup plus de serveurs, d'applications, de ressources de stockage et d'utilisateurs connectés.

→ Le nombre des serveurs et des applications prolifère ainsi que les systèmes de stockages de données associés.

2.2 L'évolution des systèmes de stockages de données

Afin d'appréhender l'importance croissante des systèmes de stockage dans les systèmes d'information, un historique de leur architecture est aussi nécessaire. De profondes évolutions se sont opérées dans l'architecture des systèmes de stockage ces dix dernières années.

Ainsi **Hacene Latreche**, Responsable Marketing Green IT chez Bull, me faisait part lors d'un entretien accordé en juillet 2008, des principales étapes de l'évolution des architectures serveurs-stockages.

Début des années 90 : chaque serveur possède son stockage, avec ses propres disques durs internes. Les serveurs sont répartis et indépendants.

De Mi-90 à 2000 : les premiers systèmes de stockages indépendants des serveurs et rattachés directement à ces derniers se développent, c'est la technologie DAS⁴² : Direct Attached Storage.

Début des années 2000 : progressivement, les systèmes de stockage ont leur propre autonomie et sont administrés indépendamment des serveurs. Ils peuvent avoir un réseau dédié pour améliorer, isoler et sécuriser leurs flux d'échanges de données avec les serveurs. Ainsi se développent les technologies SAN⁴³ (réseau fibre channel indépendant) et NAS⁴⁴ (pouvant être basé sur le réseau Ethernet existant ou sur un réseau Ethernet dédié et indépendant) dans le data center.

Dès 2004 à aujourd'hui : les baies de stockage améliorent significativement leur performance, leur capacité et leurs dispositifs de continuité de service. Certains fonctionnent en « haute disponibilité » avec toujours plus d'utilisateurs connectés, 24h/24 et 7j/7. Ils intègrent de nouvelles fonctionnalités comme la virtualisation, la surallocation (Thin Provisioning), les snapshots, la déduplication et sont interconnectés à de multiples réseaux de stockage SAN, NAS, I-SCSI de plus en plus rapides : 8 et 10 Gbits/s.

→ *L'évolution actuelle met en évidence le fort développement des ressources de stockage en systèmes autonomes, fortement concentrés (même tendance que les serveurs) et sécurisés, utilisant des réseaux dédiés pour assurer des échanges rapides et sécurisés avec les serveurs et les utilisateurs.*

L'évolution de l'usage des Systèmes d'Informations

On constate une intensivité d'utilisation des systèmes d'information qui fonctionnent souvent 24 heures sur 24 et 7 jours sur 7 car les équipements liés aux communications et technologies internet doivent être accessibles continuellement. Ils supportent quasiment toutes les activités de l'entreprise et deviennent une véritable épine dorsale pour l'entreprise.

Le développement rapide de la numérisation, de la dématérialisation des processus et de la simulation, accroissent les dispositifs de stockage de données et donc de sauvegardes et d'archivages nécessaires à leurs traitements. Des technologies comme le Web 2, la vidéo Haute Définition, le RFID, sont quelques exemples du développement exponentiel des données numériques dans toute l'entreprise.

→ Le développement soutenu des systèmes traitant les données et la multiplicité de leur rôle (stockage, sauvegarde, archivage) font augmenter les besoins en consommation énergétique, en refroidissement et en espace dans le data center.

⁴² DAS : Direct Attached Storage

⁴³ SAN : Storage Area Network

⁴⁴ NAS : Network Attached Storage

2.2.1 Le m² du data center toujours plus cher et consommateur d'énergie

Construire un data center peut coûter près de 20 000 euros par m² à sa construction, et jusqu'à 40 000 euros en exploitation selon ses fonctionnalités. Cela est dû principalement à l'accroissement de la densité énergétique, la puissance d'un rack au m² peut atteindre 8KW, Et aux exigences de fonctionnement du data center et de sa capacité à répondre aux besoins de haute disponibilité, de continuité d'activité, jusqu'à 24h/24 et 7J/7.

Un classement de la capacité à assumer une plus ou moins importante continuité de service -par Tier de 1 à 4- a été réalisé par un spécialiste des data centers, l'Uptime Institute⁴⁵, qui utilise des critères comme la redondance des équipements de secours électrique et leur autonomie (voir en annexe le tableau synthétique). Depuis 1993, l'Uptime Institute publie des études et livres blancs sur l'optimisation et l'amélioration de la disponibilité du data center.

La forte augmentation de la densité énergétique dans le data center est illustrée par le graphique suivant d'IDC - présenté le 23 octobre 2008 lors d'une conférence dédiée au Green IT à Paris - :

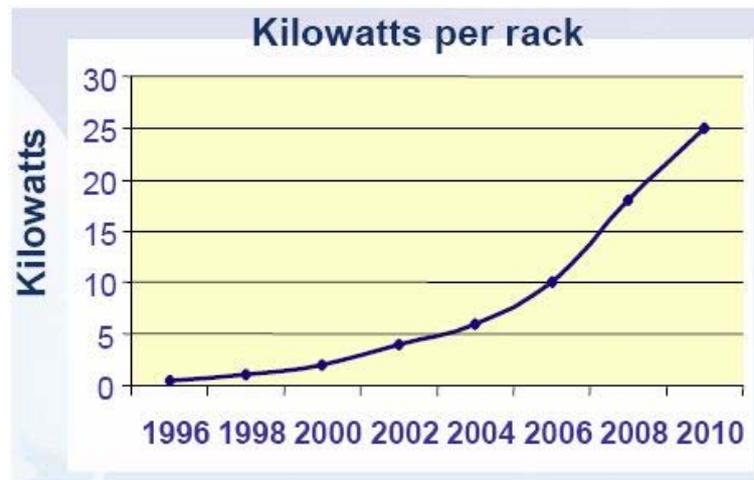


Figure 5. Evolution énergétique moyenne par rack, source IDC

De 1 000 watts au m² en 2000, puis 6 à 8 000 en 2006, les prévisions dépasseraient les 25 000 watts au m² en 2010 : cette évolution et accélération du besoin en électricité est autant liée au développement des serveurs « lames » (blades) associant puissance et compacité qu'à celui des systèmes de stockage beaucoup plus performants et capacitifs.

Une étude faite par le groupement des utilisateurs de data centers (Data Center User Group-DCUG-) en 2006 et conduit par la société Emerson⁴⁶, affirme que 96 % des data centers auront atteint leur consommation énergétique possible maximum d'ici 2011.

On estime en général à 1 Watt le besoin en climatisation et refroidissement pour faire fonctionner 1 Watt en équivalent d'équipement informatique. Ainsi, 50 % de l'électricité du data center sert à refroidir et à ventiler les systèmes d'information.

⁴⁵ <http://uptimeinstitute.org> : livre blanc « Tier Classifications Define Site Infrastructure Performance »

⁴⁶ <http://www.liebert.com> (Emerson Network Power)

IDC estime à 200 000 euros la facture mensuelle d'électricité d'un data center de 2 000 m² et ajoute que d'ici 2015, les coûts opérationnels et énergétiques seront supérieurs aux coûts d'acquisitions.

→ Dans le dossier très complet publié par Matt Stansberry : « *Energy efficient computing in the 21st Century, The Green Data center* », Nicolas Carr affirme que lorsque le public aura compris comment la consommation d'énergie impacte gravement l'environnement et que les data centers gaspillent quotidiennement l'énergie et sont responsables d'une grande part des émissions polluantes, on peut s'attendre alors à ce que Greenpeace prenne aussi d'assaut les salles informatiques !

2.2.2 Évolution de la consommation énergétique des systèmes de stockage

Selon une étude publiée en 2007 par l'agence de protection de l'environnement américaine (EPA), les systèmes de stockage ont le taux de croissance en consommation énergétique le plus élevé : 191 % et une consommation de l'ordre de 3,2 milliards de kWh.

Évolution de la densité des systèmes de stockage

Comme pour les serveurs, la densité des systèmes de stockage évolue significativement et l'intégration de leurs composants est de plus en plus forte, notamment les supports de stockage : disques durs, mémoires, lecteurs et bandes magnétiques...

Des systèmes de stockage, à base de disque dur S-ATA hautement capacitif, peuvent concentrer dans un seul cabinet plusieurs centaines de Tera-octets de données.

Même les bibliothèques de bande magnétique connaissent une augmentation importante de leur nombre de lecteurs, de slots et de capacité des cartouches.

2.2.3 Analyser la consommation énergétique du data center

Continuant à approfondir la problématique énergétique avec **Ahcene Latreche**, Responsable Marketing de l'offre Green IT de Bull, ce dernier partage son expérience passée de l'exploitation du data center et souligne les difficultés à comprendre comment est consommé son énergie :

- Combien d'énergie rentre au total dans le data center ? Comment est-elle consommée ?
- Quelle part de l'énergie est consacrée aux systèmes de climatisation, aux systèmes d'alarmes et incendies, aux générateurs, aux onduleurs ? Sans oublier l'éclairage, la surveillance vidéo, les dispositifs d'identification et de sécurité ?
- Quelle part restante est finalement consommée par les systèmes informatiques ?

Différentes formules permettent d'évaluer les valeurs de ses parts

Le ratio « énergie totale utilisée par le data center / énergie consommée par les systèmes informatiques du data center » est le PUE⁴⁷ : « Power Usage Effectiveness ».

⁴⁷ PUE : Power Usage Effectiveness

$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}} = \frac{\text{Total énergie utilisée par le data center}}{\text{Total Energ.consommée par les systèmes inf.}}$$

Et réciproquement :

$$DCE^{46} = \frac{1}{PUE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}} = \frac{\text{Total Energ.consommée par les systèmes inf.}}{\text{Total énergie utilisée par le data center}} \times 100\%$$

Des formules de calcul⁴⁸ pour mesurer l'efficacité de la consommation énergétique sont proposées par l'association « The Green Grid »⁴⁹. Fondée en février 2007, elle a pour objectif de réduire l'impact de l'informatique sur l'environnement et compte parmi ses membres certains des plus grands constructeurs et acteurs informatiques.



Le schéma suivant illustre les formules du Green Grid :

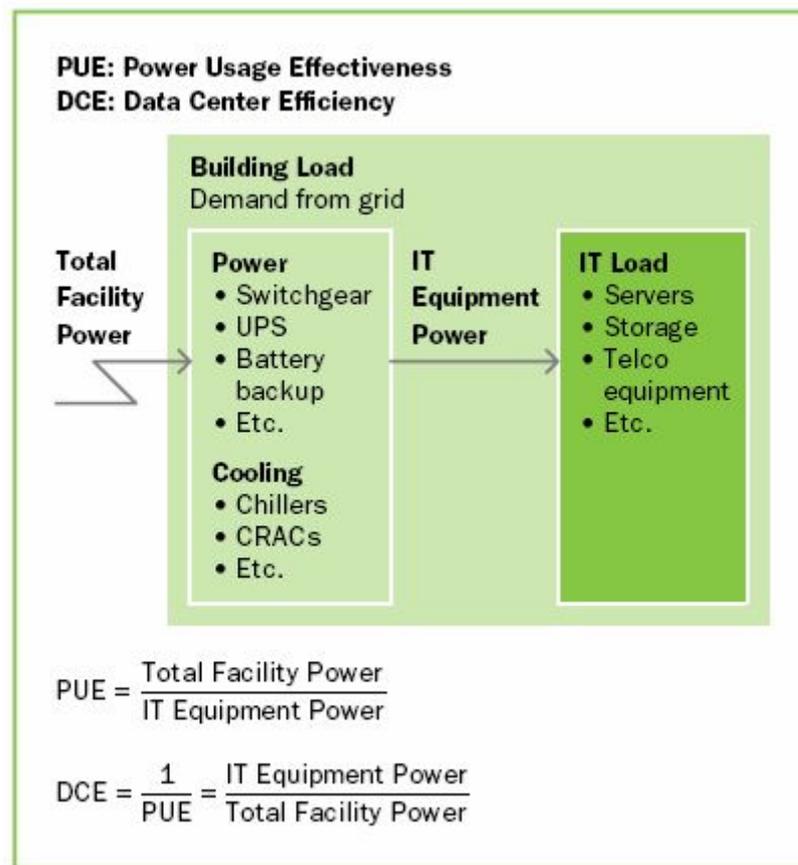


Figure 6. Métriques du Green Grid

Ce schéma met en évidence de gauche à droite :

- le total d'électricité entrant dans la salle informatique ;

⁴⁸ DCE : Data center Efficiency, Livre blanc : « Green Grid Metrics : describing Data Center Power Efficiency »

⁴⁹ <http://www.thegreengrid.org>

- la part consommée par tous les composants distribuant et « protégeant » l'énergie : alimentations électriques, batteries, onduleurs ;
- dans le même groupe : la part consommée par tous les composants participant au refroidissement des équipements : climatiseurs, ventilateurs,... ;
- il reste ainsi une part d'énergie enfin utilisée par les équipements constituant les systèmes d'informations du data center (« IT Equipment Power ») : serveurs, stockages, infrastructures réseaux et télécommunications.

Pour être plus précis dans l'usage des métriques, il conviendrait de tenir aussi compte des composantes suivantes nécessaires au fonctionnement quotidien du data center : éclairage et veilleuses, surveillance : dispositifs d'alarmes, écrans et caméras vidéo, écrans de supervision, climatisation : composants de contrôle et de distribution d'eaux.

On obtient ainsi selon le Green Grid :

$$\text{L'indicateur DCIE}^{49} = \frac{1}{\text{PUE}} = \frac{\text{Total Énergie consommée par les systèmes inf.}}{\text{Total énergie utilisée par le Data Center}} \times 100\%$$

Le DCIE est donc un pourcentage permettant d'évaluer l'efficacité de la consommation énergétique des systèmes informatiques dans le data center.

Si à 100 % d'utilisation, un data center consomme 10 MégaWatts, son utilisation type moyenne consommera environ 70 % ou 7 MW, dont 3 à 4 MW pour l'équipement informatique. On constaterait donc un PUE compris en 1,75 et 2,33.

La société APL France, spécialiste des problématiques des data centers, estime plus précisément à 1,2 million d'euros la facture électrique annuelle moyenne de 2000 m² utiles informatiques à 1 KW au m². Dans ce cas, le PUE dans le data center est donc de 2,3.

→ *Le Green Grid mentionne qu'un PUE approchant de 1 serait très idéal, indiquant une efficacité énergétique proche de 100 %. **Mais actuellement, de nombreux data centers ont un PUE de 3 ou plus, et améliorer le PUE entre 1,6 et 2 est réalisable à terme, un PUE inférieur à 3 étant possible en commençant par modifier le design et l'implantation des systèmes dans le data center.***

2.2.4 Mesure énergétique des infrastructures de stockage

Afin d'anticiper les besoins du data center ou de pouvoir comparer les offres de différents constructeurs, c'est souvent en se référant à la documentation technique que l'on évalue la consommation générale du système de stockage et donc du besoin associé en dissipation thermique.

Mais le résultat du calcul n'est pas aisé, il faut additionner les consommations d'éléments pouvant être des éléments dissociés, par exemple pour une baie de stockage classique : les contrôleurs, tiroirs disques et cabinet ainsi que les éventuels éléments réseaux comme les commutateurs Fibre Channel et Ethernet.

Afin de faciliter ce calcul fastidieux, certains constructeurs fournissent désormais des outils logiciels de type « power calculator » pour faciliter l'anticipation des besoins en climatisation du data center en fournissant à l'acquéreur d'un système de stockage les consommations et dégagements de chaleur exacts.

On peut constater la difficulté d'utilisation de ces premiers calculateurs énergétiques pour les non-spécialistes des baies de stockages. En effet, lors d'échanges avec **Nabil Debbou**, en stage chez Bull, et qui avait parmi ses objectifs d'étudier les outils existants, nous avons pu constater ensemble leur manque de simplicité et de convivialité.

Complexité de la mesure énergétique des systèmes

L'information mesurée principalement dans une baie de stockage est aujourd'hui la température prise près des alimentations. Ces informations ne sont pas assez détaillées pour révéler la température moyenne réelle du système et adapter sa climatisation plus finement.

Il est complexe de mesurer l'efficacité énergétique de systèmes de stockages dont les fonctionnements et rôles sont très différents. Ainsi, une librairie de sauvegarde à base de bandes magnétiques est presque à l'état de veille lorsqu'elle n'est pas sollicitée alors qu'une baie de stockage à base de disques durs est en général constamment « en ligne » et complètement en fonctionnement pour réagir rapidement aux requêtes.

Les métriques d'efficacité énergétique

Le développement de métriques d'efficacité énergétique pour le stockage se confronte aux choix de nombreux paramètres. Faut-il prendre en compte :

- le niveau de consommation comparé à une volumétrie en Go ?
- Avec quelles conditions de charges de travail ? Quelles performances d'accès ? De débit en lecture, en écriture ?

2.2.5 Outil de reporting et management énergétique

Les systèmes de stockage sont désormais accompagnés de nombreux logiciels d'administration et de supervision. Mais il est encore rare de trouver dans ces outils, une évaluation en temps réel ou non, de la consommation énergétique générale d'un système.

L'idéal étant un outil d'évaluation énergétique capable d'aller sonder des systèmes issus de constructeurs différents, de remonter leurs consommations énergétiques et de générer un rapport global résumant « l'activité énergétique » du data center, détaillé système par système.

Il est difficile pour les DSI de mettre en place des indicateurs assez précis pour mesurer, système par système, les consommations mensuelles.

➔ *Le manque d'outils de reporting et d'indicateurs « individualisés » aux systèmes de stockage ne contribue pas à distinguer les bons des mauvais élèves énergétiques dans le data center.*

2.2.6 Difficulté d'approvisionnement énergétique

La saturation du centre de données lié au développement des T.I.C. et la forte contrainte du coût économique de son extension ou de sa création pour accueillir de nouveaux systèmes, ont eu pour conséquence de développer des solutions matérielles plus denses et moins encombrantes comme les serveurs lames et les systèmes de stockage hautement capacitifs.

Cette optimisation de l'espace entraîne une densité énergétique sans précédent et une problématique nouvelle d'approvisionnement de l'énergie. Également, ce niveau de densité

énergétique génère une chaleur exceptionnelle qui nécessite des moyens en climatisation impactant pour l'environnement.

Lors de projets d'extensions, des entreprises ont constaté des difficultés de la part de leur fournisseur d'électricité pour assurer une augmentation d'apport d'électricité à leur centre de données. Notamment lorsque des conflits d'intérêts énergétiques se posent dans une même zone géographique : industries, hôpitaux, transports et particuliers...

→ L'évolution technologique très dense des serveurs et des systèmes de stockage entraîne de nouvelles problématiques de consommation énergétique et de refroidissement. Face aux problématiques de coûts économiques, de saturation d'approvisionnement et de distribution d'énergie vers les data centers, il devient indispensable d'améliorer leur rendement énergétique et de réduire leur impact sur l'environnement.

2.3 Encombrement spatial

Des systèmes dispersés

Diverses raisons, notamment le développement des serveurs « distribués », composés d'architecture X86, ont contribué à faire rapidement croître la taille des centres de données. En effet, l'usage a été d'allouer un serveur à une application, augmentant significativement leur nombre dans le centre de données. Ce développement complexe a généré des infrastructures isolées, peu optimisées et encombrantes, composées de silos.

→ *Les rôles des systèmes de stockage se sont diversifiés : stockage haute performance et/ou haute capacité, sauvegarde sur disque ou bande magnétique, archivages, cd-roms..., plusieurs générations de ces systèmes se côtoient même parfois dans le data center.*

Les réseaux dédiés aux systèmes de stockage contribuent aussi à saturer les data centers

Les systèmes de stockage interconnectés aux serveurs, ont aussi développé parallèlement leur réseau de stockage. **Francis Bianchin**, Product Manager Bull, rappelle que les cartes HBA⁵⁰ servant à la connexion des serveurs vers les commutateurs Fibre Channel et/ou les baies de stockages occupent de nombreux emplacements dans les serveurs. Ces cartes encombrant les serveurs, pareillement, les câbles en fibre optique interconnectant des dizaines à des centaines de ports, prolifèrent dans le data center.

Il faut aussi tenir compte des cartes Ethernet, qui comme les cartes Fibre Channel, voient leurs générations successives se développer et se côtoyer dans le data center.

→ *Toute technologie pouvant regrouper en un seul réseau unifié, les différents réseaux de stockage existants, serait un gain d'espace et d'énergie pour les serveurs et le data center.*

⁵⁰ HBA : Host Bus Adapter

2.4 Pollutions

Le centre de données est responsable par sa consommation énergétique et ses dispositifs de refroidissement, d'importantes émissions de CO₂, mais aussi du traitement de ses déchets électroniques pouvant être toxiques.

Consommation énergétique : émissions de CO₂, électricité

McKinsey et Uptime Institute dans un rapport commun publié⁵¹ en 2008 estiment -en millions de tonnes par an- les émissions de CO₂ à :

- 142 Mt pour l'Argentine, 146 Mt pour les Pays-Bas, 178 Mt pour la Malaisie

- 170 Mt pour les data centers

Ils affirment aussi que les centres de données consomment environ **0,5 % de l'électricité mondiale** et qu'un seul data center de taille moyenne consomme autant que 25 000 foyers.

Des centres de données vieillissants

Les nouvelles réglementations environnementales obligent en amont, dès la fabrication, les constructeurs à améliorer leur produit. Mais les centres de données ont une durée de vie moyenne de 12 ans. Par conséquent, le data center connaît de nombreux retraits de systèmes, notamment pour les systèmes de stockage. Il faut identifier, traiter et recycler ses déchets, conformément aux lois.

Directive européenne RoHS sur les substances toxiques

À partir du 1er juillet 2006, tous les équipements informatiques et télécommunications doivent respecter cette réglementation sur certaines substances toxiques. La société Farnell, leader mondial en fabrication de produit électronique a publié un guide législatif et technique⁵² introduisant les enjeux de la directive et son application pas à pas dont voici un extrait :

Substance	Application
Plomb	Solders/Soudures
	Revêtement des pattes de composants
	Peintures utilisées comme pigments et agents desséchants
	PVC utilisé comme stabilisant
	Piles et batteries (sortant du champ d'application de la Directive RoHS)
Cadmium	Revêtements déposés par électrolyse
	Soudures spéciales (par ex. dans certains types de fusibles)
	Contacts électriques, relais, commutateurs
	Stabilisant PVC
	Pigments à base de plastique, verre et céramique
Certaines matières à base de verre et de céramique	
Mercure	Lampes
	Capteurs
	Relais
Chrome hexavalent	Revêtements passivés sur métaux
	Dans les peintures anticorrosion
PBB et PBDE	Ignifuges dans les matières plastiques

Figure 7. Substances toxiques

⁵¹ <http://uptimeinstitute.org/content/view/168/57> « McKinsey / Institute Report on Data Center Energy Efficiency Released »

⁵² http://fr.farnell.com/images/fr_FR/RoHS_Manual_FR.pdf

Nous pouvons reconnaître dans notre quotidien, certains de ces composants. Nous les croyons sans danger mais ils sont très pollués :

	Un potentiomètre peut renfermer du cadmium		Plomb dans une soudure ou le revêtement d'un contact
	Une ampoule, du verre ou une soudure peut contenir du plomb		Boîtiers plastiques, PBB, PBDE, cadmium et plomb
	Un connecteur plastique ou l'isolant d'un câble peuvent contenir du plomb ou du cadmium		Condensateur électrolytique ; plomb dans le revêtement des pattes et dans le capot en plastique si PVC
	Condensateur MLCC : le plomb présent dans la céramique est exempté, mais pas le plomb sur les pattes		Cadmium ou plomb dans les matières plastiques et plomb dans les revêtements déposés par électrolyse

Figure 8. Composants toxiques

Directive européenne WEEE/RAEE sur les déchets électroniques

Près de 10 000 micro-ordinateurs deviennent obsolètes chaque jour, selon Forrester⁵³. Nous compterons plus de 1 milliard d'ordinateurs en 2008 et plus de 2 milliards d'ici 2015 -donc à priori autant de disques durs. 20 à 50 millions de tonnes « DEEE » sont générées dans le monde chaque année.

→ De leur fabrication à leur utilisation et jusqu'à leur retrait, les systèmes de stockages impactent l'environnement. L'entreprise doit en tenir compte, limiter cet impact, respecter les normes environnementales et ainsi contribuer au développement durable.

2.5 Analyse du cycle de vie (écobilan)

Il est quasiment impossible d'obtenir des documents proposant une analyse de cycle de vie d'un système de stockage récent ou même seulement d'un disque dur. Les constructeurs leaders sur le marché sont souvent des intégrateurs de composants qui font appel à des sous-traitants.

Les constructeurs commencent à être sensibilisés à l'impact environnemental de la fabrication de ces composants. Mais rien ne les oblige à fournir des analyses de cycle de vie

⁵³ Forrester Research : « Worldwide PC adoption Forecast to 2015 »
<http://www.forrester.com/research/document/Excerpt/0,7211,42496,00.html>

et/ou les mesures et engagements environnementaux pris pour le fonctionnement de leur usine et ainsi constituer un bilan général écologique d'un système de stockage.

Problématique de transparence du bilan environnemental de fabrication du produit

Il s'agit de pouvoir estimer l'impact environnemental, comme la consommation énergétique pour la fabrication des produits. Quelle est la nature de l'énergie ? Nucléaire ? Gaz ? Charbon ? Quelle a été la consommation en eau, en pétrole, pendant les étapes de la production et de l'assemblage ?

L'analyse de cycle de vie des produits se voit donc réduite à la partie « recyclage et déchet », pour laquelle en dehors de l'application stricte des directives RoHS et WEEE, les constructeurs montrent peu d'initiatives environnementales, par exemple, la reprise des anciens matériels avec un recyclage et/ou traitement des éléments polluants.

→ *La sensibilisation aux aspects verts des étapes de fabrication et de recyclage des systèmes de stockage est encore peu développée pour être un critère de sélection et d'achat, soit pour les décideurs informatiques, soit pour les responsables RSE⁵⁴, en charge de l'image de l'entreprise qui doit démontrer sa contribution au développement durable et au respect de l'environnement.*

2.6 Architectures de stockage

Les directions métiers, face aux défis de leurs activités, exigent des applications et des données toujours plus « disponibles ». La demande de systèmes en « haute disponibilité » se généralise et démultiplie les besoins en énergie et participe largement à la croissance des données avec des architectures dont les composants suivants doivent être redondants : serveurs-stockages-réseaux.

Aussi le plan de secours ou de continuité d'activité, peut à nouveau « doubler » cette architecture principale « haute disponibilité » (et les données) sur un site secondaire.

2.6.1 Des infrastructures dispersées

De même que les architectures des serveurs, les systèmes de stockages se sont souvent développés en silos indépendants. Ils utilisent des technologies distinctives selon les rôles assumés : stockage, sauvegarde, archivage, protection distante...

Chacun de ces rôles peut requérir plusieurs systèmes de stockage dans le data center, multipliant les composants de l'infrastructure et par conséquent leur besoin en énergie et refroidissement. Différentes générations de chacune de ces technologies cohabitent même parfois ensemble dans le data center.

Un autre exemple concerne les bases de données, qui ont souvent bénéficié de serveurs de données dédiés à une application, multipliant les systèmes de stockages attachés. Pareillement, les différentes versions d'une application (ex. : Oracle 9, 10..) ont souvent chacune leur propre espace de stockage.

⁵⁴ RSE : Responsabilité Sociétale d'Entreprise – Corporate Social Responsibility

2.6.2 Sous utilisation

Lors de la conférence dédiée au Green IT du 23 octobre 2008, IDC a communiqué les résultats de son étude parue le même mois : « *Baromètre du Green IT* »⁵⁵, faite auprès d'une centaine d'entreprises françaises de plus de 1000 salariés et en complément, plus de 350 entreprises européennes.

De nombreuses solutions en place ou projets en cours, ont pour objectifs d'améliorer l'usage, le taux d'utilisation et l'optimisation des systèmes.

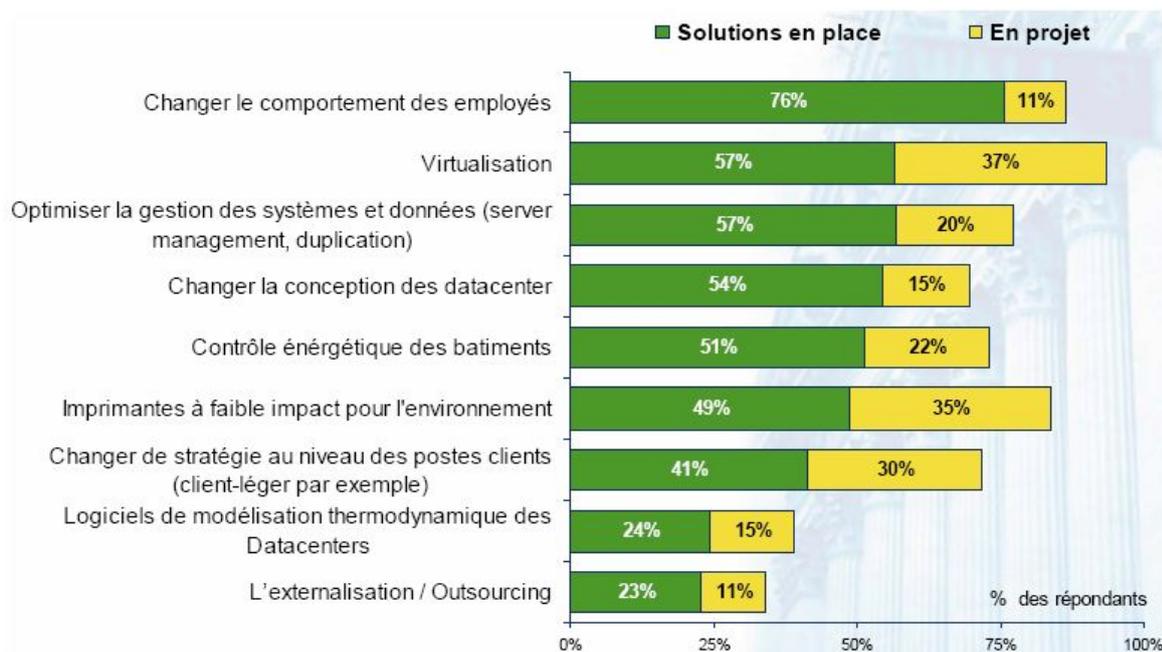


Figure 9. Sondage IDC octobre 2008

Le comportement des employés et l'usage des équipements IT, la virtualisation et l'optimisation de la gestion des systèmes et des données sont les principales solutions mises en place ou en projet, selon IDC.

Le taux d'utilisation des principales familles de serveurs est en moyenne le suivant :

- 5 à 15 % pour les serveurs Intel
- 20 à 30 % pour les serveurs Unix
- 70 à 80 % pour les mainframes

Jean-Marc Rietsch, président de la Fédération européenne de l'ILM, du Stockage et de l'Archivage (FedISA⁵⁶) illustre l'usage et l'administration « quasi » irrationnels des systèmes de stockages par l'exemple suivant :

« Une baie de stockage de 10 To donne 8 To allouables, 4 sont utilisés tandis que 2 To deviennent obsolètes. Il reste donc 2 To disponibles alors on rachète une nouvelle baie. Soit 6 To de perdus, sans parler de la sauvegarde associée... »

⁵⁵ www.idc.com « Baromètre du Green It » réf.: #IDCWP28Q

⁵⁶ <http://eifisa.eu> (FedISA et EIFISA : European Institute For ILM, Storage and Archiving)

➔ *Le centre de données actuel a un taux d'utilisation moyen de 56 % -selon Uptime Institute dans une étude⁵⁷ conjointement réalisée avec McKinsey, détaillant la problématique de sous-utilisation- les systèmes de stockage sont parfois aussi sous-utilisés que les serveurs.*

2.7 Explosion du volume des données

La demande en stockage a augmenté de plus de 50 % par an ces dernières années, plus rapidement même que l'importante décroissance du coût du stockage. Les deux graphiques ci-dessous extraits de la publication de McKinsey « Meeting the demand for data storage » illustrent ces tendances :

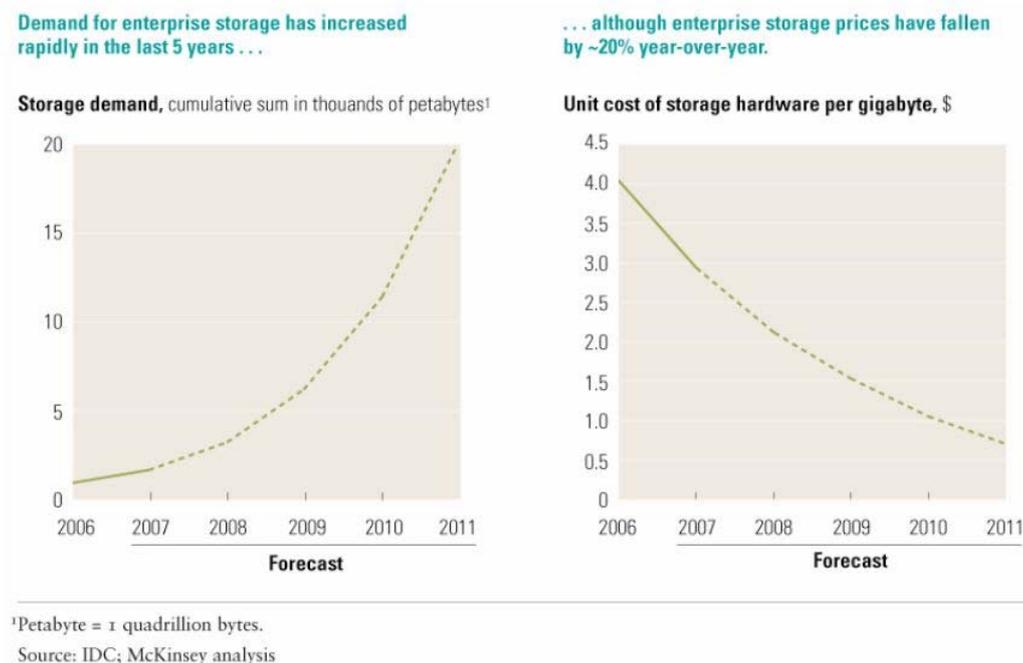


Figure 10. Evolution de la demande en stockage, IDC & McKinsey

La croissance exponentielle des données est due aux fantastiques avancées permises par les systèmes d'informations :

- l'usage « étendu » du mail et des pièces jointes offrant l'instantanéité de la télédiffusion ;
- chaque direction métier de l'entreprise possède des applications complexes qui intègrent chacune leur propre « silo » de données. Les « métiers » veulent capitaliser et exploiter cet actif essentiel, mais souvent sans savoir quelles données sont superflues, inactives ou obsolètes ;
- les applications de type E-commerce, E-Marketing, Helpdesk font proliférer les données ;
- les applications de CRM, Supply Chain, E-Sourcing, et tout ERP, non seulement acquièrent, mais transforment et surtout créent sans cesse de nouvelles données ;

⁵⁷ <http://uptimeinstitute.org/content/view/168/57> « McKinsey / Institute Report on Data Center Energy Efficiency Released »

- les nouveaux outils pour les systèmes d'informations sont très « datavores » : les bases de données relationnelles ou décisionnelles, la Business Intelligence (B.I.) comme le Data Mining⁵⁸ ;
- les données « multimédias » très consommatrices d'espaces de stockage viennent accroître la masse de données :
 - les images et sons issus des systèmes associés au domaine médical, géographique et aussi la vidéo surveillance, la téléconférence et le RFID ;
 - les communications internes et externes de l'entreprise, le plus généralement produites et distribuées au format numérique ;
 - les contenus multimédias créés pour internet ou transmis par ce dernier comme les webcasts et podcasts.

Selon IDC dans une étude publiée en avril 2008 : « L'optimisation du stockage : enjeux, difficultés et stratégies gagnantes des entreprises européennes »⁵⁹ faite auprès de 100 grands comptes européens, la croissance des données devrait encore atteindre plus de 30 % en moyenne en 2008 et en 2009.

Les principaux responsables identifiés sont :

- La **messagerie** : généralisation de l'outil et diffusion de pièce jointe en forte augmentation ;
- la **dématérialisation** des processus et la mise en œuvre de **nouvelles applications** souvent critiques pour l'entreprise : applications métiers, applications mobiles, ERP, CRM, décisionnel ;
- l'augmentation des débits réseau favorise les échanges de **fichiers multimédias** très consommateurs de stockage ;
- les contraintes **réglementaires** (archivages, conservations plus longues)
- la **croissance organique** de l'entreprise

→ *De nouvelles problématiques surviennent face à la prolifération des masses de données et leurs flux en constantes augmentations : comment distinguer les données actives et inactives ? Comment les associer aux activités ? Quelles en sont les propriétaires ? Quelles sont les données obsolètes ? Comment décider si elles sont supprimables ? Comment s'assurer que les données sont crédibles ? auditables ?*

2.7.1 Diversité des données

À la problématique de la croissance des données, s'ajoute celle de la nature des données et de leur diversité.

En effet, chaque application a ses spécificités ainsi que les données qu'elles créent ou utilisent : taille de bloc, accessibilité, fréquence en lecture ou écriture, besoin en bande passante et/ou en entrées-sorties.

⁵⁸ Data Mining : processus d'extraction de connaissances à partir de grands volumes de données

⁵⁹ <http://www.bull.com/fr/storage/wp.html>

Le tableau ci-dessous illustre **la complexité de cette diversité** :

Applications	Lecture / Écriture	Performance	Profil entrée-sortie (I/O)
Décisionnelle	L	Sensible Mo/s	Grands blocs, séquentiels
Transactionnelle	L / E (3/1)	IO/s, latence sensible	Pages, aléatoire
Audio vidéo	L (E)	Streaming, bande passante	Streaming, grands blocs, séquentiel
Collaborative	L (E)	Pas trop sensible	Petites ES, aléatoire
HPC	L/E (1/2)	Sensible Mo/s	Grands volumes, grands blocs, séquentiels

Évolution de la taille des données

L'évolution constante de la taille de certains formats de données audio vidéo haute définition, communément appelé « Rich Media », accroît les problématiques de stockage, mais aussi de transports (bande passante réseau) et de traitements des données.

➔ *Ne pas tenir compte ou ne pas savoir distinguer les besoins en performances de ses différentes formes de données peuvent amener les DSI à stocker les données sur des systèmes de stockages **surperformants et inadaptés**, les plus coûteux et consommateurs d'énergie.*

2.7.2 Données redondantes

Les versions

À des fins de sécurité, de test et de production, il est d'usage de conserver plusieurs versions d'un programme et ses données. La cohabitation de différents systèmes d'exploitation au sein du data center peut aussi contribuer à développer les données redondantes.

Les copies redondantes

Le SNIA dans l'illustration suivante met en évidence les contraintes de l'exploitation d'une volumétrie de donnée et ses conséquences sur la volumétrie réelle finale :

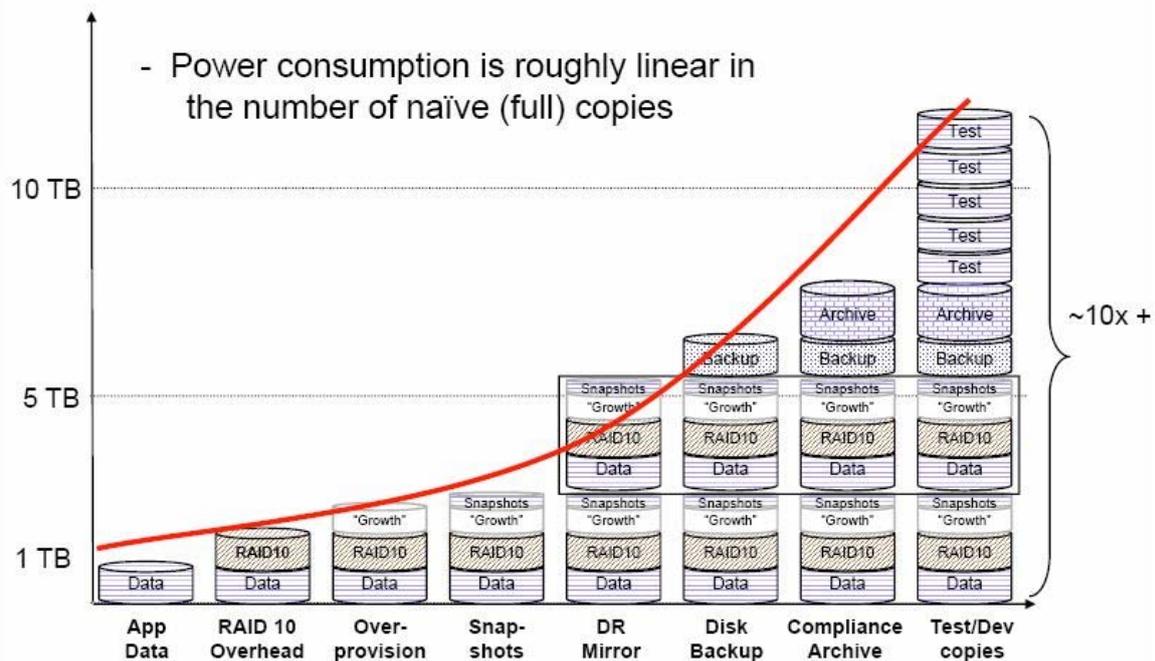


Figure 11. Consommation de la volumétrie, SNIA

En effet, les problématiques suivantes vont avoir plus ou moins d'effet sur l'espace utile nécessaire aux données d'un volume (« Data » sur le schéma, volume de données issue d'une application) :

- 1/ Le choix du type de RAID⁶⁰, ainsi le RAID 10 (RAID 1+0) performant pour les bases de données exigeantes, mais ayant pour conséquence de doubler le nombre de disques (et la volumétrie) nécessaire à sa mise en oeuvre.
- 2/ La surallocation (« Over-provision ») qui est une réserve de volumétrie anticipant les besoins de l'application.
- 3/ Les copies instantanées (« Snapshots ») utilisées à des fins de sauvegardes locales et du volume de données, consomment quelques pourcentages de volumétrie supplémentaire.
- 4/ La réplication ou miroir (« DR Mirror ») du volume de données vers un site distant, exigée par le plan de secours ou de continuité d'activité. Cette réplication peut doubler à nouveau toute la volumétrie issue des 3 premières étapes.
- 5/ La sauvegarde (« Disk Backup ») du volume de données.
- 6/ L'archivage légal (« Compliance Archive ») – selon l'activité, ses normes ou ses contraintes réglementaires - du volume de données.
- 7/ Les copies (« Test ») du volume de données initial (« Data » sur le schéma) nécessaires aux tests et développement et pouvant doubler plusieurs fois la volumétrie initiale.

L'illustration résume donc les conséquences des différentes opérations quotidiennes sur un volume de données, non optimales, mais nécessaires à l'environnement de production.

⁶⁰ Redundant Array of Inexpensive Disks ou Redundant Array of Independent Disks (matrice redondante de disques bon marché ou matrice redondante de disques indépendants)

La messagerie et les pièces jointes

Extrait du livre blanc « The Diverse and Exploding Digital Universe » d'IDC en mars 2008, l'exemple de l'envoi d'un courrier électronique accompagné d'une pièce jointe à quatre destinataires représente un volume de données impressionnant : près de 31,5 Mo pour 1,1 Mo initial dans le cadre d'une politique de sauvegarde classique et formelle.

En effet, le courrier initial a une taille de 0,1 Mo et le document joint, 1 Mo, mais il faut aussi tenir compte pour le mail original : de la copie locale (+1,1 Mo), de la sauvegarde : (+1 Mo), du serveur redondant (+2,1 Mo). Donc l'ensemble sauvegardé sur bande pour archive est égal à 4,2 Mo, le total généré par le mail de l'expéditeur est de 9,5 Mo.

Il faut ensuite tenir compte de l'envoi du courrier électronique à 4 destinataires : 4 copies locales (+4,4 Mo), 4 copies aux serveurs (+4,4 Mo), serveurs de sauvegarde (+4,4 Mo), sauvegardes sur bande pour archive (+ 8,8 Mo) soit un total de 22 Mo.

Un total global de **31,5 Mo** (9,5 + 22) a été généré pour un envoi initial de **1,1 Mo** soit un facteur de près de 30 fois.

→ *Des fonctionnalités logicielles pouvant optimiser et réduire ces données additionnelles et souvent redondantes, sont donc indispensables pour réduire la volumétrie totale de stockage nécessaire, tout en respectant les contraintes de production des activités.*

2.7.3 Données dormantes ou persistantes

Face à la croissance des données, comment distinguer les données dormantes ? Comment reconnaître les données actives pour l'entreprise, des données non actives ?

Selon le Arun Taneja Group⁶¹, 70 % des données stockées ne sont pas où elles devraient être (espaces de productions ou de sauvegardes) et deviennent des données « persistantes ».

Certaines données peuvent rester persistantes dans l'espace de stockage de production, dédié aux applications critiques de l'entreprise et engendrent les problèmes suivants :

- saturation de l'espace de stockage le plus performant et souvent le plus demandé par les directions métiers et le plus sollicité par les applications ;
- consommation de l'espace le plus coûteux matériellement et énergétiquement ;
- engorgement des sauvegardes en encombrant inutilement et en augmentant significativement l'espace de stockage de production qui fait naturellement l'objet de sauvegarde quotidienne ;
- allongement des délais de restauration, soit parce que les données à restaurer deviennent trop volumineuses, soit parce que les données persistantes ne sont plus distinguées des données critiques et que l'on préfère dans le doute restaurer la totalité plutôt que la seule partie nécessaire ;

Selon **Steve Duplessie**, de l'Enterprise Strategy Group, plus de 85 % des données stockées dans les espaces de stockage de production sont « non transactionnelles » ou « post-transactionnelles ».

Selon la société Copan, les données issues de l'activité « critique » de l'entreprise évoluent très vite à un deuxième stade dit « vital ». La probabilité de réutilisation des données dans l'espace « critique » décroît rapidement et devient faible au bout d'environ 30 jours. Les

⁶¹ <http://www.tanejagroup.com>

données de l'espace « vital » ont donc tendance à devenir persistantes, et exigent beaucoup moins de performances.

Sachant que le budget alloué au stockage n'augmente que de 6,3 % en moyenne chaque année, les besoins en stockage augmentent plus vite que les moyens financiers alloués. Les DSI doivent impérativement utiliser le plus rationnellement possible les infrastructures de stockage performantes et critiques pour l'entreprise.

Les DSI appréhendent donc de multiples problématiques :

- Quels sont les supports de stockages les plus efficaces pour chaque type de données ?
- Quelles sont les données inactives ? Quel est leur cycle de vie ?
- Quelles fréquences de sauvegardes sont nécessaires ? Avec quels délais de conservations ?
- Quand et quelles données peuvent être effacées ?

Selon IDC, seulement 22 % des grands comptes interrogés ont mis en place des outils de hiérarchisation des données.

Lors d'un entretien sur le sujet avec **Philippe Reynier de Bull**, celui-ci soulignait : « Quelle est la valeur énergétique d'une donnée ? ». En effet, il convient de comparer la valeur de la donnée à celui de l'espace de stockage occupé, selon le stade de son cycle de vie. Lorsqu'elle reste dans le même espace de stockage, la valeur énergétique de la donnée reste constante tandis que sa valeur réelle a pu croître ou décroître.

→ *Prendre en compte l'évolution des données et leur cycle de vie permettrait une réelle optimisation énergétique car les systèmes de stockage conçus pour les données critiques sont surperformants, capables de répondre aux très fortes contraintes des bases de données tout en offrant une haute disponibilité souvent proche de 99,9 à 99,99 % de l'activité et sont donc les plus « énergétivores ».*

2.7.4 Restitution d'enquête sur les problématiques de gouvernance des données

The Data Warehousing Institute Research⁶², un des leaders en formation et conseil sur la Business intelligence et le Data Warehousing, dans son enquête « Data Governance Strategies », publiée au 2e trimestre 2008, a interrogé environ 400 professionnels des problématiques de management des données. Pour TWDI, les principales raisons légitimant une véritable gouvernance des données sont les suivantes :

- l'accroissement des réglementations et des lois pressant les entreprises pour que leurs données soient conformes légalement et auditables (« accountability ») ;
- le changement de l'organisation qui doit traiter ses données comme un véritable actif de l'entreprise. Il s'agit de savoir qui est propriétaire des données et comment elles sont utilisées. Le besoin devient croissant de mettre en place une direction spécialisée sur les données, comprenant à la fois des représentants des directions métiers, de la DSI et de la direction générale pour pouvoir prendre des décisions transversales aux activités ;
- des activités et opérations nécessitent un management spécifique en amont car elles sont « drivées » par les données, comme la Business Intelligence (B.I.), la CRM (Customer Relationship Management), et toutes opérations de transformation de l'entreprise comme les fusions et acquisitions ;

⁶² Une division de « 1105 Media, Inc. » <http://www.twdi.org>

- pour canaliser les données, des procédures et des règles doivent être mises en œuvre et gérer. Aussi des connaissances spécifiques pour évaluer les risques et les coûts de pertes de données sont nécessaires.

L'enquête classe par ordre d'importance les bénéfices attendus de la Data Governance, basés sur 1 960 réponses de 394 répondants :

- 60 % : qualité des données - 58 % : cohérence des définitions de données
- 52 % : données considérées comme un actif de l'entreprise
- 38 % : prise de décision basée sur les données
- 33 % : usage approprié des données
- 26 % : collaboration entre les équipes, les divisions
- 25 % : plus de responsabilités dans l'usage des données
- 23 % : qualité des « master » données
- 22 % : le partage des données
- 20 % : qualité des « méta » données
- 20 % : visibilité dans l'entreprise par les données
- 18 % : changer les processus du management de l'usage des données et leur gestion
- 17 % : transformations du business
- 16 % : la sécurité des données
- 12 % : les services clients basés sur les données
- 12 % : les données appropriées pour les rapports réglementaires
- 10 % : confidentialité des données

L'enquête révèle en effet que 33 % des entreprises sont déjà dans une phase d'étude ou d'implémentation et que 40 % l'envisagent. 88 % des entreprises interrogées commencent à prendre des initiatives pour s'engager vers une Data Governance.

Gérer la croissance des données est une problématique majeure

En effet, choisir les composants les plus efficaces et les architectures associées à des fonctionnalités les plus optimisantes possible auront peu d'effets si dans le même temps l'entreprise n'arrive pas à canaliser les flux de données en très forte hausse.

→ *Les données deviennent multiples et complexes, souvent redondantes, et à chaque stade de leur cycle de vie, peuvent nécessiter des infrastructures de stockages différentes. Une gouvernance des données devient nécessaire pour appréhender, canaliser et optimiser cette masse proliférante d'informations et contribuer à la rationalisation des ressources de stockages dans le data center.*

2.8 Évolution du rôle de la DSI

J.C. Henderson et N.Venkatraman, dès 1993, avaient anticipé cette évolution dans leur étude « Strategic Alignment Leveraging Information Technology for Transforming Organizations »⁶³ en élaborant un modèle à quatre dimensions d'alignement stratégique :

- stratégie business (externe) : choix des produits et marchés, compétences distinctives, partenariats et alliances ;

⁶³ <http://www.research.ibm.com/journal/sj/382/henderson.pdf>

- stratégie de technologie de l'information (externe) : domaines technologiques, compétences, gouvernance de l'IT ;
- organisation (interne) : infrastructures, processus métiers et managériaux ;
- systèmes d'information (interne) : architectures, processus de développement, portefeuille d'applications, compétences et contrôles.

Alain Le Billan, directeur technique avant-vente Bull, grâce à son expérience dans la direction de grands projets informatiques, souligne cet élargissement du rôle de la DSI :

- elle porte des responsabilités plus polyvalentes, proactives et transversales dans les directions métiers à l'inverse de l'alignement traditionnel où elle exécutait principalement une fonction de support ;
- elle contribue à conquérir ou s'adapter aux marchés « externes » et aux objectifs business et doit savoir remettre en question ses réalisations ;
- elle n'a plus que des « clients », comme les directions métiers, pour lesquels elle apporte des prestations, dans une dynamique de collaboration, en s'appuyant sur leur expertise métier.

La problématique principale de la DSI est de réussir cette adaptation stratégique, technologique, opérationnelle, compétitive.

Sa contribution est attendue pour qualifier les besoins des directions métiers ; cartographier les applications et les données associées ; aider à évaluer les risques financiers de pertes de données ; définir une politique de protection de données en comprenant les exigences des activités des directions métiers et leurs contraintes de sauvegarde, d'archivage et de plan de reprise d'activité.

➔ **La DSI est en pleine mutation** : d'une « Direction des Systèmes d'Information », elle évolue progressivement vers une « Direction des Services d'Information ». Elle est contributrice de réduction de coûts et surtout d'améliorations de l'organisation et des processus. Elle doit s'organiser pour optimiser et rationaliser l'infrastructure de stockage tout en améliorant les services délivrés aux directions métiers.

2.9 Différencier les performances Green

Distinguer les différents résultats atteints par les solutions « vertes »

Dans un registre complémentaire à cette approche, **Greg Shulz**, fondateur et consultant senior de la société Storage IO Group⁶⁴ (analyste en technologie de l'information) détaille la notion d'usage effectif en plusieurs catégories comme le montre l'illustration suivante :

⁶⁴ <http://www.storageio.com>

Boost Energy Efficiency – Useful Work per Energy Use

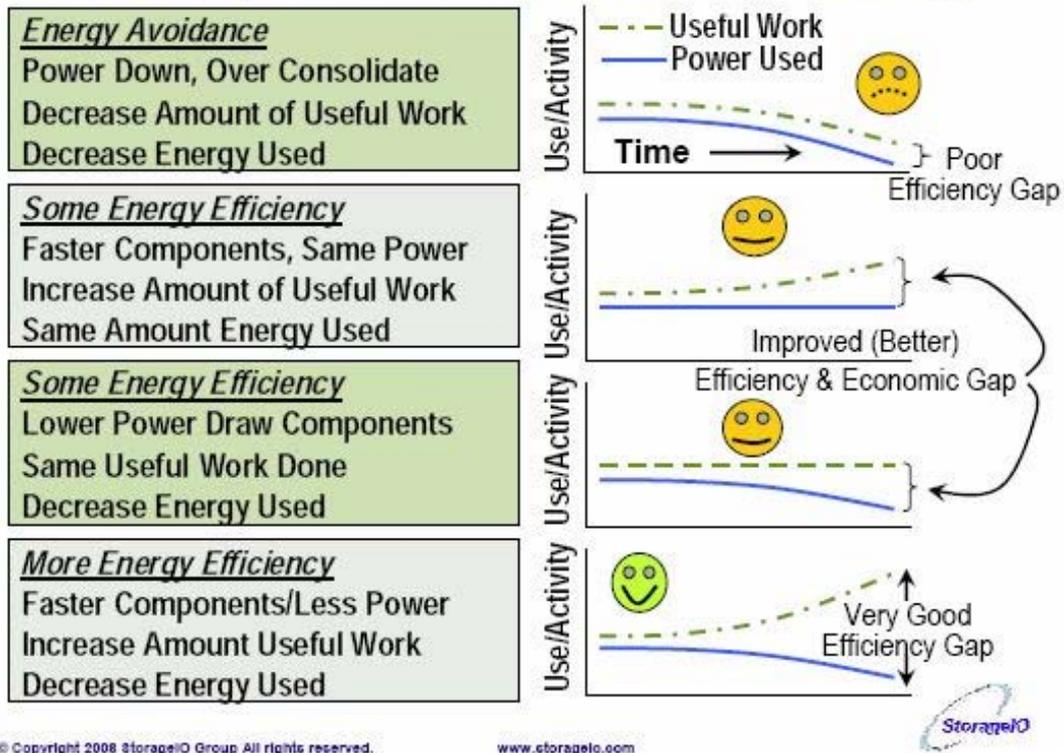


Figure 12. Les différentes efficacités Green IT, StorageIO Group

On distingue dans l'illustration du groupe StorageIO ci-dessus, quatre catégories distinctives possible de dispositifs contribuant au Green Storage :

1/ Par suppression d'énergie : par coupure d'alimentation ou par retrait de système après consolidation. Le niveau de l'énergie baisse comme le niveau d'usage.

2/ Par plus d'efficace énergétique : grâce à des composants plus performants consommant la même énergie. On augmente le niveau d'usage possible mais on dépense la même énergie.

3/ Par plus d'efficace énergétique : grâce à des composants plus économes, mais offrant le même niveau de performance. Ces dispositifs font décroître la consommation énergétique pour une activité IT similaire.

4/ Par beaucoup plus d'efficace énergétique : les composants sont plus performants et plus économes en énergie. Ils augmentent le niveau de performance et l'usage possible tout en diminuant la consommation énergétique.

Greg Shulz est un conférencier reconnu pour son approche sur le Green Storage et a publié plusieurs documents⁶⁵ sur le Green IT et le Green Storage.

⁶⁵ <http://greendatastorage.com/events.html>

2.10 La Responsabilité Sociétale d'Entreprise

La loi sur les N.R.E. (Nouvelles Régulations Economiques) oblige depuis 2002, les sociétés françaises cotées en bourse, à « préciser dans leur rapport annuel la manière dont elles prennent en compte les conséquences sociales et environnementales de leurs activités ⁶⁶ ».

Benjamin Bergeron, en charge de la Responsabilité Sociétale d'Entreprise (RSE) de Bull, souligne l'importance d'une combinaison équilibrée de 3 domaines, les 3 « P » :

- Profitability – Créer la prospérité économique est la fondation ;
- People – Apporter de l'équité sociale, les individus sont les acteurs clés du succès ;
- Planet – Assurer une qualité environnementale, réduire le gaspillage, la pollution ;

Les problématiques de développement durable ont de multiples natures :

- éthiques : entreprise citoyenne, « Bien agir », normes morales, corruption ;
- sociales : capital humain, niveau de compétence, formation, chômage, insertion ;
- environnementales : lutte contre la pollution, préservation, offres / produits socialement et environnementalement respectables.

→ La R.S.E. encourage les infrastructures permettant de diminuer l'impact de l'entreprise sur son environnement. Elle contribue à stimuler les critères de choix d'investissements « éco-responsables » pour l'environnement, l'éthique et l'équité sociale, car l'entreprise joue sa réputation avec des parties prenantes aux exigences différentes : salariés, actionnaires, clients, partenaires, fournisseurs.

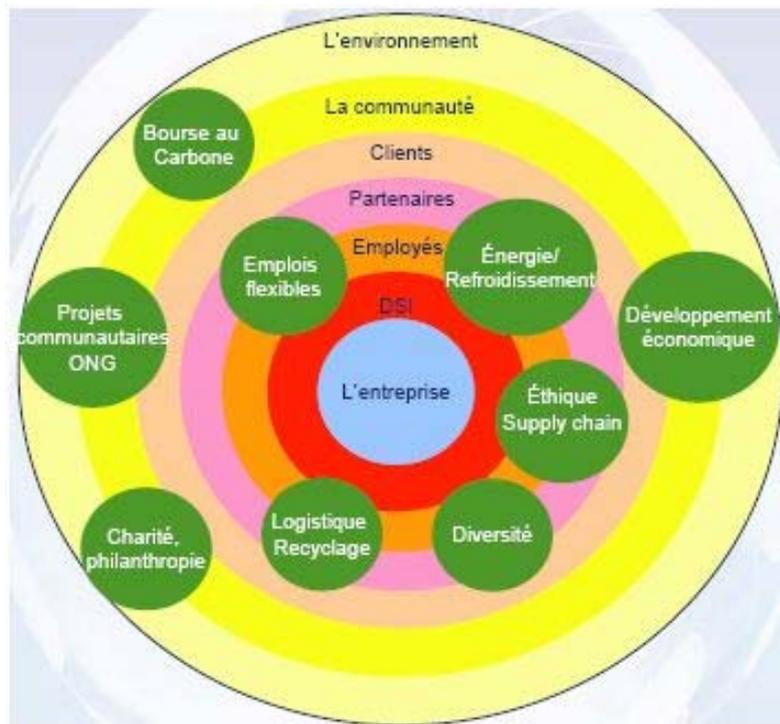


Figure 13. Illustration d'EMC de la R.S.E.

⁶⁶ http://www.ecologie.gouv.fr/article.php3?id_article=4030

2.11 Méthodologie de collecte et d'identification des problématiques et facteurs clés

Afin d'identifier les principaux facteurs clés et critères « Green Storage », les démarches de collecte et d'analyse d'informations se sont appuyées sur différentes sources :

- des interviews d'experts techniques du secteur du stockage, de spécialistes du Green IT, de représentants de DSI et de responsables d'exploitation ;
- la littérature académique issue de la bibliothèque HEC et de l'École des Mines de Paris et extraite des bases de données (comme Forrester, McKinsey, Business Insights, Gartner et IDC), des dossiers et livres ;
- des articles professionnels : sites internet spécialisés, revues, dossiers, presses ;
- des rencontres et entretiens non formels lors de salons et colloques spécialisés sur le thème Green IT ;
- des échanges et discussions avec mes tuteurs académiques et professionnels ainsi que les enseignants et intervenants tout au long du cycle du Mastère.

FACTEURS CLES DE SUCCES « GREEN STORAGE », RECOMMANDATIONS ET MEILLEURES PRATIQUES

Seront abordés...

- l'implémentation dans le data center ;
- les technologies optimisant l'usage de l'espace ;
- les composants et systèmes plus économes en énergie ;
- les architectures : consolidation, virtualisation et mutualisation ;
- les fonctionnalités optimisant les volumétries ;
- les aspects de rationalisation : réseaux, espaces de stockage, cycle de vie des données, usage de la sauvegarde et de l'archivage ;
- la canalisation du développement des données par une « Data Governance » ;
- un catalogue de services de stockage pour la DSI et les Directions Métiers ;
- les audits, outils et programmes d'améliorations ;
- l'intégration d'indicateurs et de nouvelles métriques.

"Les pattes du canard sont courtes, il est vrai ; mais les allonger ne lui apporterait rien."

~Tchouang Tseu (-315)

Suite à la collecte et à l'analyse des problématiques, ce chapitre approfondit les principales recommandations et meilleures pratiques pour rendre plus « éco-responsables » le choix des composants, des fonctionnalités, des architectures ainsi que le déploiement et l'utilisation des systèmes de stockage dans l'entreprise.

3.1 Implémentations, technologies et composants

3.1.1 Implémentation des systèmes dans le data center

« **Les meilleures pratiques en matière de dissipation thermique** », regroupées dans l'étude IDC « Baromètre du Green IT » exposent différentes voies d'amélioration du rendement énergétique et de la climatisation grâce à des implémentations spécifiques des systèmes dont voici quelques exemples :

- plancher surélevé, réduction de la hauteur de plafond ;
- refroidissement par eau, disposition alternée des ailes (couloirs) chaudes et froides ;
- le "free cooling" consistant à refroidir un bâtiment par ventilation en utilisant l'énergie de l'air extérieur lorsque celui-ci présente une température inférieure à la température intérieure ;
- trappe arrière échangeur de chaleur, modélisation thermodynamique ;

Illustration de mise en place d'ailes chaudes et d'ailes froides alternées, principe de design de data center canalisant et améliorant l'extraction d'air chaud :

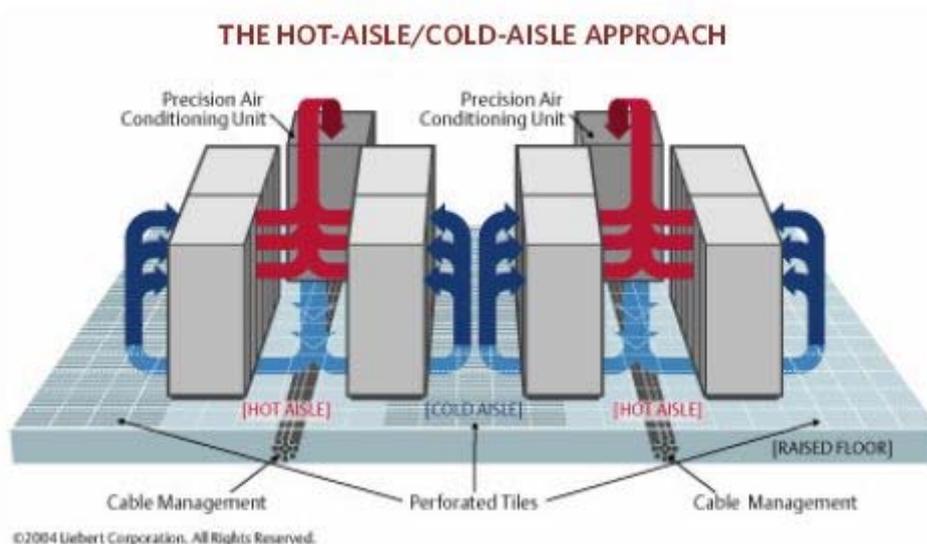


Figure 14. Principe alternance aile chaude / aile froide

Améliorer le rendement énergétique

De nombreux data centers ont un PUE (Power Use Effectiveness), ratio de rendement énergétique, supérieur à 2 voir 3, soit 3 watts consommés pour faire fonctionner 1 watt d'équipement informatique.

La facture électrique⁶⁷ annuelle d'un data center de 2000 m² utiles à 1 kW/m² et avec un PUE de 2,3 est d'environ 1,2 million d'euros. Après la mise en place de nouvelles implémentations et équipements efficaces, si le PUE est amélioré et ramener à 1,8, l'économie est de 260 000 euros par an.

Utiliser les équipements de haute disponibilité pour les activités les plus critiques

Le recours systématique à la redondance et à la haute disponibilité, multiplie les systèmes en fonctionnement, souvent à un faible taux d'utilisation, et augmente significativement la consommation énergétique.

L'ensemble des équipements redondants du data center doivent être contrôlés et justifiés selon la nature des activités et des impératifs des directions métiers. Voir en annexe, le classement par tiers de l'Uptime Institute des data centers selon leurs niveaux et dispositifs de sécurisation.

Optimiser localement les ressources du data center

La surface totale des salles informatiques du site de Bull Trélazé est de 3 100 m². Les bureaux de l'agence de Trélazé sont chauffés par la chaleur récupérée à partir des dispositifs de climatisation des salles informatiques.

→L'application de recommandations et meilleures pratiques d'implémentations des équipements dans le data center, peut améliorer significativement son rendement énergétique, réduire la facture d'électricité et son impact sur l'environnement. Les économies réalisées, permettent aussi de financer de nouveaux équipements plus efficaces.

3.1.2 Technologies réduisant l'empreinte spatiale dans le data center

Certaines caractéristiques technologiques peuvent améliorer le rendement spatial des systèmes de stockage dans le data center :

- une plus forte densité de disques dans un seul tiroir ou rack, en parvenant à supprimer les problèmes liés aux vibrations générées par leur vitesse de rotation ;
- la capacité à densifier le nombre maximum possible de disques dans un seul cabinet ou rack et à accepter des technologies de disques durs différentes : Fibre Channel, S-ATA, SAS, et obtenir ainsi avec un seul système la mise à disposition de différents types d'espace de stockage ;
- la capacité de diminuer le nombre de contrôleurs disque nécessaire pour gérer un ensemble de disques durs le plus grand possible ;
- les contrôleurs ayant de grandes possibilités de connectivité (connexions entrantes et sortantes) et la plus large en termes de protocoles de communication, peuvent ainsi assumer plusieurs rôles : SAN, NAS, et largement s'interconnecter aux serveurs, parfois en s'affranchissant du besoin de nouveaux switches.

⁶⁷ Source <http://www.apl-france.fr>

➔ *Même si de nouveaux systèmes de stockage hautement capacitifs ont un « packaging » très dense et plus consommateur d'énergie, ils ont l'avantage de diminuer le nombre de systèmes de stockage nécessaire et le besoin en refroidissement associé dans le data center. Leurs multiples capacités leur permettent de remplacer plusieurs systèmes de stockage.*

Optimisation des infrastructures de réseaux de stockage avec la technologie Fibre Channel sur Ethernet : FCoE

Cette technologie performante (10 Gb/s) permet d'utiliser le protocole Fibre Channel sur Ethernet et de réduire les réseaux de stockage physiques dans le data center en faisant converger les différentes connectivités des serveurs dans une seule connexion physique.

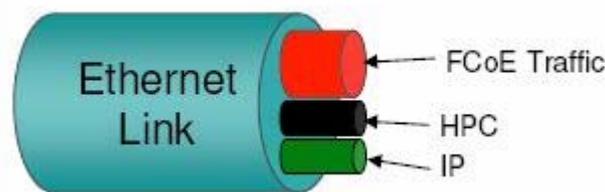


Figure 15. Coupe illustrée d'un câble FCoE

Regrouper les différents réseaux de stockage existants est un gain d'espace et d'énergie pour le data center, notamment en :

- diminuant le nombre de câbles réseau nécessaires par 2 ou par 3 ;
- réduisant le nombre de cartes HBA : les interfaces physiques des serveurs ;
- simplifiant l'administration et la maintenance des réseaux de stockage.

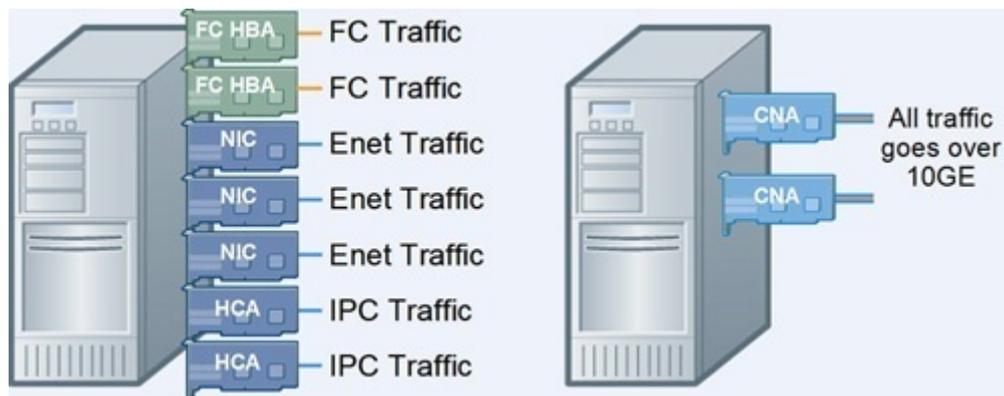


Figure 16. Bénéfice du FCoE

La partie gauche de l'illustration représente la situation actuelle : les cartes occupent une place importante dans les serveurs (en tenant compte de la redondance nécessaire pour disposer d'une architecture « haute disponibilité ») pour connecter jusqu'à 3 types de réseaux différents. Dans la partie droite, la nouvelle technologie FCoE a l'avantage de faire converger l'ensemble des réseaux sur une seule carte (ou deux pour la redondance).

3.1.3 Technologies et composants optimisant l'efficacité énergétique

De nouvelles technologies de stockage innovantes se sont récemment développées, dont les consommations énergétiques sont plus efficaces.

Ces technologies ont différentes stratégies de réduction énergétique :

- la technologie SSD : mémoire flash électronique - pas de composants mobiles ;
- la technologie LowPower S-ATA : des disques durs capacitifs « basses consommations », plus économiques ;
- la technologie à variation de vitesse de rotation de disque « low spin » : adaptant leur vitesse et donc leur performance aux besoins de l'activité ;
- la technologie MAID : capable de suspendre le fonctionnement des disques durs dont les données ne sont pas accédées ;
- de nouveaux composants émettant moins de chaleurs ou capables de fonctionner à des températures plus hautes sans refroidissement.

3.1.3.1 La technologie Solid State Drive

Des unités électroniques constituées de mémoire flash de type NAND⁶⁸, portent les performances de l'espace de stockage primaire (production) à un niveau inégalé et pourraient remplacer progressivement les disques durs fibre channel, encore bien moins coûteux que les SSD, mais aux consommations énergétiques très fortes.

Les bénéfices principaux des disques SSD comparés aux disques durs sont les suivants :

- performances des traitements Entrées / Sorties (E/S=I/O⁶⁹) supérieures de 10 à 30 fois ;
- temps de réponse inférieurs de 9 fois (1 ms contre 6 à 9) ;
- accès « multi-channel », traitement parallèle des E/S, vitesse de transfert ;
- faible consommation énergétique due à l'absence de pièces mécaniques.

Le constructeur EMC a été parmi les premiers à introduire cette technologie dans ses baies de stockage en 2008. Si leur coût est encore très élevé, on peut allouer aux disques SSD uniquement la partie la plus gourmande en transaction des applications et laisser la partie la plus volumineuse, mais moins exigeante, aux disques durs. On améliore ainsi significativement la performance globale tout en réduisant le nombre de disques durs précédemment nécessaires aux traitements des entrées-sorties.

⁶⁸ NAND : abbreviation de "not" - "and" soit " non " - "et", opérateur logique de l'algèbre de Boole

⁶⁹ I/O : Input/Output , Entrée / Sortie lors de transaction lecture / écriture de donnée

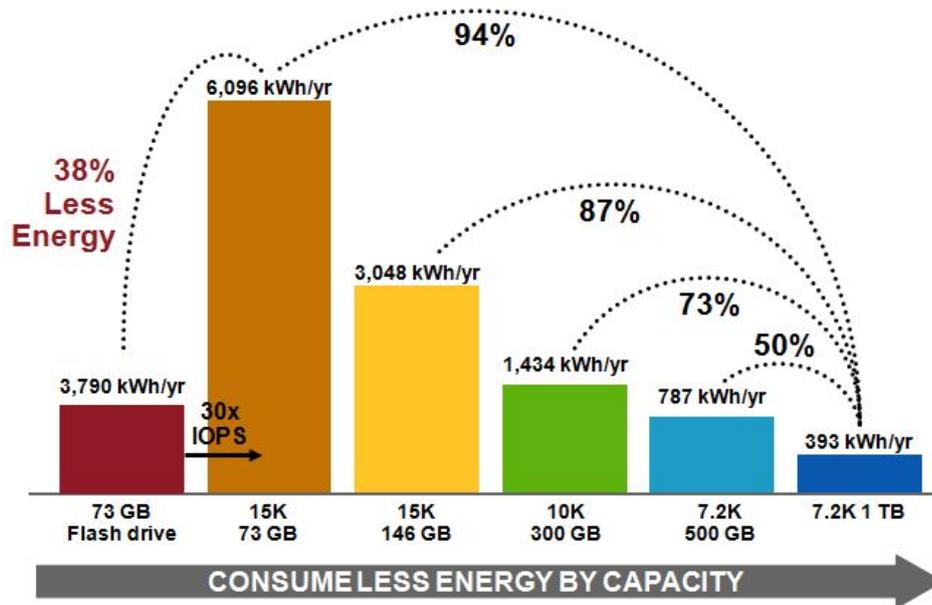


Figure 17. Consommation des technologies disque et capacités, EMC

Cette illustration du constructeur EMC positionne les disques « Flash Drives » 73GB (SSD) à gauche, pour une volumétrie d'un Téraoctet de données, comme ayant le meilleur ratio « consommation énergétique / capacité de traitement transactionnel », comparé aux autres technologies de disques durs (sur le schéma, les capacités des disques durs Fibre Channel : 73, 146, 300 Go et disques S-ATA : 500 Go, 1 To).

Les disques SSD créent un nouveau « tier » (catégorie) de stockage appelé « tier 0 » de par leur performance inégalée pour les activités transactionnelles comparée au tier 1 habituel à base de disques durs Fibre Channel ou SAS.

3.1.3.2 Les disques durs basses consommations

Des disques durs basses consommations apparaissent comme le nouveau disque dur 2,5 pouces Seagate 300 Go, 10 000 tours par minute pour 6,2 W en fonctionnement et 3,6 W en veille ou celui de Toshiba s'appuyant sur 2 plateaux tournants à 5 400 tours par minute pour 400 Go et 2,5 pouces, ne consommant en moyenne que 1,9 W.

Ces disques durs répondent idéalement aux besoins fortement capacitifs (stockage secondaire) tout en optimisant le ratio capacité/consommation. Certains sont capables d'adapter leur fonctionnement selon l'activité : soit réduire leur rotation ou quasiment arrêter leur fonctionnement lorsque les données qu'ils hébergent ne sont pas accédées.

Cette autre illustration ci-dessous du constructeur EMC représente les disques durs de technologie S-ATA II 5 400 tours par minute, pour une volumétrie d'un Téraoctet de données, comme ayant la plus faible consommation énergétique pour leur capacité :

1 To de données sur des disques de différentes capacités et performances

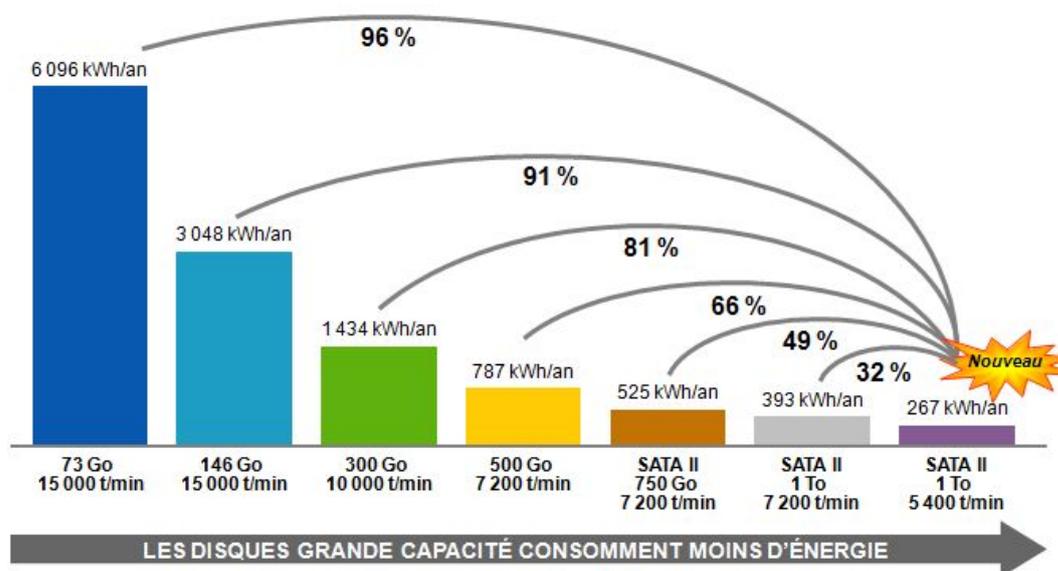


Figure 18. Disques durs : consommation énergétique et capacité, EMC

Sur le schéma, les disques durs de 15 000 et 10 000 tours par minute sont de technologie fibre channel, les disques de 5 400 à 7 200 tours par minute sont de technologie S-ATA II.

3.1.3.3 Les systèmes MAID

La technologie MAID, à l'origine proposée par le constructeur Copan, signifiant Massive Array of Idle Disks, est en principe un système de stockage « massif » ou « dense » intégrant dans leur châssis les caractéristiques suivantes :

- plusieurs dizaines à centaines de disques durs S-ATA fortement capacitifs offrant un bon ratio d'occupation spatiale dans le data center pour la capacité totale offerte sur le nombre de m² occupé ;
- destiné aux applications dites : « écrire une fois » et « lire occasionnellement » ;
- les disques durs dont les données sont accédées sont mis en fonction alors que ceux dont les données ne sont pas utilisées sont suspendus.

Dès lors que les données des systèmes MAID sont accédées plus fréquemment, leur profil énergétique devient moins efficient et s'approche de celui des baies de stockage conventionnelles. Un temps d'accès plus long aux données même s'il est optimisé par la mémoire cache du système MAID, réserve pour le moment cette technologie à des usages ni transactionnels ni critiques.

→ La moyenne de disques durs fonctionnant simultanément dans un système MAID peut aller jusqu'à 25 %, réduisant significativement la consommation énergétique du système et son besoin en climatisation.

3.1.3.4 Les bibliothèques de bandes magnétiques

Le changement progressif actuel de dispositifs de sauvegardes à base de bibliothèques de bandes magnétiques vers des sauvegardes à base de baies de disques durs fortement capacitives (en raison d'opérations de sauvegarde et de restauration plus fiables et rapides) pourrait

connaître un ralentissement du fait de la grande efficacité énergétique de ces systèmes ne consommant réellement que lorsqu'ils sont sollicités pour des sauvegardes ou des restaurations.

Ainsi « l'ancienne » technologie à base de bande magnétique se retrouve aujourd'hui en position de « leader énergétique » pour la conservation à long terme, car une fois que les données ont été écrites, elles sont conservées avec une facture électrique proche de zéro. Ce support de stockage est un des meilleurs choix énergétiques pour une rétention de longue durée.

3.1.4 Évaluer la consommation énergétique

Superviser et comprendre les consommations énergétiques des systèmes de stockage

Si la DSI souhaite comprendre et améliorer la consommation énergétique de ses infrastructures de stockage, elle doit mettre en place des indicateurs de consommations associés à chaque système. Des relevés réguliers permettent alors d'avoir une vision plus fine des évolutions de consommations liées aux charges d'activités.

Afin de construire ces relevés, les systèmes de stockage proposant des outils de reporting de la consommation électrique (charge, pics, moyennes...) ainsi que leur intégration à d'autres outils synthétisant la consommation globale du data center, donnent aux administrateurs les moyens de comprendre et de localiser les différentes consommations énergétiques.

Gérer les ressources de stockage et leur énergie selon l'activité

La capacité des infrastructures de stockage à être mise en veille, partiellement ou non, selon l'activité de l'entreprise et les besoins réels de ses directions métiers est une composante essentielle et directe de la réduction de la consommation énergétique du data center.

Des nouveaux métriques d'efficacité énergétique pour le stockage

Connaître l'efficacité des systèmes, leur usage et leur performance, est une problématique abordée par le Green Grid. Ce dernier propose une nouvelle mesure plus innovante, le DCPE⁷⁰ :

$$\text{DCPE} = \frac{\text{Useful Work}}{\text{Total Facility Power}} = \frac{\text{Fonctionnement utile}}{\text{Total énergie utilisée par le Data Center}}$$

Cette approche est comparable à celle faite par l'agence américaine EPA associée à l'organisme SPEC⁷¹, mettant actuellement au point un standard d'évaluation de performance des serveurs en rapport avec leur consommation énergétique : quantité de transactions réalisées et watts dépensés pendant une période définie.

Le Green Grid souhaite étendre la mesure des systèmes, non seulement aux serveurs et aux éléments de stockage, mais aussi à tout le data center. Dans son livre blanc « The Green Grid Productivity Indicator », est souligné l'intérêt de mesurer le taux d'utilisation de chaque infrastructure du data center : serveur, réseau, stockage, alimentation et refroidissement pour obtenir un indicateur de productivité globale :

⁷⁰ DCPE = Data center Performance Efficiency

⁷¹ www.spec.org Standard Performance Evaluation Corporation

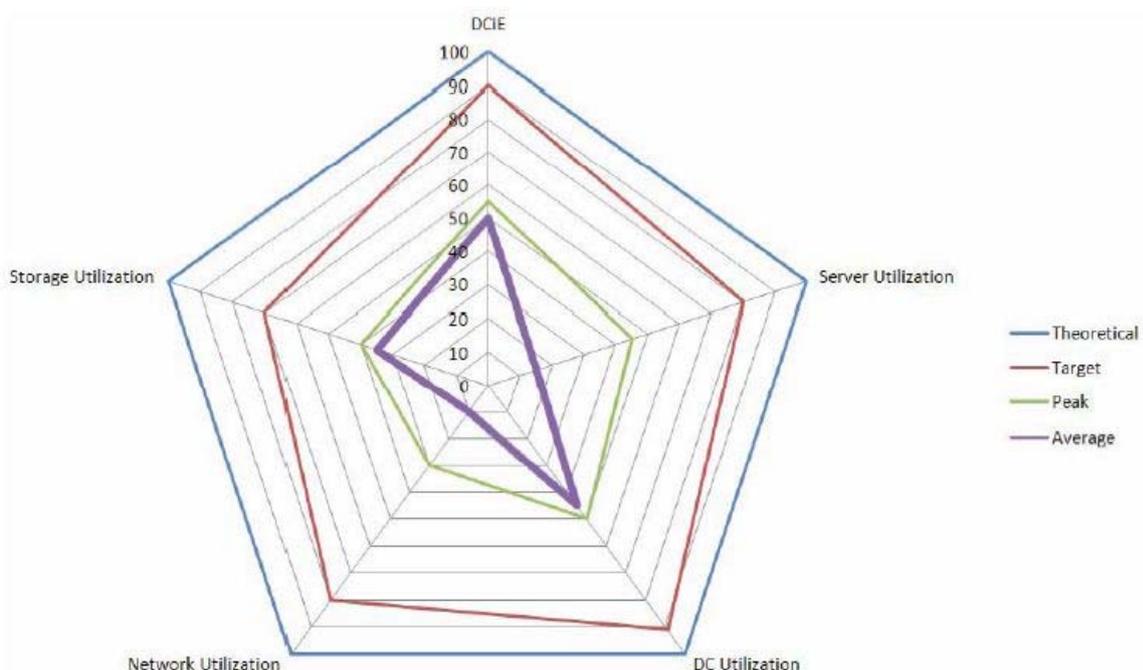


Figure 19. Indicateur radial du Green Grid

Cet indicateur de productivité sous forme radiale permet de visualiser instantanément le taux d'utilisation théorique et cible, ainsi que sa moyenne et son maximum. Le tableau ci-dessous indique ces différents paliers :

Objectifs de productivité

	Théorique	Cible	Max.	Moyenne
DCIE ⁷²	100	90	55	50
Utilisation des serveurs	100	80	45	15
Utilisation du data center	100	91	50	45
Utilisation du réseau	100	80	30	10
Utilisation du stockage	100	70	40	35

L'intérêt de rationaliser ces objectifs est de mieux situer un objectif de productivité raisonnable par rapport aux bénéfices attendus.

SNIA : mesure des équipements inactifs

À défaut de pouvoir obtenir un consensus sur une mesure standardisée d'efficacité énergétique d'équipements actifs, du fait de la difficulté de définir un standard de charge associé à un type d'activité, l'association SNIA spécialisée dans les problématiques liées au stockage, propose dans un premier temps, un standard de mesure d'équipements inactifs.

⁷² Data Center Infrastructure Efficiency : total de l'énergie utilisée pour l'informatique divisé par le total de l'énergie du data center – contraire du PUE : Power Usage Effectiveness

3.2 Architectures et fonctionnalités

Ce chapitre approfondit un deuxième niveau de recommandations et meilleures pratiques, contribuant à l'optimisation des systèmes de stockage. Il s'agit des choix d'architectures opérationnelles, de la mise en œuvre des systèmes de stockage et leur implémentation dans le data center en intégrant des fonctionnalités optimisant à la fois leur utilisation, leur capacité, leur performance et leur rendement énergétique.

3.2.1 Consolidation

La consolidation des données de plusieurs systèmes de stockage vers un seul système centralisé et unique apporte des gains notoires :

- administration : passage de supervisions multiples de systèmes hétérogènes à une administration centralisée et simplifiée ;
- maintenance : moins de pièces détachées et de support logiciel différents et coûteux ;
- coût de formation des équipes d'exploitation et des utilisateurs diminué
- réduction des systèmes de stockage en îlot, requérant souvent leur propres alimentations et climatisations ainsi que des m² dans le data center ;
- réduction du besoin en refroidissement en regroupant les différents systèmes de stockage sur un seul système plus efficient.

Les différentes composantes de stockage peuvent avoir chacune un rôle à jouer lors d'un projet de consolidation :

- le stockage primaire : un regroupement de différentes baies de stockage en une seule, aura pour conséquence de réduire le nombre de dispositifs de climatisation, facilitera l'interconnexion aux serveurs ;
- le stockage secondaire : associer plus de disques durs à un ou deux contrôleurs de stockage diminuera le nombre de systèmes nécessaires ainsi que la climatisation associée ;
- les bibliothèques de sauvegarde : associer un dispositif de bande virtuelle (VTL)⁷³ et de cache sur disques durs permet d'optimiser la sauvegarde et de réduire le nombre de bibliothèques ou de lecteurs de bande magnétique nécessaires ;
- les réseaux de stockage : par exemple, utiliser 2 ports 4 Gb/s au lieu de 4 ports 2 Gb/s, ou même 1 seul port 8 Gb/s si les critères de haute disponibilité le permettent.

➔ *La consolidation, en retirant de multiples systèmes de stockage, retire aussi différents équipements d'alimentation et de refroidissement, améliorant ainsi le rendement énergétique du data center. Elle libère aussi de précieux m².*

3.2.2 Les fonctionnalités logicielles optimisant les volumes de données

3.2.2.1 La compression

La compression est aujourd'hui utilisée constamment dans les technologies de télécommunication. Les technologies de stockage mettent en œuvre cette fonctionnalité d'optimisation de la volumétrie de deux manières possibles :

- compression sans perte (par ex. de type Lempel-Ziv) ;
- compression avec perte, également dite non conservative.

⁷³ Virtual Tape Library – Librairie de bande virtuelle

Certaines données de type audio vidéo (souvent responsables de l'engorgement des disques durs personnels ou professionnels) sont déjà compressées par les applications, par exemple les formats JPG, MPEG, MP3 (compression avec perte) et ne pourront tirer parti de ces algorithmes logiciels ou matériels.

La compression peut entraîner une baisse de performance si la donnée doit être encodée quand elle est écrite et décodée (décompressée) quand elle est lue. Une compression sans perte peut générer de l'ordre de 30 à 50 % de réduction de volumétrie sur certains fichiers.

Il peut être opportun de compresser les données sur des disques de type fibre channel, afin d'accroître leur capacité et de tirer parti de leur performance, si la compression et la décompression ne gênent pas l'activité.

3.2.2.2 *La déduplication en mode fichier ou bloc*

Dédoublonner les fichiers identiques

La fonctionnalité de type « Single Instance Storage » élimine dans un volume, tous les fichiers redondants pour n'en conserver qu'un seul avec les pointeurs associés.

Il existe des outils auditant les volumes de données pour détecter les données en double. L'espace de stockage d'un serveur de messagerie (envoi de mail avec pièce jointe à de multiples destinataires) ou d'un serveur de fichiers, peut comprendre de nombreux fichiers bureautiques identiques, occupant inutilement les volumes de données dès lors que l'on peut associer à un seul fichier de multiples pointeurs vers leurs propriétaires.

Alors que la compression des données s'opère au niveau du bit, la déduplication peut agir au niveau fichier et aussi au niveau du bloc de données. Elle peut réduire le besoin de stockage d'une volumétrie de données pour un ratio⁷⁴ de 15 :1.

Le principe est de conserver dans un volume de données uniquement les blocs de données originaux et en générant seulement des pointeurs pour les duplicats (blocs redondants). Les blocs de données redondants sont identifiés et référencés vers un seul bloc identique et original avec leur pointeur. Ainsi les blocs redondants ne sont pas pris en compte lors de la sauvegarde (sur disque ou bande).

La déduplication en mode fichier peut être moins efficace, en effet dès lors que deux fichiers sont identiques à 99 %, deux copies seront tout de même conservées. Les ratios de déduplication sont alors seulement de 3 ou 4 :1 Les meilleurs résultats étant obtenus avec les postes de travail et les ordinateurs portables stockant surtout des fichiers bureautiques.

➔ Dédupliquer un volume de stockage améliore la capacité utile et a aussi pour conséquence de réduire et d'optimiser les sauvegardes associées.

3.2.2.3 *Les types de RAID*

Des types de RAID comme le RAID 1 ou 10 sont plus consommateurs de ressources en disques durs, car ils protègent les données en les dupliquant constamment.

Le RAID 6 autorise la panne de 2 disques durs dans un groupe RAID sans perte de données, tout en préservant une performance acceptable, et ne nécessite pas un doublement de la volumétrie pour sa mise en oeuvre. Dès la perte d'un premier disque dur détectée, un

⁷⁴ Quotient ou ratio de déduplication (ou compression) = taille initiale / taille finale

nouveau disque dur peut être ajouté au groupe RAID afin d'éviter qu'un deuxième disque tombe en panne et que tout le groupe de données soit à reconstruire.

Éviter cette indisponibilité de plusieurs heures lors de la reconstruction du groupe RAID est un critère distinctif important pour les activités.

Moins de disques durs dédiés à la protection des données améliorent la capacité du système tout en respectant les mêmes contraintes de productivité et a pour conséquence d'optimiser la consommation énergétique liée.

3.2.2.4 *Les snapshots et clones*

La fonctionnalité de snapshot⁷⁵ consiste à prendre une « photo », un instantané d'un volume logique. Même s'ils enregistrent uniquement les modifications apportées au volume cible, on peut utiliser la fonctionnalité de snapshot quotidiennement comme un premier niveau de sauvegarde « locale » économique en espace et restaurable rapidement. L'avantage est de ne pas utiliser en plus un autre système de sauvegarde pendant la journée et de laisser le dispositif de sauvegarde classique opérer la nuit.

Les développements et tests font partie intégrante des activités du data center et peuvent demander d'importantes duplications de volumes de données pour réaliser leurs simulations. Certaines fonctionnalités permettent de créer des clones dynamiques de volume de données en ne consommant comme espace supplémentaire, que l'écriture des données modifiées et apportent un gain de volumétrie considérable.

3.2.2.5 *Le « Thin Provisioning »*⁷⁶

Dans une architecture serveur-stockage classique, les serveurs ont un espace de stockage alloué, basé sur l'anticipation des besoins des applications. Si l'espace de stockage est saturé par les données issues de l'application, celle-ci ne pourra continuer à fonctionner et entraînera une défaillance applicative.

Les applications requièrent une provision plus ou moins importante en espace de stockage. La conséquence est une énergie dépensée inutilement par des disques durs en fonction dont les volumétries sont sous utilisées.

Le « Thin Provisionning » évite cette surallocation inutile et consommatrice de ressources. Il s'agit de l'allocation « fine » des ressources de stockage à la demande.

L'illustration d'EMC ci-dessous du « Thin Provisionning » expose le principe d'utilisation maximale des espaces de stockage tout en donnant à l'utilisateur la même qualité de service :

⁷⁵ Snapshots : copies instantanées

⁷⁶ « Thin Provisionning » : surallocation ou surprovisionnement

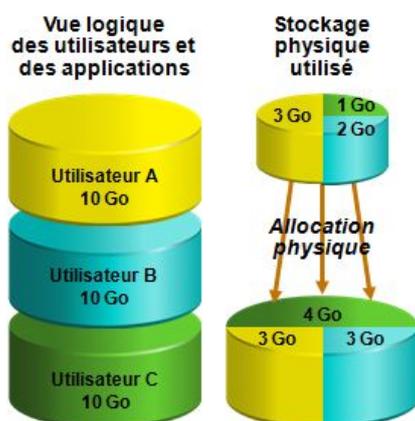


Figure 20. Thin Provisionning, EMC

Le résultat est la diminution notable de la sous-utilisation typique générée par la plupart des applications ou des espaces alloués aux utilisateurs. La fonctionnalité permet de fournir du stockage à la demande (capacity planning) en réduisant la quantité de disques durs nécessaire à l'activité.

3.2.2.6 Les volumes ajustables dynamiquement

Cette approche est complémentaire au Thin Provisionning pour l'optimisation de l'usage des capacités de stockage. Il s'agit de réduire ou d'agrandir dynamiquement un espace de stockage selon le total de données générées par l'application ou les utilisateurs.

Habituellement un administrateur alloue par avance des espaces de stockages aux serveurs et aux applications. Les volumétries sont figées et incluent généralement une réserve, ces espaces de stockage ne sont que rarement utilisés à leur maximum.

La fonctionnalité d'espace ajustable dynamiquement, et son effet sur chaque espace de stockage, génère une optimisation globale significative.

3.2.2.7 Combiner les fonctionnalités logicielles

La combinaison des fonctionnalités logicielles évoquées ci-dessus permet la réduction significative d'un volume de données. L'illustration suivante faite par le SNIA, met en évidence les bénéfices de ces fonctionnalités pour faire face à la problématique de la redondance des copies d'un volume (« Data » dans le schéma) nécessaires à l'activité de production :

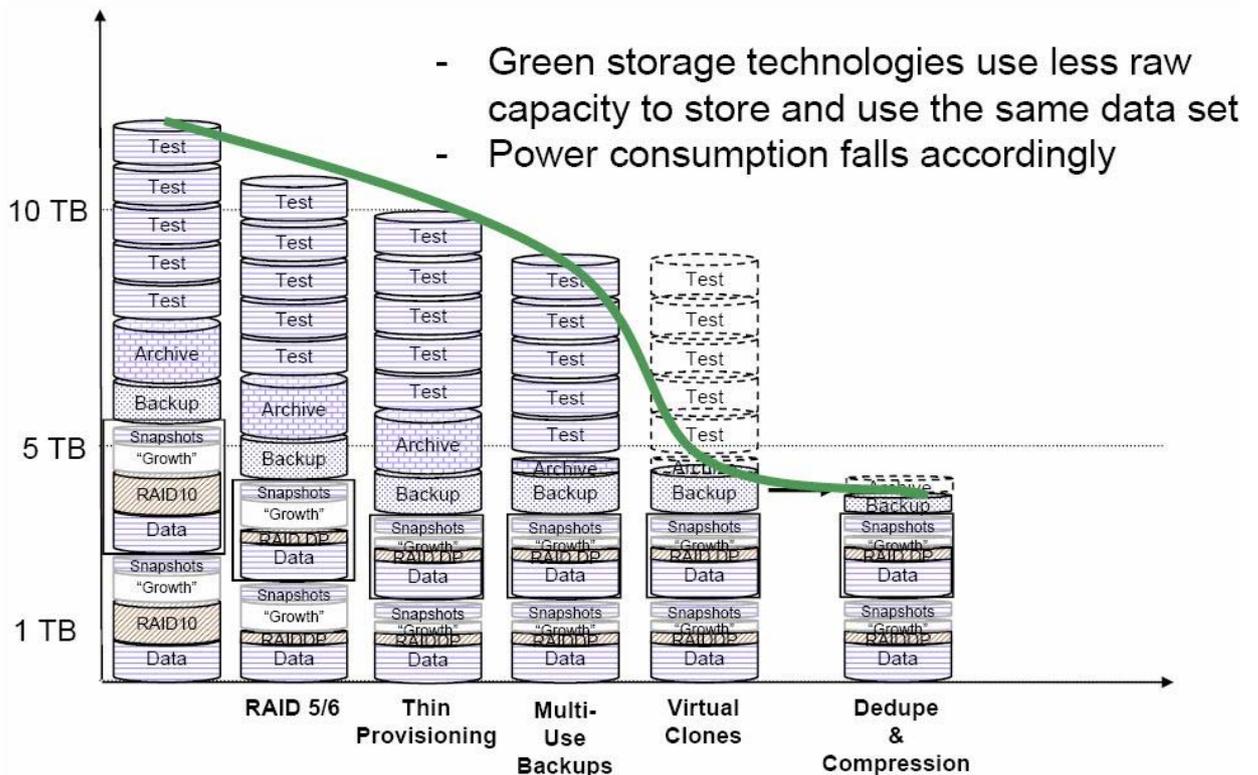


Figure 21. Combinaison des fonctionnalités, SNIA

1/ La première colonne de gauche correspond aux volumétries additionnelles générées par le type de RAID appliqué, à la réplication distante (PRA), à la sauvegarde locale et distante, à l'archivage et enfin aux volumes dupliqués utilisés pour les besoins en tests et développements.

2/ L'application du RAID 5 ou du RAID 6 optimise l'espace nécessaire au stockage du volume de données.

3/ Le « Thin-Provisioning » ajuste optimalement la taille d'un volume de données partagé.

4/ La sauvegarde à usage multiple (« Multiple Use Backups ») permet de distinguer les données à sauvegarder des données à archiver- la capacité de gérer de longues rétentions.

5/ La fonctionnalité de clone « virtuel » (« virtual clones ») permet de ne pas dupliquer inutilement le volume de données initial (« Data ») à des fins de tests et de développements. Le clone virtuel saura créer une copie virtuelle du volume initial, en écrivant uniquement les données modifiées au fur et à mesure.

6/ La déduplication et la compression optimisent finalement l'ensemble du volume de données en éliminant les blocs et les fichiers redondants et en accroissant la densité utile.

➔ La combinaison des fonctionnalités RAID 5 ou 6, snapshots et clones, thin provisioning, sauvegardes à usages multiples, déduplication et compression, peut générer ensemble une optimisation de 50 % de la volumétrie.

3.2.3 Les systèmes disques

3.2.3.1 Contrôleurs et technologies disques

Certains systèmes de stockage sont capables d'utiliser de multiples protocoles comme CIFS, NFS, FCP ou I-SCSI et de combiner différentes technologies de disque en leur sein : SSD, Fibre Channel, SAS, S-ATA. Ces deux caractéristiques combinées permettent de traiter avec un seul système, différents types d'espaces de stockage et d'associer selon la typologie des données et leurs exigences, l'espace au meilleur rapport coût-énergie-performance.

Ainsi, le constructeur Network Appliance (NetApp) permet sur un seul contrôleur d'utiliser les principaux protocoles existants : CIFS (Windows), NFS (Unix, Linux), I-SCSI (SAN sur IP), FCP (SAN sur Fibre Channel) et de combiner différents types de technologies disque (FC, SAS ou S-ATA) dans un même cabinet.

→ *Les contrôleurs multi protocoles et capables d'accueillir différentes technologies de disques, permettent en assurant différents rôles, de réduire le nombre de systèmes de stockage nécessaire à chaque type de donnée. Selon sa valeur, son besoin en performance et accessibilité, chaque donnée est associée à une technologie disque adaptée et d'un point unique d'administration. Cette centralisation facilite l'association des données à l'espace de stockage ayant le meilleur ratio énergie / performance.*

Des centres de données voient le nombre de leurs bibliothèques de bande magnétique s'accroître ainsi que les lecteurs de bandes associés, pour répondre aux problématiques de temps alloués de sauvegarde (« fenêtres ») qui sont souvent réduits par l'extension des créneaux horaires des activités ou saturés par l'augmentation des volumétries.

Les directions métiers imposent non seulement à la DSI des garanties de protection de leurs données de plus en plus fortes, mais désormais imposent aussi des garanties de restauration des données dans des temps de plus en plus courts pour les données classées critiques.

L'emploi de disques durs pour les opérations de sauvegarde se développe, du fait d'un besoin croissant de sauvegarde et de restauration rapide voir quasi-instantanée des données.

L'usage de disques basses consommations et fortement capacitifs est essentiel pour limiter la consommation.

3.2.3.2 Librairie virtuelle de bande - Virtual Tape Library (VTL)

Un dispositif de librairie bande virtuelle (VTL) servant de cache à base de disques durs et se faisant passer pour une librairie de bandes classique, permet d'améliorer les opérations de sauvegarde/restauration en optimisant les flux réseaux et réduisant ainsi éventuellement le nombre de bibliothèques de bande magnétique ou leur nombre de lecteurs nécessaires.

La technologie VTL, utilisant des disques durs, démontre aussi une beaucoup plus grande fiabilité de ses sauvegardes et restaurations, comparée à une librairie de bande magnétique classique.

De plus, certaines bibliothèques virtuelles ou VTL, peuvent supporter un mode de réplique incremental bloc vers un autre système distant, ce qui est particulièrement efficace pour externaliser un volume de données et constituer un premier niveau de Plan de Reprise d'Activité, même avec un réseau intersite modeste.

→ *Le rôle des systèmes de stockages à bases de disques durs s'étend car tout en offrant de meilleures performances, ils sont capables d'améliorer leur consommation énergétique (disques durs basses consommations, MAID), de rationaliser le nombre de systèmes nécessaires : grâce à des contrôleurs multi protocoles, à leur capacité à mixer les types de disques et à optimiser les opérations de sauvegarde –VTL et de Plan de Reprise d'Activité.*

3.2.4 Virtualisation

Introduction sur les avantages de la virtualisation

L'avantage principal de la virtualisation des infrastructures serveurs-stockages est de pouvoir mieux appréhender et répartir les charges d'activités grâce à un « pool » global de ressources. La sous-utilisation générale de ces ressources leur autorise une capacité de traitement et de stockage maximale grâce à la virtualisation.

La virtualisation permet aussi d'associer à une activité normale les ressources appropriées et de pouvoir absorber les pics d'activités en ajoutant dynamiquement de nouvelles ressources si nécessaire. Les ressources non sollicitées peuvent donc être mises en « veille » d'une part et d'autre part, celles qui sont utilisées le sont au meilleur taux d'utilisation possible.

Virtualisation des ressources physiques de stockage

Les fonctionnalités de virtualisation permettent d'ajouter une couche logique sur les infrastructures de stockage physiques afin de gérer un espace de stockage unique, une capacité globale.

Le principal avantage est de pouvoir allouer une partie de cette capacité dynamiquement aux applications et aux utilisateurs, selon leurs besoins et de façon optimale, c'est-à-dire sans allouer aux volumétries des réserves excessives, et en réduisant ou agrandissant les espaces selon les besoins des différentes activités.

Une fois les ressources de stockage virtualisées, les fonctionnalités logicielles déjà évoquées comme le Thin-Provisionning, les snapshots, les clones virtuels, voient leurs effets positifs « d'optimisations des volumétries » se renforcer et se mettre en œuvre plus efficacement car jouant sur des volumétries plus grandes.

L'exemple d'une mutuelle d'assurance

Lors d'un projet de consolidation et de virtualisation des serveurs et du stockage d'une importante mutuelle d'assurance, la nouvelle architecture mise en place a fait preuve de nombreuses améliorations :

1/ Réduction des infrastructures serveurs-stockage dans le centre de données.

Gain d'espace évitant la création d'une nouvelle salle informatique

- 40 serveurs de type Intel remplacés par 2 à 3 serveurs en cluster -haute-disponibilité-, intégrant les dispositifs de virtualisation
- 3 systèmes de stockage non clusterisés, remplacés par un seul, en cluster.
- Un serveur mainframe retiré et ses données migrées vers les nouveaux systèmes.

2/ Diminution de la consommation énergétique grâce à de nouveaux serveurs et systèmes de stockages plus économes, pouvant être mis en veille et adapter leur refroidissement selon l'activité.

3/ Remplacement de l'ancien réseau de stockage par un nouveau SAN -composé de commutateurs Fibre Channel nouvelle génération- apportant **un gain de plus de 50 %** de consommation énergétique pour 3 fois plus de bande passante.

4/ La virtualisation permet de mettre en œuvre plus économiquement un dispositif de plan de reprise ou de continuité d'activité (PRA) par l'utilisation à cet usage du 3e serveur de virtualisation placé dans une deuxième salle distante. Un accès moins coûteux à un PRA est un facteur contribuant à la sécurité des données de l'entreprise et, en cas de sinistre, à la reprise garantie de son activité et donc à son développement durable.

5/ Économies en maintenance et en formation : la consolidation et la virtualisation vers une technologie identique, réduit les besoins en formation et les stocks de pièces détachées nécessaires. L'utilisation systématique de technologies clusterisées et d'architectures haute-disponibilité réduit très fortement le nombre d'interventions de maintenance.

6/ L'amélioration globale de la continuité d'activité, des performances, de la disponibilité des données contribue à la satisfaction des directions métiers. Les systèmes d'informations virtualisés et consolidés sont désormais disponibles entre 99,9 et 99,99 % de l'activité. Cette performance contribue au développement durable de l'entreprise et à sa compétitivité.

L'illustration ci-dessous, expose succinctement les différents bénéfices obtenus dans une architecture consolidée et virtualisée à l'aide des outils VMWare :

	Avant VMware	Après VMware
Serveurs	• 1000	• 50
Stockage	• En attachement direct	• SAN et NAS hiérarchisés
Réseau	• 3000 câbles/ports	• 300 câbles/ports
Installations	• 2000 racks • 4000 modules d'alimentation	• 10 racks • 20 modules d'alimentation



Figure 22. Bénéfice de la virtualisation, VMWare

3.2.5 Réseaux de stockage

Réseau SAN : Storage Area Network

Les réseaux de stockage SAN sont constitués de carte fibre channel (HBA⁷⁷ serveur), de câbles en fibre optique et de commutateurs (switchs) pour interconnecter les serveurs aux systèmes de stockage.

Les dernières générations de SAN améliorent **leur génération de chaleur tout en offrant de nouvelles performances : 4 et 8 Gbits/s**. Une baisse de température interne des composants de l'ordre de 10°C, peut doubler leur durée de vie et permettre l'utilisation

⁷⁷ HBA : Host Bus Adapter, carte d'interface serveur-stockage

d'alimentation plus économique, et ainsi améliorer significativement la fiabilité des équipements. Les dernières technologies SAN (par exemple chez le constructeur Brocade) tendent à répondre à cet objectif, tout en **améliorant leur compacité** avec une forte évolution du nombre de ports par U⁷⁸.

Certains switches SAN autorisent **des configurations minimales** initiales de quelques ports, **extensibles** par ajout de module selon l'évolution des besoins, permettant de meilleurs dimensionnements et évolutivités du réseau de stockage.

Un nombre important de ports fibre channel sur les contrôleurs des baies de stockage peut selon le contexte, apporter une connectibilité suffisante pour s'affranchir de switches SAN.

➔ Les SAN peuvent être consolidés, leur performance et taux d'utilisation améliorés par la virtualisation (fabrique virtuelle, VSAN). L'efficacité énergétique, la compacité et l'évolutivité des dernières générations de commutateurs contribuent à la réduction de la consommation électrique et au désencombrement du data center.

Réseaux NAS : Network Attached Storage

Les systèmes de stockage NAS sont interconnectés avec les clients et serveurs par le biais du réseau Ethernet de l'entreprise. Les dernières générations de cartes Ethernet voient leur performance atteindre des vitesses proches du SAN avec des réseaux de 1 ou 10 Gbits/s.

En effet, ces derniers peuvent aussi porter les flux des nouvelles infrastructures Voix sur IP (VoIP) et se doivent d'offrir des performances « haut débit ».

Il devient opportun d'étudier la possibilité d'utiliser ces nouveaux réseaux pour les flux de données de l'infrastructure de stockage afin d'éviter l'implémentation de nouveaux équipements (SAN ou commutateurs Ethernet dédiés).

➔ *Les infrastructures réseau récentes ont des performances permettant de prendre en charge certains besoins réseau des systèmes de stockage. Cela réduit les équipements nécessaires et contribue à rationaliser la consommation énergétique du centre de données*

3.2.6 Architecture organisée par « Tiers » et « HSM énergétique »

De multiples îlots de stockages cohabitent souvent dans le data center. Cela est dû principalement à la diversité des types de données, de leur usage et de leur valeur pour l'activité.

On distingue 3 catégories principales de données :

- les données transactionnelles généralement hébergées dans un SAN offrant des espaces de stockage performants, disponibles, sécurisés à base de disques SSD, FC ou SAS ;
- les données non structurées généralement hébergées sur des espaces de stockage de type NAS ou DAS, privilégiant principalement l'accessibilité et la disponibilité à base de disques durs FC, SAS et surtout S-ATA ;

⁷⁸ U : le U est une dimension de 1,75 pouce (44,454mm) utilisé comme unité de mesure dans les racks informatiques

- les données archivées et/ou réglementaires généralement hébergées sur des espaces de stockage de type CAS⁷⁹ privilégiant l’immuabilité, l’intégrité, la sécurité surtout à base de disques S-ATA et S-ATA WORM.

Aussi les technologies de stockage disque ont des caractéristiques distinctes :

- Solid State Disk (SSD) : pour les données structurées issues d’applications fortement transactionnelles. Le coût des disques SSD reste très élevé mais ils sont capables de traiter de beaucoup plus grandes quantités d’entrées-sorties et donc de remplacer de nombreux disques durs FC, pour un bénéfice important en performance et en consommation énergétique ;
- Fibre Channel (FC) : pour les données structurées issues d’applications, transactionnelles, souvent accédées, ou critiques. Bon MTBF⁸⁰ mais coût élevé, forte consommation ;
- Serial Attached SCSI (SAS) : pour les données structurées ou non. Performances proches du Fibre Channel, MTBF et consommation moyenne ;
- S-ATA : pour des données en générales non structurées, type fichiers ou peu accédés. Disque au MTBF très moyen, faible coût et consommation.

Il s’agit d’aligner la valeur des données **à la valeur d’acquisition et de fonctionnement** de l’espace de stockage adéquat, par une répartition **en catégorie ou « Tier »** :

- Tier 1 : données les plus critiques, haute disponibilité 24h/24 7J/7 traitement transactionnel élevé (bases de données), sauvegarde intensive ;
- Tier 2 : données non critiques, issues de bases de données transactionnelles n’exigeant pas de disponibilité 24h/24 ou de sauvegarde intensive ;
- Tier 3 : données non critiques et plus rarement accédées, souvent de type non structuré, fichiers, sauvegarde plus espacée.

On peut aussi considérer pour la protection des données des catégories supplémentaires :

- Tier 4 : Sauvegarde et restauration sur site (librairies de bandes ou de disques)
- Tier 5 : Sauvegarde hors site, souvent à des fins de reprise d’activité (librairies de bandes ou de disques)

Les outils **HSM** : « **Hierarchical Storage Management** » permettent de stocker les données de façon hiérarchisée et de les déplacer automatiquement selon des critères de durée, d’accessibilité, de date, ect...

On peut ainsi constituer l’illustration ci-dessous, les flèches pointillées montrant les déplacements dynamiques des données vers différents espaces / tiers de stockage selon plusieurs critères : coût, performance, temps d’accès, donnée active ou inactive, consommations, l’objectif étant de mettre en œuvre un « **HSM⁸¹ énergétique** » :

⁷⁹ Content Adressed Storage :

⁸⁰ MTBF : Mean Time Between Failures – Temps moyen entre panes ou entre deux défaillances

⁸¹ HSM : Hierarchical Storage Management – Stockage de données de façon hiérarchique

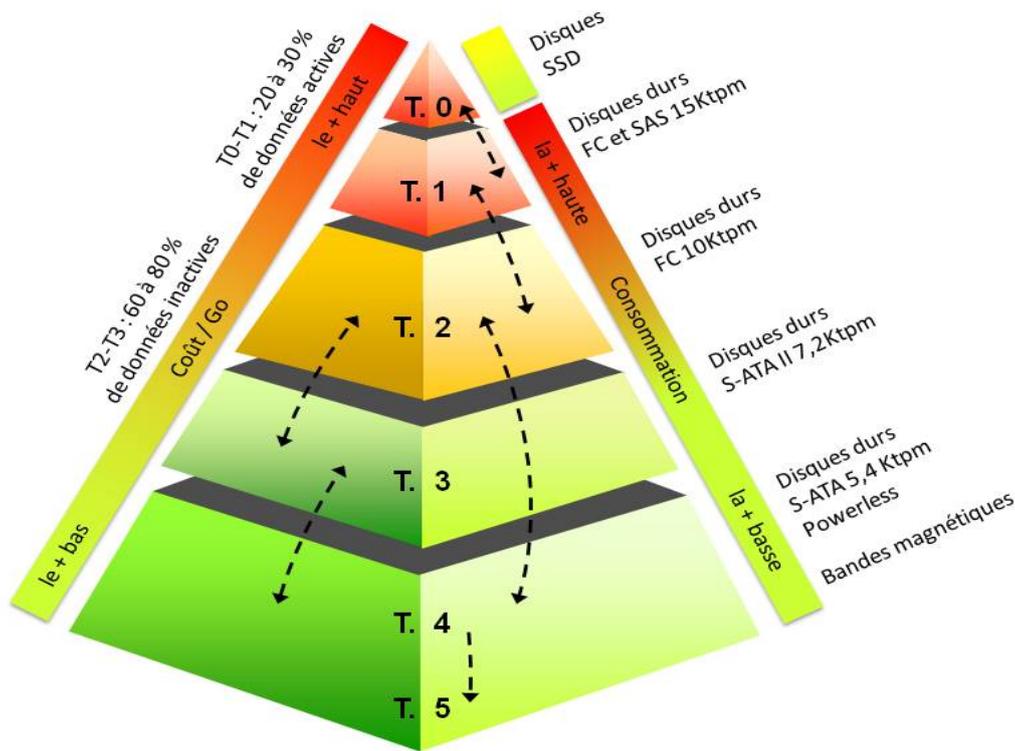


Figure 23. Management hiérarchique et énergétique du stockage, F-Laura

On constate généralement dans les zones de stockage T0 à T1, qu'assez rapidement, près de 20 à 30 % seulement des données sont réellement actives. Pour les espaces T2 à T3 on peut atteindre 60 à 80 % de données inactives. Le déplacement dynamique des données dans les différentes zones de stockage est donc un facteur d'économie substantielle pour les entreprises dont les volumétries de données sont importantes et dont le cycle de vie est changeant.

C'est aussi par conséquent un facteur d'optimisation et de contribution à la réduction de la consommation énergétique, en ne laissant dans les espaces de stockage T0 à T1 -les plus « énergétivores »- que les données essentielles, critiques et actives pour l'entreprise.

Les baies de stockages pouvant mixer différentes technologies de disques durs, plus ou moins énergétivores intègrent alors dans un même cabinet, différents tiers de stockages et facilitent ainsi les opérations de migrations dynamiques des données par des outils HSM, selon les prérogatives de l'administrateur.

➔ *Le stockage organisé par tiers a pour bénéfice de renforcer les procédures de rétention des données afin que celles-ci soient dans l'infrastructure de stockage correspondant à leur valeur. Cette organisation réduit ou canalise la volumétrie de l'espace de production, le plus énergétivore, en déplaçant au maximum les données en fin de cycle de vie, vers les espaces aux consommations énergétiques faibles comme les bandes magnétiques ou les disques durs S-ATA.*

3.2.7 Contrôle du cycle de vie des données

Les données sont conservées indéfiniment si l'entreprise n'a pas fixé de durée de conservation des données. Pire encore, en cas de site de secours, les données peuvent être répliquées (Plan de Reprise d'Activité) et sauvegardées totalement plusieurs fois par mois inutilement.

On peut associer les directions métiers à un audit des applications existantes pour établir un classement des données selon leur importance et le type d'espace de stockage qu'elles occupent. Ce premier diagnostic peut révéler l'inadéquation des données aux ressources.

L'entreprise doit reconnaître que toutes les applications ne sont pas critiques et n'ont pas besoin d'avoir leurs données dans des espaces de stockages ultraperformants et très consommateurs d'énergie.

ILM : Information Lifecycle Management – Gestion du cycle de vie de l'information

L'objectif des outils ILM est de pouvoir évaluer l'importance des données et leur type d'usage afin de les placer sur le support de stockage offrant les meilleurs compromis en termes de coût au Go, de performance (débit, temps d'accès, entrées-sorties) et de consommation énergétique.

Les outils HSM contribuent à faire face à la croissance des données et à l'optimisation des ressources de stockage :

- en agissant **en amont de la création** des données et en identifiant et canalisant les nouveaux flux. Les données sont cartographiées et leur cycle de vie est contrôlé. L'avantage est une meilleure compréhension de l'origine des données (quelles activités ? quelles directions métiers ?) et de leur croissance ;
- en rationalisant l'usage des ressources, notamment en renforçant les contrôles sur les espaces les plus consommateurs énergétiquement comme la zone de stockage de production ou espace primaire ;
- en détectant les données non accédées qui sont dormantes ou persistantes et en les migrant vers l'espace de stockage le plus économique ou en les archivant définitivement.

Le schéma d'EMC ci-dessous illustre le déplacement dynamique des données selon le stade de leur cycle de vie vers différents tiers de stockage :

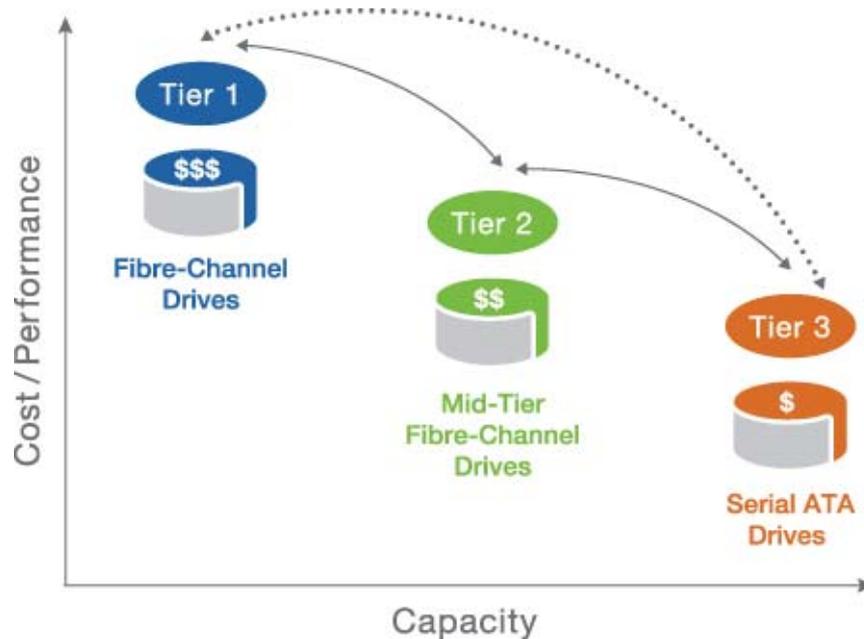


Figure 24. Tiering dynamique, EMC

→ L'association de la hiérarchisation des données selon leur cycle de vie à des espaces de stockages organisés par tiers selon leur performance, canalise l'usage des ressources les plus énergétivores, et contribue à réduire la consommation énergétique dans le data center

3.2.8 Rationalisation des investissements selon la valeur des données

L'enjeu principal de la DSI et de la direction corporate est de considérer le risque de perte de données et le niveau d'investissement adéquat pour y remédier, sans surévaluer le TCO⁸² final de l'infrastructure de stockage.

Paul P. Tallon et Richard Scannell dans leur étude publiée en 2007 : « Information Life Cycle Management », s'appuient sur l'exemple du secteur financier : « Value-At-Risk » (« Var ») pour l'appliquer aux risques de pertes de données. Le principe de « valeur à risque » est de donner une estimation du montant des pertes qui ne devrait jamais être dépassée sur un portefeuille d'actifs ou de fonds de placement.

Ainsi, le « VaR » définit par avance à partir de quel seuil un investissement perdra sa rentabilité et donc quel montant maximum de perte est envisageable pour maintenir le placement, en espérant qu'ensuite il atteindra l'objectif de bénéfice précédemment fixé.

L'adaptation du principe du « VaR » à l'ILM -à la gestion du cycle de vie des données- est d'augmenter ou de diminuer le TCO de l'infrastructure de stockage selon deux facteurs :

- le risque de perte de données pour les activités et la valorisation précise des impacts pour chacune des directions métiers ;
- la valeur des données, qui est changeante et doit donc être ré-ajustée périodiquement.

La valeur des données pouvant rapidement croître ou décroître selon les circonstances des activités de l'entreprise, l'illustration suivante montre la dynamique nécessaire pour ajuster

⁸² Total Cost of Ownership : coût total de possession

les investissements des ressources de stockage à un valeur, régulièrement réactualisée des données :

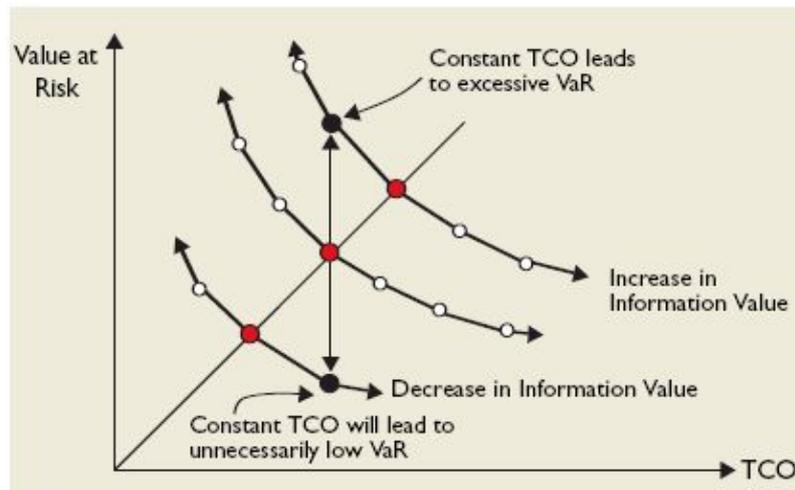


Figure 25. Répondre au changement de valeur des données, source ACM

3.2.9 Mutualisation

Face à la spécialisation des systèmes de stockage et à la multiplicité de leur rôle, il est opportun d'utiliser toute solution capable de regrouper et mutualiser les fonctions suivantes : sauvegarde locale, réplication distante, HSM, archivage, protections dédiées aux applications : Oracle, Exchange, déduplication, etc.

Les principaux bénéfices de la mutualisation sont de :

- réduire le nombre de serveurs et de systèmes de stockage associés à chaque rôle ;
- faciliter et réduire la maintenance, l'administration et leurs coûts par la centralisation ;
- améliorer l'évolutivité générale de la solution et sa capacité d'adaptation aux charges des différentes activités.

Une solution comme celle proposée par l'éditeur CommVault (Bull Calypso) offre une véritable mutualisation en termes de protection et de management des données. De plus, cette solution autorise une mise en œuvre initiale allégée, incluant les fonctionnalités basiques, à laquelle on peut implémenter au gré des besoins, par module, de nouvelles fonctionnalités (ex. : commencer par la sauvegarde, puis le HSM et ensuite le module archivage).

L'illustration suivante met en évidence la mutualisation des fonctionnalités en une architecture unique et commune :

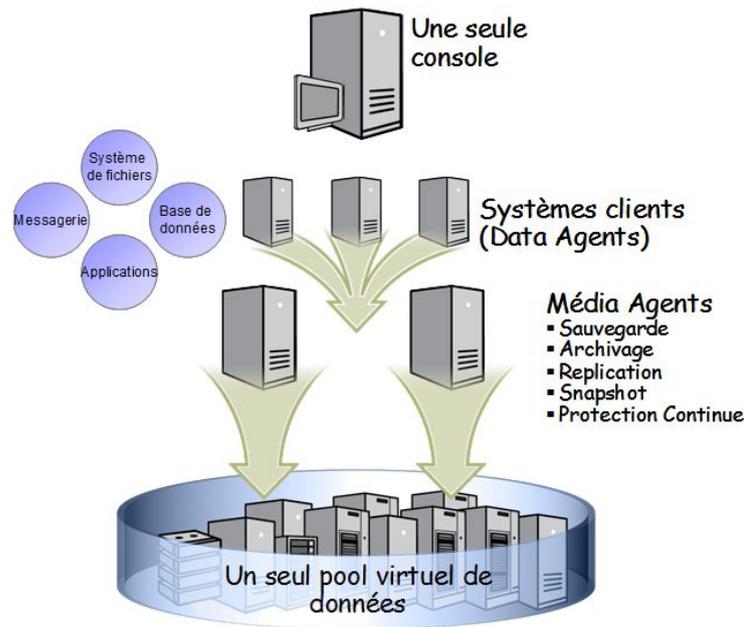


Figure 26. Exemple de mutualisation avec la solution Calypso

→ Les solutions « multifonctionnelles » ont pour effet de consolider différentes fonctionnalités de management des données en une seule plate-forme, réduisant l'encombrement du data center et sa consommation énergétique. La possibilité d'incrémenter les fonctionnalités selon les besoins des activités améliore encore la rationalisation de leur usage.

→ Les mises en oeuvre de la consolidation, de la virtualisation, de la gestion du cycle de vie, de la hiérarchisation, de la mutualisation, améliorent le « juste usage » des ressources, et contribuent à optimiser et rationaliser l'infrastructure de stockage afin de réduire son empreinte dans le data center.

3.3 Canaliser la croissance des données

Introduction

Une infrastructure de stockage plus efficace en termes de consommation d'énergie et d'espace dans le data center est un premier niveau de contribution « green ». Un second niveau concerne les choix d'architectures et de fonctionnalités améliorant leurs capacités et rationalisant leur usage.

Mais face à l'extraordinaire croissance des volumétries et aux nouvelles contraintes de protection et de conservation des données, il est nécessaire d'approfondir un troisième niveau pour canaliser, voire réduire le plus en amont possible, le volume des données des systèmes d'information.

3.3.1 La Gouvernance des données

Ainsi en complément de la gouvernance d'entreprise et de la gouvernance des systèmes d'information, la gouvernance des données est une politique spécifique sur chaque type de données qui constituent son patrimoine informationnel : l'ensemble de ses choix stratégiques, des règles d'usage et des procédures mises en place, compose la « Data Governance ».

Une politique stricte et globale sur les données de l'entreprise

Son rôle est de comprendre et canaliser les flux de données en amont, pour contribuer à l'optimisation et à la productivité des ressources de stockage en aval. L'enjeu étant aussi de rationaliser les demandes et exigences des directions métiers afin d'atteindre les objectifs de développement des activités de l'entreprise tout en utilisant efficacement ses ressources de stockage.

La « Data Governance » peut rationaliser en amont les flux de données et leur infrastructure de stockage dans les systèmes d'information, en agissant sur :

- la « Storage Compliance » : conformité de l'infrastructure ;
- la « Data Compliance » : conformité des données ;
- la « Data Privacy » : respect des données privées ;
- la « Data Protection » : continuité d'activité et développement durable ;
- la « Legacy Data Migration » : migration des données d'anciens systèmes ;
- le « Data Management » : authentification, classification, qualité des données.

3.3.1.1 La « Storage Compliance » - conformité de l'infrastructure

La conformité légale environnementale des infrastructures de stockage

Dans une perspective de développement durable, la Data Governance définit des règles et procédures pour prendre en charge et vérifier l'atteinte des nombreux objectifs de conformité légale comme les normes RoHS, WEEE/DEEE et REACH.

Comme le data center, l'infrastructure de stockage est aussi vieillissante

Il faut détecter et identifier les matériels sensibles et conformément aux réglementations, assurer la collecte, le recyclage et la valorisation des déchets générés par le retrait des systèmes. En France, des sociétés spécialisées comme Ordi Recycling, Screelec, Micronov traitent et valorisent les déchets électroniques.

S'agissant de la valorisation d'un système, plusieurs types de traitements existent comme le réemploi (remise en état initial), le reconditionnement (modification de composants pour améliorer les performances), le démantèlement.

L'entreprise en remettant en état des systèmes pour en faire don à des associations, se donne une image positive et contribue au développement durable.

3.3.1.2 La « Data Compliance » : conformité des données

La «Data Compliance» est l'ensemble des règles et procédures intégré à la gouvernance des données permettant à l'entreprise de mettre en conformité ses données et les pratiques d'utilisation, de protection et de conservation associées, avec les normes et réglementations légales dont la liste ne cesse de s'allonger, par exemple :

- Sarbarne Oxley (SOX) : section 404, Management Assesment of Internal Controls ;
- protocole de Bale II, loi 2004-575 dite de sécurité dans l'économie numérique, loi 2003-706 dite de sécurité financière, recommandations B.C.E.⁸³ ;
- DCSSI : Direction Générale de la Sécurité des Systèmes d'Information. Directives n°485,495,1223,4201/SG et recommandations n°400, 600 et 901 ;
- HIPAA (Etats-Unis, santé), ect.

Conformités aux lois « financières » et « sectorielles »

L'enjeu de la « Conformité des données » est de définir une transparence suffisante et donc une organisation permettant l'accès aux données de façon auditable en définissant préalablement les données qui feront partie du périmètre.

Dans le cas des contraintes légales telles que SOX, Bâle II ou HIPAA (secteur santé), l'objectif majeur pour l'entreprise est d'avoir préalablement identifié les données critiques, enjeux des audits de contrôle, afin de démontrer sa capacité à les sauvegarder, à les restaurer ou simplement à les conserver sur une période donnée et sans dépassement de la période de rétention fixée.

Les audits de conformité sont donc généralement réalisés dans le but d'obtenir la preuve du respect des processus et d'atteinte des objectifs évoqués ci-dessous et que nous pouvons lister dans l'ordre du process : identification des données critiques, identification de la durée de rétention, actions à accomplir à l'échéance de la durée de rétention.

La « Data Compliance » renforce les procédures de classement, d'inventaire et d'identification contribuant à comprendre et limiter les flux de données.

➔ La « Data Compliance » identifie les réglementations sur les données auxquelles doit répondre l'entreprise. Elle définit les moyens et les pratiques nécessaires pour répondre aux exigences de conformité. En renforçant les procédures d'inventaire, de classement et d'identification des données, cela contribue à comprendre et maîtriser leurs flux.

3.3.1.3 La « Data Privacy » - respect des données privées

Autant aux États-Unis qu'en Europe, la législation sur les données privées se développe et particulièrement en France avec la **CNIL**⁸⁴. Elle impose un cadre strict réglementaire touchant à différentes facettes du traitement des données :

- collecte à caractère personnel : si cela identifie une personne, il faut son consentement ;
- finalité des traitements : chaque fichier doit avoir un objectif précis. Les informations exploitées doivent être cohérentes avec celui-ci ;
- durée de conservation : les données personnelles ont une durée de péremption qui fait l'objet d'une déclaration initiale ;
- sécurité des fichiers : « tout responsable de traitement informatique de données personnelles doit adopter des mesures de sécurité physique (locaux) et logiques (systèmes d'information) adaptées à la nature des données et aux risques présentés par leur traitement » ;
- confidentialité des données : seuls les destinataires explicitement désignés et les tiers autorisés (police, fisc) peuvent accéder aux données personnelles contenues dans un fichier.

⁸³ B.C.E. : Banque Centrale Européenne

⁸⁴ C.N.I.L. Commission Nationale Informatique et Liberté

- déclaration des fichiers, etc.

Le non-respect d'une de ces obligations est passible de 5 ans d'emprisonnement et de 300 000 euros d'amende.

L'entreprise peut adopter une charte interne donnant des obligations à l'employé (et à elle-même) sur ses données créées au quotidien :

- mise à disposition d'un espace de stockage personnalisé dont la taille est limitée ; cet espace peut exclure tout format de fichier audio ou vidéo par exemple ;
- mise à disposition d'un espace de stockage pour les données critiques, des fichiers archivés de courriers électroniques, afin d'identifier par le volontariat les données capitales à sauvegarder (de nombreux cadres laissent sur leur ordinateur portable des données qui sont critiques pour l'entreprise, sans protection particulière ni sauvegarde).

➔ La « Data Privacy » consiste aussi à mettre en place des règles et procédures pour identifier les données sensibles et constituer une politique d'accès renforcée et confidentielle à ces données. Connaître et appliquer les obligations légales sur les fichiers de personnes permet de canaliser le développement des bases de données et leur finalité. L'enjeu est aussi de responsabiliser et orienter l'utilisateur pour stocker et sauvegarder ses données les plus sensibles.

3.3.1.4 La « Data Protection » : continuité d'activité et développement durable

Le rôle de la Data Governance est d'évaluer les risques encourus en cas de perte ou de non-restauration des données dans des délais acceptables pour les directions métiers. Elle doit identifier et contrôler de façon concertée la nature des données à :

- sauvegarder et notamment la fréquence nécessaire selon leur valeur ;
- archiver selon les obligations légales et les enjeux de l'entreprise ;
- répliquer vers un plan de secours selon leur valeur pour l'activité.

➔ L'implication d'une réflexion globale en amont sur la protection des données et les risques associés permet de qualifier les données sensibles et non sensibles et d'optimiser l'ensemble des flux liés aux opérations de protection des données.

3.3.1.5 La « Legacy Data Migration » - Migration de données des anciens systèmes

Lors d'évolution du data center ou après des opérations de consolidation et virtualisation, d'anciens mainframes ou serveurs peuvent être retirés. C'est alors que de nombreuses données se « révèlent », et la légitimité de leur survie doit être contrôlée avant toute migration.

Le bénéfice est bien entendu de migrer les données éligibles de tout ancien système souvent gourmand en énergie et en espace, vers de nouveaux systèmes de stockage consolidés plus efficaces énergétiquement.

➔ La Data Governance doit être active sur les opérations de migration des données. En effet, elles favorisent l'inventaire et l'identification des données non accédées ou dont le temps de conservation est périmé. Ces données peuvent faire l'objet d'un archivage sur des supports de stockage plus « économiques » ou d'une suppression définitive.

3.3.1.6 Le « Data Management » - Contrôle et qualité des données

Canaliser les volumes de données

L'enjeu est en premier lieu de changer et améliorer les comportements liés au stockage des données.

Les points suivants doivent ainsi être contrôlés :

- toutes les données conservées sont-elles périodiquement vérifiées comme encore utiles aux activités ?
- les directions métiers sont-elles récompensées selon leur effort à supprimer leurs données obsolètes ?
- les utilisateurs sont-ils formés et sensibilisés au management de leurs données ? Suppriment-ils aussi régulièrement leurs données obsolètes ou dupliquées ?

Localiser et mesurer les données dormantes

De nouvelles mesures comme celles qui sont proposées par l'Uptime Institute⁸⁵, permettent de mettre en évidence les améliorations faites sur l'usage du stockage des données, notamment pour les données dites dormantes :

$$DH-UR = \frac{\text{Number of terabytes of storage holding important, frequently accessed data (within the last 90 days)}}{\text{Total Terabytes of storage actually deployed}}$$

DH-UR signifie "Deployed Hardware Utilisation Ratio", le ratio d'utilisation des équipements déployés. Le résultat obtenu en divisant le total des données accédées dans les 90 jours et la volumétrie totale déployée est un premier indicateur d'efficacité du Management des données.

Gérer le cycle de vie des données

Les données ont besoin de Data Governance à tous les âges et stades de leur usage pour pouvoir passer d'un statut « actif » à un statut « d'archive » avec tous les contrôles nécessaires pour autoriser le passage d'un stade à l'autre.

Les données non structurées deviennent rapidement fixes et persistantes. Leur indexation est nécessaire (outil HSM, ILM) avec l'ajout de « métadonnées » caractérisant par exemple, leur statut, leur durée de rétention, leur criticité et leur propriétaire.

Le bénéfice est de réduire considérablement des temps de recherche (très coûteux et peu efficaces sur les données stockées « en vrac »), en filtrant régulièrement les informations des métadonnées, l'espace de stockage s'épure des données à archiver ou à effacer.

Renforcer les contrôles d'accès

Renforcer l'identification des utilisateurs et les autorisations d'utilisations en lecture et écriture des données a pour effet de maîtriser l'usage d'un espace de stockage. L'objectif étant à nouveau de canaliser les créations de données vers les différentes catégories de stockage les mieux adaptées, selon la typologie des données et leur criticité.

⁸⁵ [http://uptimeinstitute.org/wp_pdf/\(TUI3009F\)FourMetricsDefineData_center.pdf](http://uptimeinstitute.org/wp_pdf/(TUI3009F)FourMetricsDefineData_center.pdf)

➔ *Des fichiers bureautiques, des sauvegardes, de nombreuses données non transactionnelles, ne doivent pas se trouver dans la zone de stockage la plus performante et la moins efficiente énergétiquement. De même, des données non accédées depuis un an pourraient sûrement être définitivement déplacés, archivés ou effacés selon les directives strictes de la Data Governance.*

3.3.2 Une « Green Data Governance »

Une gouvernance appliquée en amont à canaliser les anciens et nouveaux flux de données est contributrice d'un meilleur usage et optimisation possible des ressources de stockage.

La « Data Governance » peut donc intégrer plusieurs facettes, plus spécifiquement « vertes », essentielles pour réussir une politique globale « Green Storage » :

- la « Green Storage Compliance » : recenser la conformité des infrastructures de stockage avec les réglementations comme RoHS, WEEE et développer les procédures de traitement et de recyclage des matériels usagés ;
- la « Green Data Compliance » : assurer la conformité légale des données, maîtriser les durées de péremption, migrer les données sensibles vers le support adéquat et savoir supprimer définitivement les données qui l'exigent ;
- la « Green Data Privacy » : renforcer les règles d'appartenances, d'accès, d'usage et de conservation des données ; savoir respecter les obligations de la CNIL, vérifier les durées de conservation, supprimer les données sensibles, contribuer au respect social humain (« social equity ») et donc au développement durable de l'entreprise ;
- la « Green Data Protection » : définir une politique globale de protection, de sauvegarde, et réplique pour le plan de continuité d'activité. ; évaluer les risques, qualifier les données éligibles pour chacune des protections, faire l'arbitrage entre les enjeux des directions métiers et les moyens possibles mis à disposition par la DSI ; garantir la continuité d'activité de l'entreprise est vitale pour son développement durable ;
- le « Green Data Management » : vérifier l'alignement des technologies les plus adaptées à la nature des données et aux exigences des activités métiers ; mettre en place des dispositifs et règles pour gérer le cycle de vie des données, vérifier que la valeur de chaque donnée correspond à un espace de stockage adapté tout au long de son existence ; améliorer la qualité des données, en renforçant leur définition (référentiel, dictionnaires, MasterDataManagement) et en les « filtrant » régulièrement. ;
- La « Green Legacy Data Migration » : profiter des évolutions et des migrations pour contrôler les données et valider celles qui méritent de migrer ; recanaliser au passage les volumétries, les consolider sur une infrastructure de stockage unique pour gagner de l'espace dans le data center et réduire la consommation énergétique globale.

Les meilleurs choix identifiés concernant les matériels, leurs architectures et leurs fonctionnalités auront peu d'effet sur l'optimisation et la rationalisation de l'infrastructure de stockage si l'entreprise ne canalise pas l'extraordinaire croissance des données.

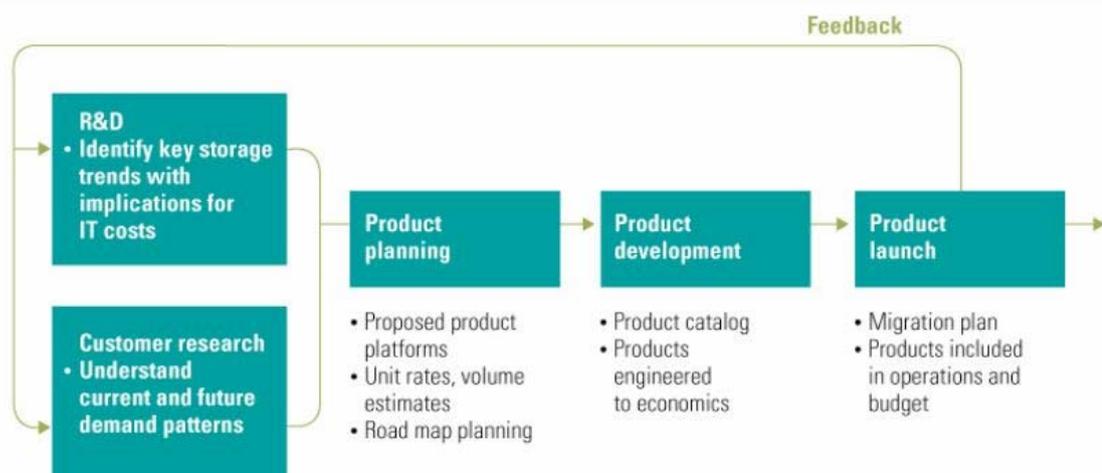
Ces « E-déchets » polluent considérablement les activités des directions métiers. La gouvernance des données peut en assumer le traitement, le recyclage ou la suppression, en accroître globalement la qualité et la valeur.

→ La gouvernance des données doit être considérée et implémentée au plus haut niveau de l'entreprise, pour maîtriser en amont et transversalement, leur prolifération dans toute l'infrastructure de stockage et réduire ses « déchets », que sont les nombreuses données dormantes et persistantes. Après le stade de l'optimisation, doit venir celui de la rationalisation.

3.3.3 Optimiser les besoins des directions métiers

La direction des systèmes d'information se doit de proposer un large choix de différentes configurations de stockage adapté à chaque spécificité des directions métiers de l'entreprise. Elle doit détecter les technologies les plus innovantes et contributrices d'efficacités spatiales et énergétiques tout en répondant aux besoins dus à la croissance et à la productivité.

La difficulté est d'organiser une dynamique de la veille technologique nécessaire, à la sélection des produits possibles, jusqu'aux lancements des solutions définitives. Ainsi, la société d'études et d'analyses McKinsey, propose dans l'illustration ci-dessous, un processus d'implémentation des offres de stockage, le plus largement communiqué entre la DSI et ses clients internes :



Source: IDC; McKinsey analysis

Figure 27. Ex. développement produit, IDC & McKinsey

Les principales phases de l'approche « product-development »⁸⁶ pour le stockage selon McKinsey, sont les suivantes :

- une veille technologique consistant à identifier les tendances des offres de stockage et leur implication sur les coûts de la DSI, complétée par une recherche permanente de la compréhension des besoins actuels et futurs des clients internes ;
- une planification des produits : élaboration des plates-formes proposées, des coûts unitaires, estimation des volumes et planning de la roadmap ;
- le développement des produits : mise au catalogue, incluant prix et contraintes ;

⁸⁶ The McKinsey Quarterly : « Meeting the demand for data storage »
http://www.mckinseyquarterly.com/Meeting_the_demand_for_data_storage_2153

- le lancement des produits : plan de migration, produits inclus dans le budget et les opérations.

3.3.4 Un catalogue de solutions de stockage orienté « Niveaux de Service »

ITIL⁸⁷ recense les bonnes pratiques pour le management des services informatiques, notamment l'amélioration de l'efficacité et de la qualité des systèmes d'information. L'illustration suivante de l'ITSMF⁸⁸ (ITIL France) permet de comprendre l'évolution attendue en termes de maturité de services :

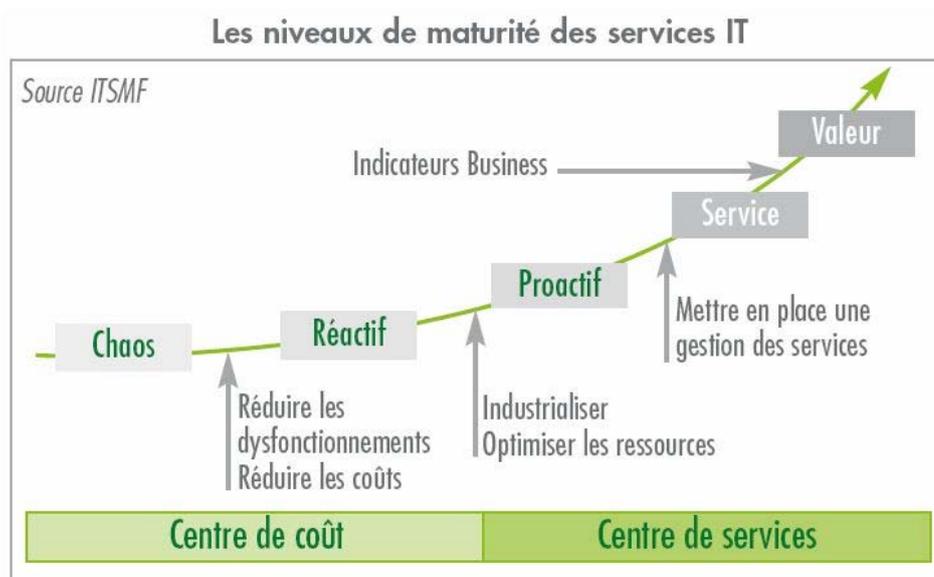


Figure 28. Niveaux de maturité des services de la DSI, ITSMF

La Direction des Systèmes d'Informations doit évoluer d'un centre de coût à un centre de services, qui tient compte des contraintes et des exigences des directions métiers. La DSI développe la valeur de son service et de l'activité métier en ajoutant des services personnalisés aux infrastructures qu'elle propose.

Améliorer la satisfaction des directions métiers

La société McKinsey illustre à nouveau efficacement un exemple de catalogue créé et mis à disposition par la DSI en concertation avec les directions métiers et compréhensible par ces dernières :

⁸⁷ Information Technology Infrastructure Library

⁸⁸ <http://www.itsmf.fr>

Service-level catalog

Disguised example of global medical-products company

Storage tier ¹	Risk/performance profile	Unit cost, \$ per gigabyte of business storage	Sample metrics			
			Performance metrics		Risk	
			I/O transfer rate, megabytes per second	Seek time (minutes)	Monthly full backups	Reliability ²
Platinum SAN • SAN array with FC drives • Mirrored, remote replication; business continuity volume		44	200	3.5	4	Five 9s
Silver SAN • Low-end SAN array with FC drives • RAID 6		14	200	3.5	4	Four 9s

Figure 29. Catalogue de stockage, 1ere partie, IDC & McKinsey

Un classement par « Storage tier » (catégorie d'espace de stockage), détermine une zone de stockage constituée par une infrastructure matérielle, logicielle et de services, caractérisée dans un premier temps selon les deux premiers critères fondamentaux suivants constituant une matrice risque / performance :

- risque de perte de données ou de continuité d'activité ;
- performance d'utilisation.

Les critères suivants précisent pour le même espace de stockage :

- le coût unitaire au giga-octet ;
- le débit de transfert en méga-octet par seconde ;
- le nombre de sauvegardes mensuelles (backup) complètes ;
- la fiabilité en terme de continuité de service (« reliability » : five 9s = 99,999% soit moins de 5 minutes d'arrêt par an).

<p>Silver SAN</p> <ul style="list-style-type: none"> • Low-end SAN array with FC drives • RAID 6 		14	200	3.5	4	Four 9s
<p>NAS</p> <ul style="list-style-type: none"> • NAS array with SATA drives • RAID 5 		9	150	8.5	4	Four 9s
<p>Near-line archive</p> <ul style="list-style-type: none"> • Low-end NAS array with SATA drives • RAID 6 		5	150	8.5	1	Three 9s

¹SAN = storage-area network; NAS = network-attached storage; FC = Fibre Channel; I/O = Input/Output; RAID = Redundant Arrays of Independent Disks; SATA = Serial Advanced Technology Attachment.
²Five 9s means 99.999% of time, or ~5 minutes downtime per year; four 9s means 99.99%, or ~52 minutes downtime; three 9s means 99.9%, or ~9 hours downtime.
 Source: IDC; McKinsey analysis

Figure 30. Catalogue de stockage, 2ème partie, IDC & McKinsey

Les 5 « tiers » de stockage proposés dans ce catalogue vont ainsi répondre explicitement et clairement à des besoins distincts d'activités, par exemple :

- un espace SAN « platinum » hautes performances et disponibilités, fréquemment sauvegardé, pour les données transactionnelles ou critiques ;
- un espace NAS aux performances et disponibilités moyennes avec un temps d'arrêt probable plus élevé au coût au giga-octet 5 fois moins élevé que l'espace précédent ;
- un espace « Near-line archive » : une zone d'archive sur disques durs, aux performances et disponibilités les plus basses, sauvegardée une seule fois par mois.

Ce catalogue d'espaces de stockage pourrait être enrichi d'autres critères facilement compréhensibles pour aider les directions métiers à faire leur choix comme :

- le temps maximal de restauration des données et de reprise d'activité ;
- une volumétrie utile calculée pour une durée et une activité propre à l'activité métier et non pas une volumétrie utile non détaillée. Exemple : 500 Go représentent 6 mois d'activités de la base commerciale et de ses actions promotionnelles.

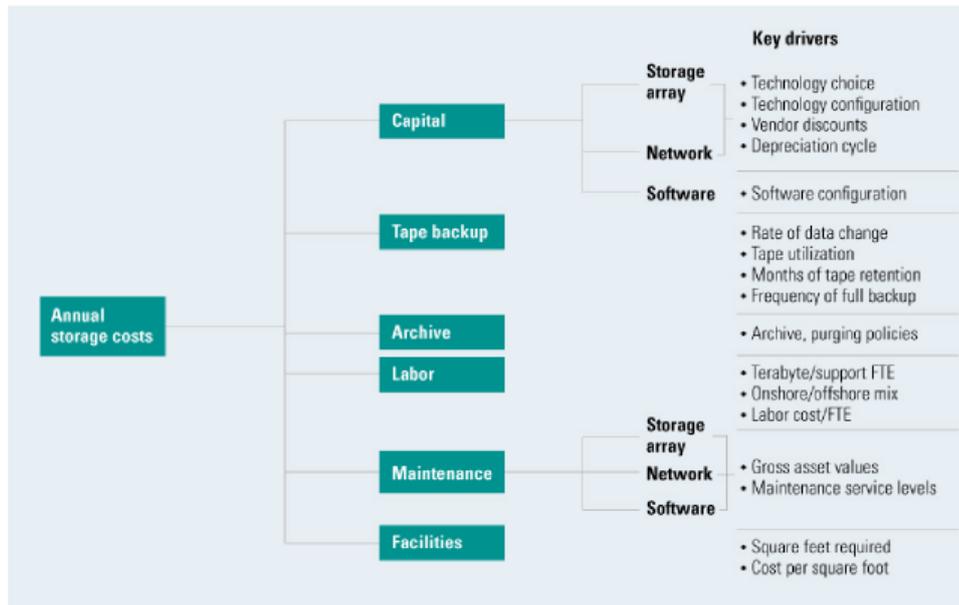
Améliorer la compréhension des services de stockage

L'objectif est de mieux communiquer et faire comprendre aux clients internes, les coûts et les configurations associés au catalogue, avec plus de transparence, pour un meilleur usage des ressources de l'entreprise.

Ainsi, l'illustration suivante de Mckinsey détaille quelques éléments fondamentaux pour un espace de stockage comme les investissements, les opérations de sauvegarde, la période de rétention et la maintenance associés de façon schématique :

Illustrative example

Application	Business unit	Volume, terabytes ¹	Service level ²
#1	A	35	SAN Gold
#2	B	20	SAN Gold
#3	C	5	NAS Silver
•	•	•	•
•	•	•	•



¹Terabyte = 1 trillion bytes.

²SAN = storage-area network; NAS = network-attached storage.

³FTE = full-time equivalent.

Figure 31. Extrait de catalogue de service de stockage, IDC & McKinsey

➔ LA DSI peut mettre en place un catalogue de « services de stockage » pour communiquer et exposer aux Directions Métiers les solutions les plus optimales tout en rationalisant leurs besoins.

3.4 Audits, outils et programme d'amélioration

3.4.1 Audit de l'infrastructure électrique et de la climatisation

Des services d'audits énergétiques permettent d'évaluer l'impact environnemental du bâtiment et de la salle concernés, et ensuite de prendre également du recul sur la situation multidimensionnelle du centre de données : à savoir la qualité de la distribution énergétique, de l'infrastructure de refroidissement, de l'impact des systèmes d'informations et leurs réseaux, et enfin du design d'implémentation de l'ensemble.

Un expert de la climatisation et de l'alimentation électrique comme APC Schneider Electric⁸⁹, dispose de logiciel d'analyse de la consommation thermique et énergétique du data center afin d'en déterminer l'efficacité. La température du data center est-elle trop froide ? La climatisation surévaluée ? En mesurant la performance des équipements existants, il peut se révéler qu'une désactivation d'une partie de la climatisation soit possible.

→La première étape peut consister à analyser les points de faiblesses du rendement du data center en commençant par faire un bilan énergétique et environnemental du bâtiment, de la salle informatique, jusqu'aux infrastructures serveurs-stockages-réseaux.

3.4.2 Audit spécifique et approfondi de l'infrastructure de stockage

Des offres d'audits spécifiques aux infrastructures de stockage sont disponibles, approfondissant et vérifiant de multiples critères d'efficacités. Ainsi la société Glasshouse⁹⁰ analyse de nombreux aspects, tant sur le plan technologique, qu'opérationnel et organisationnel. Par exemple, voici quelques points abordés lors d'un audit :

- le niveau de ressource de chaque activité et le taux d'utilisation des systèmes ;
- les politiques de sauvegarde, de restauration et de réplication adaptées aux activités ;
- la qualité de service associée à chaque typologie de client interne ;
- la politique d'accessibilité, de disponibilité et de plan de secours ;

L'objectif étant ensuite d'optimiser et de rationaliser l'infrastructure de stockage selon les réels besoins de l'entreprise et des activités métiers.

Les résultats de l'audit pourront amener ensuite une réflexion sur les objectifs d'améliorations possibles. Ainsi, par exemple, pour l'usage d'un système de stockage, les objectifs suivants seront étudiés :

- la quantité de volumétrie récupérable : en optimisant la pré-allocation, en identifiant les espaces « abandonnés » et non utilisés ;
- la possibilité de migrer des volumétries des applications hors production ou de production non critiques, vers les espaces de stockage les plus économes ;
- le potentiel de réduction ou de migration vers des espaces plus économes des volumétries liées aux réplications ;
- le potentiel de déduplication, de retrait des données inutilisées.

3.4.3 Outils de gestion de l'efficacité du data center

D'autres outils permettent une gestion active de la consommation électrique. Ainsi l'éditeur Cassatt⁹¹ offre la possibilité de prendre le contrôle sur les serveurs et de les activer selon la charge ou les besoins des directions métiers. Il est ainsi possible d'agir de façon dynamique sur le fonctionnement de l'ensemble de l'infrastructure : serveur-stockage-réseau.

L'administrateur du data center peut décider :

- d'exécuter les travaux de batch ou les applications non critiques aux heures les plus favorables en terme de coût électrique

⁸⁹ <http://apc.com>

⁹⁰ <http://glasshouse.com/Storage-153.do>

⁹¹ <http://www.cassatt.com>

- de déplacer et d'allouer par les techniques de virtualisation, les serveurs selon les charges d'activité et les demandes de la production
- de démarrer des serveurs supplémentaires selon des événements spécifiques, un horaire ou une capacité particulière

3.4.4 Programme d'amélioration basé sur le volontariat

Adhérer aux programmes basés sur le volontariat permet d'accéder facilement aux experts du domaine énergétique et d'intégrer avec leur assistance leurs meilleures pratiques.

Le Code de (bonne) Conduite européen du data center est basé sur le volontariat et couvre deux domaines principaux : l'efficacité énergétique des équipements I.T. et celle des infrastructures de climatisation, d'énergies (onduleurs, alimentations, ventilateurs...).

La première obligation est de mettre en place une procédure régulière de reporting énergétique afin de communiquer à la commission les résultats obtenus. La deuxième peut être un engagement de conformité au C.o.C.⁹² sur au moins 40 % de l'espace du data center.

L'enjeu du C.o.C. est de réduire l'énergie consommée sans impacter la performance d'activité. Pour cela, il optimise et rationalise de nombreuses facettes des systèmes d'information par **une démarche complète** :

- nomination d'une équipe responsable de l'initiative, audit d'utilisation des équipements ;
- contrôle du niveau de disponibilité alloué par système (limiter la haute disponibilité) ;
- définition de différents niveaux de disponibilité dans le data center ;
- permettre de contrôler selon la charge d'activité, les dispositifs d'approvisionnements électriques et de refroidissement ;
- **nouveaux équipements IT**⁹³ : valorisation de critères d'achats liés à l'efficacité énergétique et aux capacités de fonctionnement sur des plages de température et d'humidité plus étendues ; prise en compte des standards définis par l'ASHRAE⁹⁴ ; priorité aux systèmes capables de mesurer, reporter leur consommation et de varier leur performance selon la charge d'activité ;
- **nouveaux services liés aux architectures IT** : déployer les technologies Grid et virtualisation. Justifier, voir réduire les dispositifs de haute disponibilité ou de réplication dans le plan de reprise d'activité ; éviter les architectures serveur actives/passives ;
- **gérer les équipements et services IT existants** : commencer par un audit exhaustif et considérer l'implémentation d'une CMDB⁹⁵ et d'un catalogue de service ITIL ; détecter les services inutilisés et retirer les équipements liés ; virtualiser si possible les serveurs peu utilisés, mais contraints par des dispositifs de conformité légale ; retirer les équipements à faible valeur pour l'activité ; éteindre les équipements non utilisés ; considérer les logiciels capables d'analyser et d'organiser la répartition des charges entre les systèmes ;
- **gestion des données** : règles sur les durées de conservation et de niveau de protection des données ; organiser les utilisateurs avec des espaces de données classées selon les durées de rétention et de protection ; implémenter un guide d'utilisation du stockage par « tiers »

⁹² C.o.C : Code of Conduct

⁹³ I.T. : Information Technology

⁹⁴ ASHRAE : American Society of Heating, Refrigerating, and Air-conditioning Engineers

⁹⁵ CMDB : Configuration Management Data Base

- incluant des SLA⁹⁶ associant performance et disponibilité ; sélectionner le stockage le moins énergétivore tout en tenant compte des capacités de traitement et de restauration des activités ; réduire le volume total des données, déployer l'identification des données pour réduire le volume des données (exemple de « users clean up days ») ; utiliser les fonctions d'optimisation de volumétrie : snapshot, « thin provisioning », compression ;
- **refroidissement** : design d'implémentation des dispositifs air chaud -air froid ; management des chemins d'aérations ; alternance aile chaude et froide ;
 - management du refroidissement : adaptation des dispositifs aux changements de chaleur, coupure des équipements non utilisés, maintenance régulière ;
 - **déployer des indicateurs et métriques** : consommation énergétique (globale et équipements IT), mesures de température et humidité sur l'essentiel des composants alimentation et rack ;
 - **reporter** régulièrement les consommations et générer un rapport environnemental ; inclure les taux d'utilisation pour chaque activité des composants serveur-réseau-stockage ;
 - et sans les développer, d'autres thèmes sensibles sont concrètement détaillés dans le guide des meilleures pratiques :
 - niveau de température et humidité, management de la climatisation ;
 - implémentation et design d'une climatisation hautement efficace ;
 - air conditionné de la salle informatique, réutilisation de la chaleur ;
 - équipement et approvisionnement électrique ;
 - bâtiment, économies d'énergie, choix géographique.

→ *Le Code de Conduite du data center européen ne concerne pas que l'amélioration des dispositifs de réduction énergétique et de climatisation. Il identifie aussi, de l'infrastructure des serveurs, des réseaux jusqu'aux éléments de stockages, les principales actions d'optimisation et de rationalisation de leurs composants, architectures et management.*

→ *Pour être plus efficace et facilité, l'optimisation et la rationalisation de l'infrastructure de stockage peuvent être accompagnées et aidées par des experts, proposant des services d'audits complets, des outils de gestion dynamique du data center, mais aussi des programmes d'améliorations librement accessibles comme le code de bonne conduite européen.*

3.5 Exemple de Balanced Scorecard “Green” (BSC)

Kaplan et Norton sont les inventeurs de la “Balanced Scorecard”⁹⁷, leurs « *Tableaux équilibrés ou prospectifs* » permettent de traduire la mission et la stratégie d'une entreprise en actions et objectifs opérationnels concrets.

Cette vision structurée intègre des performances mesurables sur différentes perspectives. Elle permet de contrôler le déploiement de la stratégie sur les axes complémentaires suivants : interne et externe de l'entreprise, financier, clients-fournisseurs-employés, processus, savoir-faire et apprentissage.

⁹⁶ SLA : Service Level Agreement – Contrat à niveau de service

⁹⁷ R.S. Kaplan et D.P. Norton : « The Balanced Scorecard : Measures That Drive Performance »

Les 4 perspectives essentielles de la « BSC » sont : « financière » (gains, performances), « client » (satisfaction), « processus interne » (optimisation, révision), « connaissance et apprentissage ».

Dans l'étude « *Proactive Corporate Green Strategic Management Using Balanced Scorecard* »⁹⁸, Mlle Wen-Yi Huang a sélectionné 40 mesures clés pour des industries électroniques (Ericsson, Toshiba, Matsushita...) travaillant et ayant des contraintes liées à la norme orientée protection de l'environnement ISO 14000.

La BSC ci-dessous, extraite de l'étude, montre l'articulation de la stratégie Green sur les 4 perspectives :

- l'axe financier inclut des mesures de coûts : les processus, l'amélioration des produits et des technologies, le traitement de la pollution, ainsi que les revenus et profits ;
- l'axe client intègre des mesures pour rendre les produits plus respectueux de l'environnement, atteindre une satisfaction « verte », ainsi que la reconnaissance par l'obtention de la norme environnementale ;
- l'axe interne indique les améliorations des processus, des innovations « vertes » et de l'application du programme DfE⁹⁹ (programme du gouvernement américain pour réduire les risques vers les personnes et l'environnement dans les industries manipulant des produits chimiques) ainsi que des mesures de monitoring continu ;
- l'axe connaissance et apprentissage prend en compte la formation, les savoir-faire des employés, l'acquisition des informations, jusqu'au département R&D.

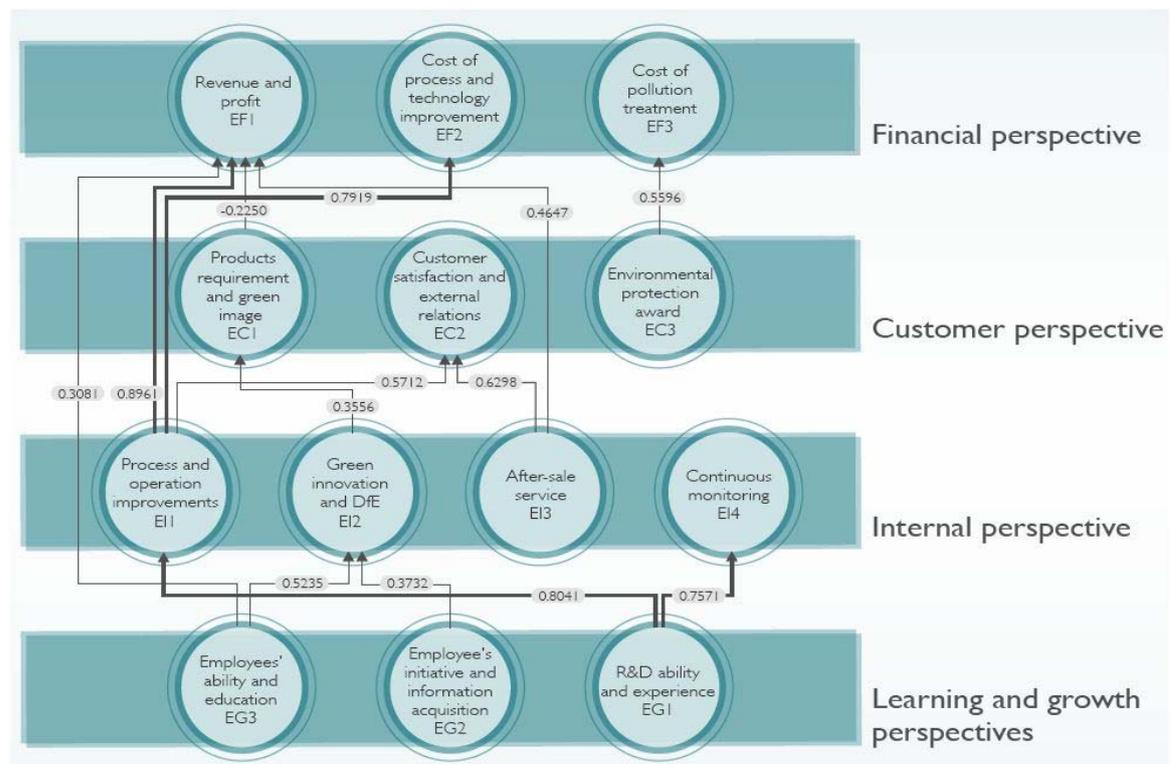


Figure 32. Balanced ScoreCard Green, National Cheng Kung University

⁹⁸ Conjointement publiée par : Dr Li-Hsing SHIH, Mr Kun-Shiang LIN, Mlle Wen-Yi HUANG ; « Resources management program », Department of Resources Eng. National Cheng Kung University

⁹⁹ www.epa.gov/dfc/ The Design for the Environment Program

→ L'avantage d'utiliser la BSC pour déployer une stratégie environnementale est de clarifier les objectifs et responsabilités de toutes les parties prenantes ; de s'appuyer sur des mesures performantes, complètes ; d'implémenter la stratégie Green avec les autres stratégies corporate en donnant une vision claire des résultats à atteindre.

3.6 Création d'une Balanced Scorecard "Green Storage"

La Balanced Scorecard présentée ci-dessous a pour objectif de dynamiser le déploiement d'une stratégie Green Storage dans l'entreprise.

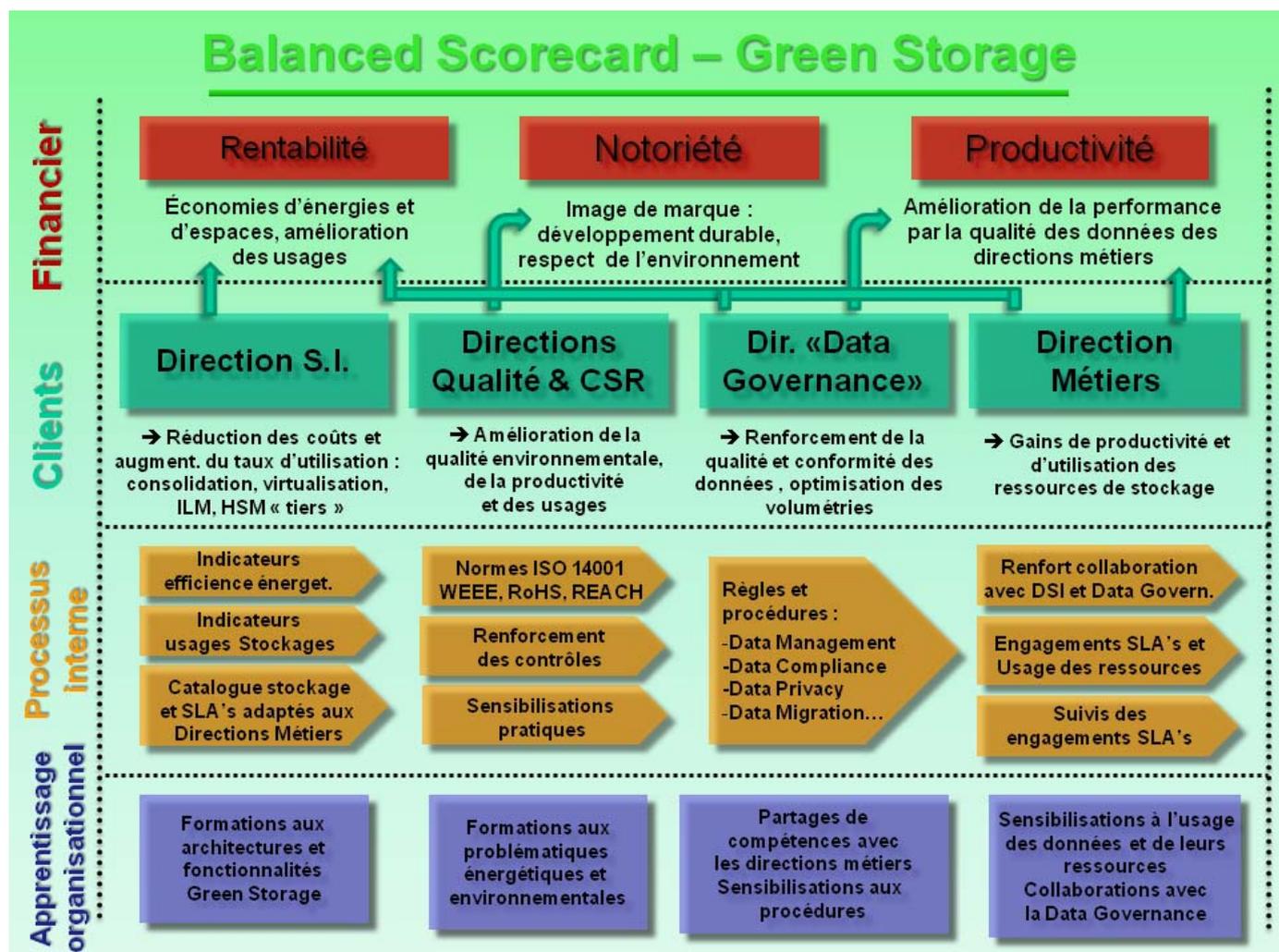


Figure 33. Balanced ScoreCard Green Storage, F-Laura

Les principales stratégies réparties dans les quatre perspectives seraient ainsi les suivantes :

L'axe Financier :

- une **rentabilité** accrue grâce aux économies d'énergies et aux gains d'espaces dans le data center. Amélioration du taux d'utilisation des espaces de données ;
- une meilleure **notoriété** grâce aux efforts faits dans les normes environnementales et le développement durable ;

- une **productivité** améliorée grâce à une qualité de donnée supérieure.

L'axe Client :

- les infrastructures de stockage sont optimisées et rationalisées par la **DSI**, grâce aux migrations, consolidations et virtualisations déployées et accompagnées d'implémentations des fonctionnalités d'ILM, d'HSM ; la réduction des coûts et l'augmentation du taux d'utilisation sont les principaux objectifs de la DSI ;

- la **Direction Qualité et la CSR** (Corporate Social Responsibility) sensibilisent l'ensemble de l'entreprise aux enjeux environnementaux (WEEE, RoHS, REACH), engagent l'entreprise vers une certification ISO 14001 ; elle met en place les nouveaux contrôles nécessaires ; leurs objectifs : amélioration de la qualité environnementale, de la productivité et des usages ;

- une nouvelle **Direction « Data Governance »** a pour objectif de renforcer la qualité et la conformité des données ; elle établit des règles et procédures sur toutes les données de l'entreprise ;

- les **Directions Métiers** doivent obtenir des gains de productivité et d'utilisation des ressources de stockage ; elles s'engagent avec la DSI et la Dir. Data Governance par des SLA¹⁰⁰.

L'axe Processus interne :

- la **DSI** crée des indicateurs d'efficacité énergétique et de taux d'utilisation du stockage ; elle élabore en partenariat avec les Directions Métiers un catalogue d'offres et de prestations « stockages » plus compréhensible et adapté aux activités spécifiques des différents métiers ;

- la **Direction Qualité** met en place les nouveaux contrôles nécessaires à l'atteinte des objectifs fixés par les normes ; elle sensibilise chaque activité de l'entreprise par la diffusion de documents personnalisés et concrets sur les bénéfices apportés par les normes ;

- la **Direction « Data Governance »** précise des procédures et règles strictes dans le management des données, contrôle leur conformité légale, définit les durées de conservation, le respect des données privatives, stimule la migration de données vers les espaces de stockage adaptés ;

- la **Direction Métiers** en collaboration avec la Dir. Data Governance, s'engage sur des SLA's ; ces contrats de niveaux de service rationalisent les besoins par des engagements clairs de chaque partie et une plus grande compréhension des usages permis par les ressources mises à disposition.

L'axe Apprentissage organisationnel :

- la **DSI** forme de nouveaux rôles : « Data Manager » et « Storage Manager », qui représenteront la DSI et intégreront la Dir. Data Governance ; ces nouveaux responsables suivent des formations spécifiques liées aux activités des Métiers et se forment aussi aux architectures et fonctionnalités « Green Storage » ;

- la **Direction Qualité** se forme aux problématiques environnementales, notamment dans les domaines énergétiques et retraitements des déchets ; elle forme des « évangélistes » qui sensibilisent chaque direction métier dans leur activité ;

- la **Direction Data Governance** diffuse à toutes les directions, les règles et procédures constituant la nouvelle politique sur les données par des documents clairs, exposant les

¹⁰⁰ SLA : Service Level Agreement (contrat de niveau de service)

procédures de façon pédagogique ; elle se forme en échangeant ses compétences avec les directions métiers ;

- les **Directions Métiers** sont formées à un usage plus rationnel des ressources de stockages et des espaces de données. Elles sont sensibilisées aux conséquences financières du non-respect des règles et des SLA.

3.6.1 Balanced Scorecard Green Storage : initiative « Data Governance »

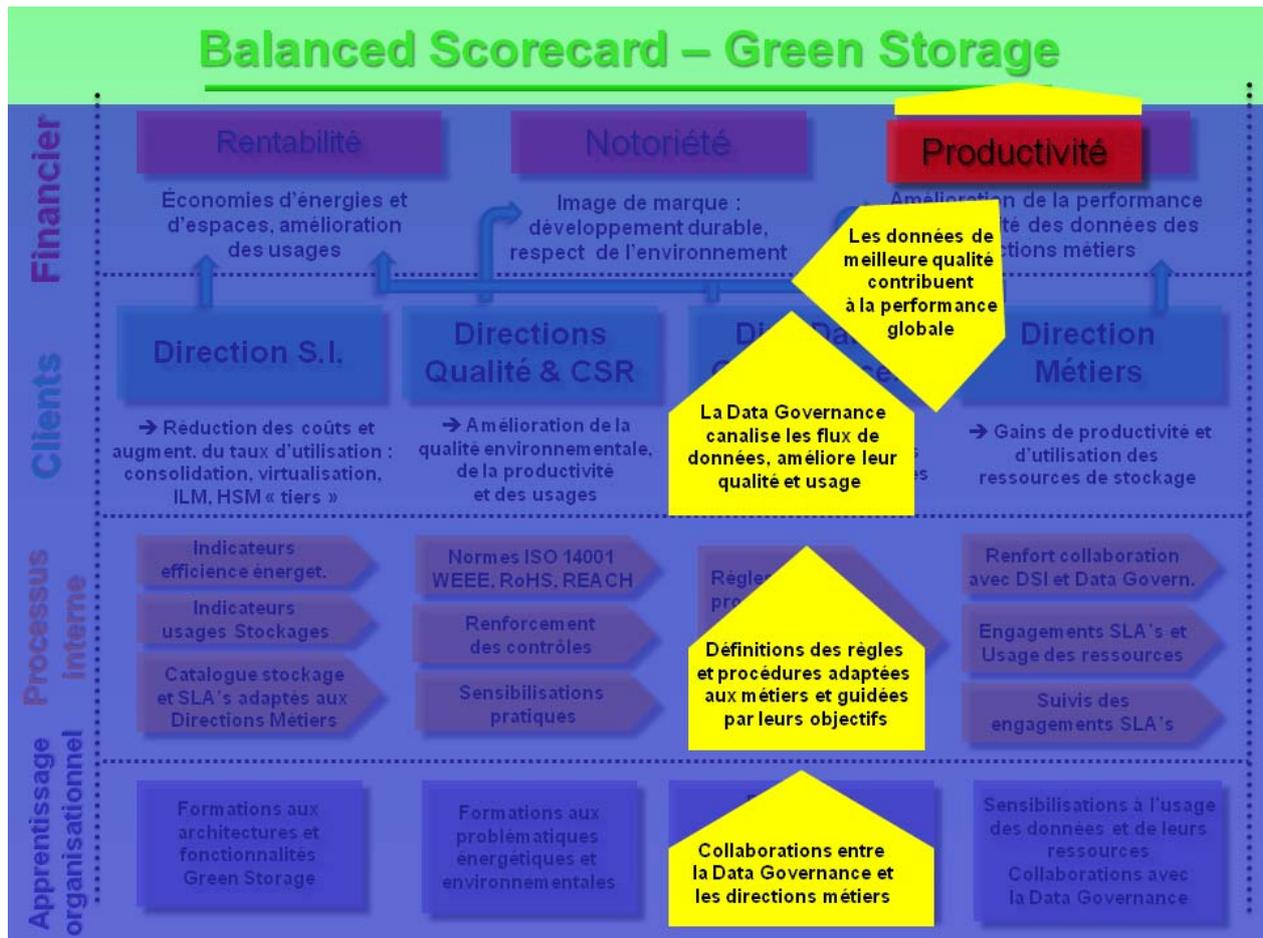


Figure 34. Balanced ScoreCard initiative Green Storage, F-Laura

Ce schéma détaille pour chaque axe de la Balance ScoreCard « Green Storage », l'initiative de la « Data Governance ». Ici, l'enjeu majeur de la direction de « Data Governance », est de s'approprier les problématiques des directions métiers, d'implémenter des règles d'utilisations et des procédures contribuant à canaliser les flux, de rendre les données de meilleure qualité pour améliorer la productivité.

3.7 L'initiative « Green Storage » du SNIA

Lancée en septembre 2007 par la Storage Networking Industry Association, l'initiative « Green Storage » regroupe plus de 25 entreprises leaders dans le secteur du stockage dont les constructeurs EMC, NetApp, Brocade, Copan, Quantum...

Ainsi, des tutoriels¹⁰¹ dédiés aux problématiques « Green Storage » ont été publiés et mis à disposition des acteurs du stockage.

Deux avancées significatives ont été faites grâce aux groupes de travail du SNIA :

Mesurer et comparer les performances « vertes » des systèmes de stockage

Le SNIA travaille sur le développement de métriques mesurant le rendement énergétique et a annoncé en février 2009 la première spécification pour mesurer la consommation d'énergies des produits de stockage green. L'objectif est dans un premier temps de comparer la consommation des systèmes en veille pour cinq classes de stockage : le stockage en ligne et nearline, les bibliothèques de médias amovibles, les bibliothèques de médias non amovibles, les systèmes dédiés de type appliance et les commutateurs réseau.

LE SNIA commence par mesurer la consommation d'équipements inactifs, car il est difficile de trouver un compromis pour les mesurer en activité. En effet, il faut pouvoir s'accorder sur la nature de l'activité et le type de charge, qui peut avantager ou défavoriser chaque technologie de stockage.

Ces nouvelles métriques peuvent stimuler la concurrence en permettant aux acheteurs et décideurs de comparer objectivement les consommations énergétiques de différents systèmes et d'ajouter des critères précis dans les procédures d'appels d'offres.

Archivage : l'initiative XAM -eXtensible Access Method

XAM est une interface standard entre l'application, le logiciel d'administration et le système de stockage pour l'archivage de données. L'enjeu est de pouvoir faire fonctionner une application compatible XAM avec n'importe quel constructeur de système de stockage compatible XAM et de rendre interopérable les composants et donc rendre plus universels les formats d'archive des données. Les conséquences positives concernent la pérennité des données archivées, leur capacité à migrer de système et à suivre plus facilement les évolutions technologiques.

3.8 Responsabilité Sociétale d'Entreprise, développement durable et Green IT

Le développement durable, la responsabilité sociétale d'entreprise et le Green IT sont interdépendants et indissociables pour la réussite de chacun.

La crise économique renforce le besoin de retour sur investissement à plus court terme et le Green IT peut aider à répondre à ce challenge.

➔ *La direction R.S.E. de l'entreprise devrait prendre en compte les systèmes de stockage pour plusieurs raisons :*

- *aider aux choix des investissements de stockage selon les critères les plus « éco-responsables » ; prendre en compte leur consommation énergétique, le traitement de leur déchet, leur recyclage, pour les rapporter et agir ensuite sur leur impact environnemental ;*

¹⁰¹ <http://www.snia.org/education/tutorials/2008/spring#green>

- veiller à l'usage éthique des systèmes : données privées, chartes d'utilisations, sécurité ;
- s'assurer que les infrastructures de stockage permettent le développement durable de l'entreprise, notamment sa capacité à faire face à un désastre (Plan de Reprise d'Activité) et à protéger son capital informationnel.

3.9 Recommandations d'associations des T.I.C.

Il est enrichissant de recueillir auprès d'associations des T.I.C. et de représentants significatifs des DSI, les recommandations et meilleures pratiques issues de leurs expériences et réflexions.

3.9.1 *Le CIGREF*

L'association du CIGREF¹⁰² – Club Informatique des GRandes Entreprise Françaises – représente plus de 100 très grandes entreprises françaises et a pour objectif de « promouvoir l'usage des systèmes d'information comme facteur de création de valeur et source d'innovation pour l'entreprise ».

Un groupe de travail du CIGREF dédié au « Green IT » rendra le résultat de ses recherches en 2009 mais dès août 2008, sur le site internet du CIGREF, **Didier Lambert** en sa qualité d'administrateur et DSI d'Essilor International, affirme que le Green IT est un pléonasme pour les systèmes d'informations. En effet, ces derniers sont initialement et légitimement des « réducteurs de consommation de l'énergie » pour de multiples activités de l'entreprise.

Par exemple, la « productique » aide à minimiser les rebuts, mais aussi contribue à maîtriser des critères de coûts, qualité et délais pour améliorer l'organisation des ressources (humaines et matérielles), l'optimisation des flux industriels, l'efficacité des procédés.

Didier Lambert affirme que les enjeux pour les DSI sont de deux ordres : trouver des solutions pour optimiser efficacement le système d'information et ses ressources informatiques, très consommateurs d'électricité et générateurs d'une forte dissipation thermique et surtout rationaliser les usages.

3.9.2 *Le CRIP*

Le Club des Responsables d'Infrastructures et de Production, présidé par **Philippe Sersot**, est particulièrement sensible aux enjeux environnementaux des centres de données dont ses adhérents ont la responsabilité quotidienne.

Un groupe de travail du CRIP, dédié aux problématiques d'optimisations des ressources de stockage, a mis en évidence l'importance de la valorisation et du classement des données le plus en amont possible comme l'expose l'illustration suivante :

¹⁰² http://cigref.typepad.fr/cigref_accueil/

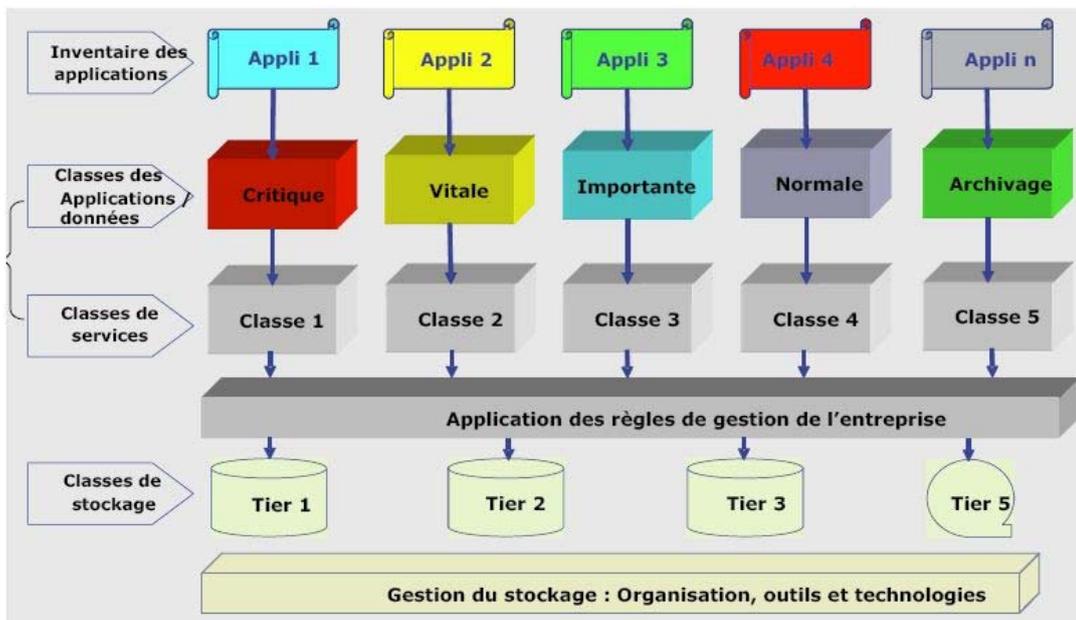


Figure 35. Classe de service de stockage, CRIP

Il s'agit de définir des « classes de services » associées à chaque valeur de donnée. Ces classes ont ensuite des espaces de stockage dédiés (tier 1 à 5) en fonction des critères de valorisation préalable : critique, vitale, importante, normale, archivage.

Une des recommandations du CRIP → Rationaliser les ressources en définissant des « tiers » de stockage liés à des « classes » de services. Ces classes se déterminent selon la valeur des données et les critères fonctionnels des directions métiers comme la perte maximale autorisée en cas de panne (RPO¹⁰³), la durée d'indisponibilité maximale admissible (RTO¹⁰⁴), le temps de réponse, etc..

3.9.3 La FedISA

Rationaliser l'usage des ressources de stockage en révisant certains usages

Jean-Marc Rietsch, président de FedISA, la Fédération européenne de l'ILM, du stockage et de l'archivage, donne sa vision en janvier 2009 des problématiques majeures en termes d'usage des infrastructures de stockage :

- Est-il prévu de migrer régulièrement les données moins accédées vers un espace de stockage secondaire ?
- Certaines pièces jointes de mail peuvent-elles être directement archivées ?
- Supprime-t-on régulièrement les données archivées après expiration de leur durée de conservation ?
- Utilise-t-on au mieux les volumétries des baies de stockage existantes avant d'en déployer de nouvelles ?

¹⁰³ RPO : Return Point Objective

¹⁰⁴ RTO : Return Time Objective

Réviser et rationaliser les rôles de la sauvegarde et de l'archivage

Jean-Marc Rietsch souligne la nécessité d'évoluer sur nos usages des systèmes de stockage et nos politiques de sauvegarde et archivage. En effet, traditionnellement, les données utiles sont d'abord sauvegardées, puis archivées dans un second temps. Toutes les données ont-elles ainsi besoin d'être archivées ?

Sa suggestion est de distinguer dans l'espace de stockage, les données encore modifiables des données figées, et de prendre directement ces dernières pour un archivage « courant et actif » sans passer par une étape de sauvegarde. Ensuite, après sélection, les données en archivage « courant & actif » sont définitivement déplacées vers un archivage « historique ».

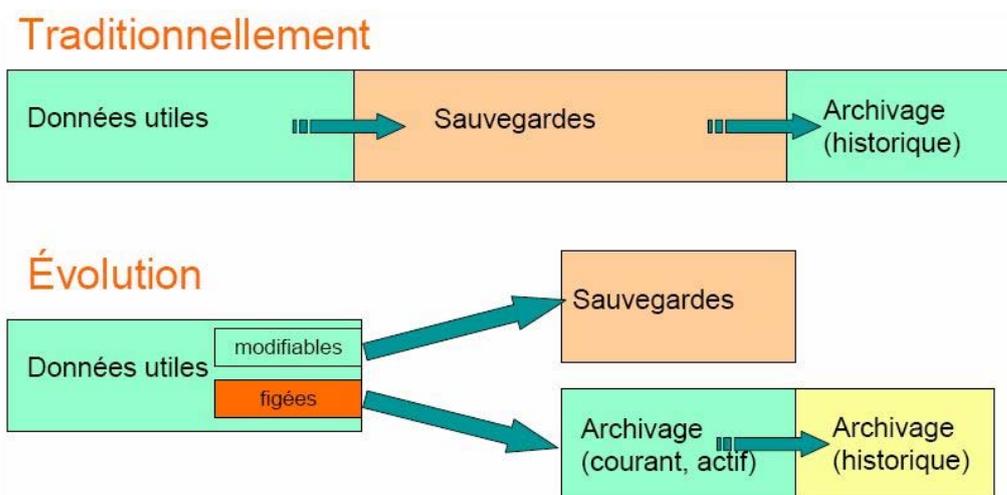


Figure 36. Usage sauvegarde et archivage, FedISA

Une des recommandations de la FedISA → Rationaliser les infrastructures de stockage demande l'acceptation d'une révision des usages et des politiques de conservation des données. Il s'agit de mieux prendre en compte le statut d'activité des données pour optimiser leur sauvegarde et archivage.

3.10 Synthèse des stratégies Green Storage

→ Ce chapitre est la synthèse des principales stratégies « Green Storage » analysée et organisée avec l'aide de **Philippe Reynier**, consultant évangéliste de la division stockage de Bull. Il permet d'établir un classement par domaine, décliné chacun en objectifs et tactiques.

3.10.1 Maximiser le taux d'utilisation des infrastructures de stockage

Objectif : Accroître le taux d'utilisation des systèmes de stockage et en réduire le nombre

Tactiques :

- consolider les différents espaces de stockage ;
- interconnecter les systèmes de stockage pour accroître l'étendue de la consolidation ;

- ➔ utiliser le plus largement possible la virtualisation pour une meilleure et plus souple utilisation des ressources ;
- ➔ consolider les infrastructures réseau, les commutateurs SAN et LAN.

3.10.2 Déplacer les données vers l'espace de stockage adapté

Objectif : Utiliser si possible les supports de stockage les plus économes en énergie et éviter le stockage de données non utilisées ou non accédées sur des supports énergétiquement inefficients

Tactiques :

- ➔ favoriser les supports de hautes capacités et les disques durs à basse vitesse (S-ATA) ;
- ➔ revoir et améliorer la politique d'utilisation de sauvegarde sur bande ;
- ➔ renforcer les procédures de stockage par « Tier » ;
- ➔ renforcer les migrations de données et l'archivage par des procédures ILM (Information Lifecycle Management) ;
- ➔ déployer les disques de technologie flash (SSD) pour les activités transactionnelles les plus lourdes (en I/Os) ;
- ➔ déployer des espaces de stockage « MAID » à base de disques durs de technologie SpinDown (arrêt de la rotation des disques durs non utilisés).

3.10.3 Accroître la densité utile du stockage

Objectif : Utiliser différentes techniques pour optimiser l'utilisation des volumétries

Tactiques :

- ➔ utiliser les snapshots pour avoir plusieurs versions de sauvegardes locales des environnements de tests et préproductions en consommant très peu d'énergie et d'espace de stockage ;
- ➔ différer l'allocation physique totale des espaces à l'aide du « Thin-Provisionnement » ;
- ➔ accroître la densité des données par la compression et la déduplication.

3.10.4 Evoluer vers une infrastructure de stockage orientée service

Objectif : Responsabiliser et rationaliser les niveaux de services proposés aux utilisateurs en les impliquant

Tactiques :

- ➔ revoir les objectifs de niveaux de service avec les directions métiers. Renégocier les objectifs utilisateurs et les attentes opérationnelles selon leur activité ;
- ➔ aligner les infrastructures et les activités en veillant à délivrer uniquement le niveau de service contractuel ;
- ➔ revoir les paramètres RAID, les répliquions, les politiques de versions et de performances ;
- ➔ éduquer les utilisateurs et renforcer l'adoption d'approvisionnement d'espace de stockage plus réaliste.

➔ **Synthèse des stratégies Green** : maximiser le taux d'utilisation des infrastructures de stockage, déplacer les données vers l'espace de stockage adapté, accroître la densité utile et évoluer vers une infrastructure orientée service.

3.11 Synthèse illustrée des stratégies “Green Storage”

L’illustration suivante à pour objectif de synthétiser les principales stratégies « Green Storage » (fondations), selon les 3 angles d’approches et d’améliorations approfondies dans la thèse (piliers), chapeautées par les directives de la RSE et les réglementations et normes existantes.

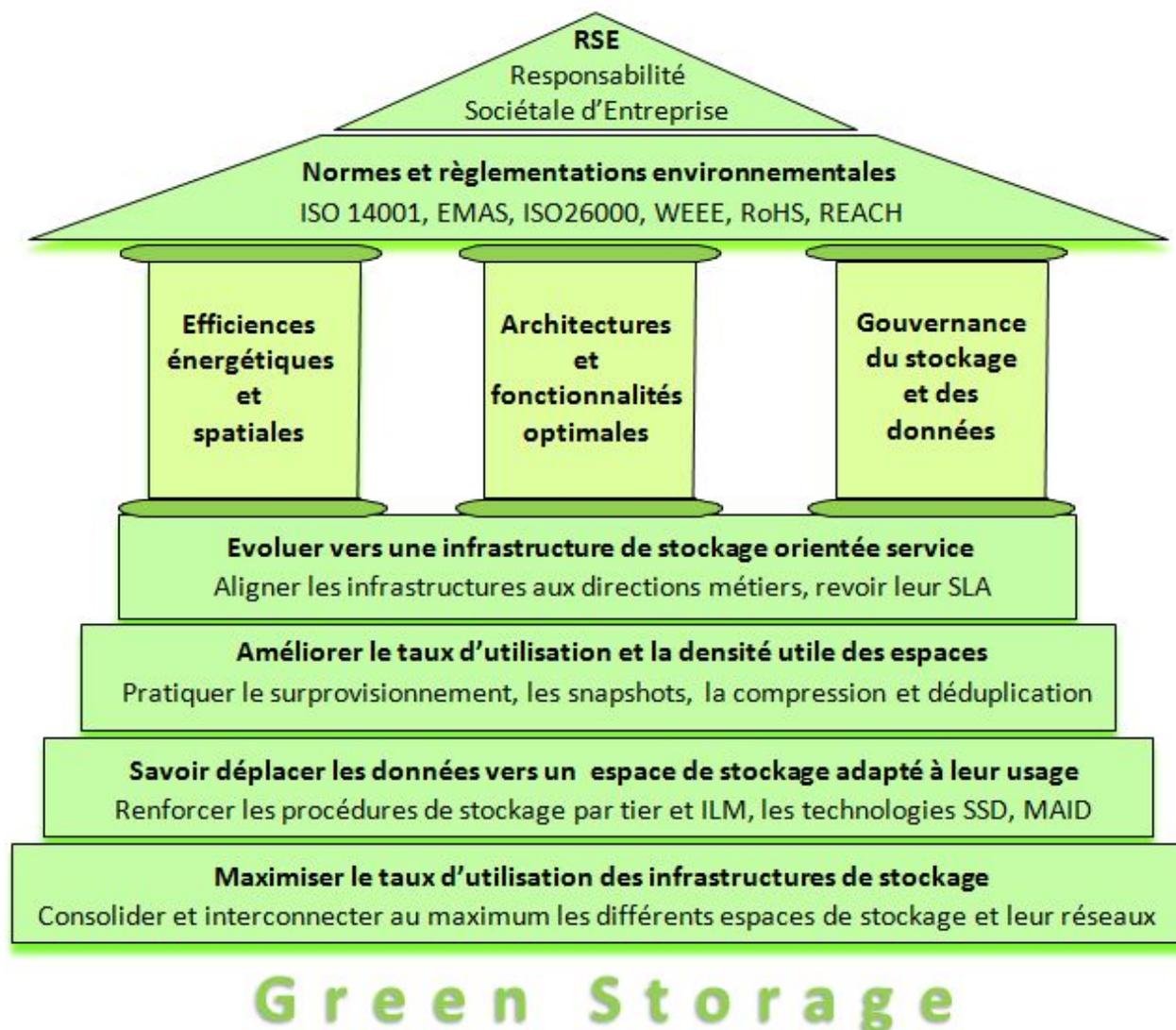


Figure 37. Synthèse Green Storage, F-Laura

3.12 Synthèse et classement des facteurs clés de succès

Un classement non exhaustif, récapitule par domaine et par type de contribution, l'ensemble des meilleures pratiques et facteurs clés de succès identifiés.



Les critères suivants sont utilisés pour répartir et classer les facteurs clés de succès :

- E-** = réduction de la consommation énergétique
- S-** = réduction spatiale (footprint)
- C-** = réduction d'émission de chaleur
- V-** = réduction de la volumétrie de donnée
- R-** = réduction des données
- OE** = optimisation énergétique
- OS** = optimisation spatiale
- OP** = optimisation de performance
- L+** = conformité légale green
- D+** = développement durable

Les critères OE, OS, OP, indiquent une contribution d'amélioration plutôt indirecte du facteur clé de succès.

Les facteurs représentés sont les plus essentiels, la liste n'est pas exhaustive.

Classement des facteurs clés de succès d'implémentation dans le data center et des matériels et composants des systèmes de stockage :

Facteurs	E-	S-	C-	V-	R-	OE	OS	OP	L+	D+	Commentaire
Implémentation des matériels											Design du data center
Plancher surélevé						X					
Réduction hauteur de plafond						X					
Refroidissement par eau						X					
Alternance aile chaude - aile froide						X					
Modélisation thermo-dynamique						X	X	X			
Composants Matériels											
Baie capacitive – multi-technologies disque						X	X	X			Packaging disque dense
Contrôleur multi protocoles						X	X	X			

FCoE	x	x					x	x	x			Unification des réseaux
Disque SSD	x	x	x				x	x	x			Solid State Disk
Disque basse consommat.	x	x	x				x	x				S-ATA et 2,5"
Disque à vitesse variable	x			x			x					Selon activité
MAID	x			x			x					Arrêt des disques aux données non accédées
Librairie de bande magnétique	x			x			x					
VTL							x	x	x			Virtual Tape Library

Classement des facteurs clés de succès liés aux architectures et fonctionnalités :

Facteurs	E-	S-	C-	V-	R-	OE	OS	OP	L+	D+	Commentaire
Consolidation	x	x	x			x	x	x			Centraliser et interconnecter
Virtualisation	x	x	x			x	x	x		x	Facilite aussi un PRA
Compression				x		x	x				
Déduplication				x	x	x	x				Mode bloc ou fichier
Thin Provisioning				x		x	x				Surallocation
Mutualisation	x	x	x	x		x	x	x			Ex. :sauvegarde, stockage hiérarchique, archivage..
Architecture par « Tiers »	x					x		x			HSM
Gestion du cycle de vie					x	x	x	x			ILM

Classement des facteurs clés de succès apportés par la gouvernance des données :

Facteurs	E-	S-	C-	V-	R-	OE	OS	OP	L+	D+	Commentaire
Storage Compliance									x	x	Conformité de l'infrastructure de stockage
Data Compliance									x		Conformité des données
Data Privacy					x				x	x	Respect des données privées
Data Protection									x	x	Continuité d'activité
Legacy Data Migration						x	x	x			Migration de données des anciens systèmes
Data Management						x	x	x			Contrôle et qualité des données, référentiel, MasterDataManagmt

Autres recommandations et meilleures pratiques :

Facteurs	E-	S-	C-	V-	R-	OE	OS	OP	L+	D+	Commentaire
Outil de reporting de consommation						X					Suivi énergétique
Outil de management actif de l'énergie	X					X					Mise en veille, extinction, adaptation à la charge
Usage des nouveaux métriques						X		X			
Catalogue de stockage orienté service	X	X	X	X	X	X	X	X			
Intégration de critères verts dans les appels d'offres	X	X	X	X	X	X	X	X	X	X	
Programme C.O.D.	X	X	X	X	X	X	X	X		X	Code de bonne conduite européen du data center

3.13 Limite des solutions et recommandations proposées

Difficulté de mise en œuvre des initiatives Green IT

L'enquête réalisée en 2008 par IDC -« Baromètre du Green IT »- auprès de 459 DSI européens de plus de 1000 employés, a révélé les principaux obstacles auxquels sont confrontés les entreprises soutenant des initiatives Green IT :

- certaines initiatives sont jugées comme risquées et ne disposent pas de métriques pour juger du retour sur investissement ;
- une pénurie de compétence et d'expertise dans le domaine « vert » limite ou retarde le démarrage des projets ;
- la responsabilité interne « Green IT » n'est pas assez ou non définie ;
- le manque de mesure assez incitative pour diminuer la consommation énergétique.

➔ *Les constructeurs, intégrateurs et sociétés de services doivent développer leurs compétences et leurs expertises pour répondre aux attentes des entreprises et les aider à franchir les difficultés de mise en œuvre d'une stratégie « verte ».*

Pollution : limites des Directives Européennes DEEE et RoHS

Le respect des Directives européennes est une obligation pour les Directions des Systèmes d'information mais leurs lacunes sont importantes :

- pas de principe du pollueur-payeur, car c'est le consommateur qui paie le coût du recyclage avec l'écotaxe ;
- ainsi, les producteurs ne subissent pas assez de pression pour une « éco-conception » et la réduction des quantités de déchets ;

- par ailleurs, c'est 6 nouvelles exemptions qui ont eu lieu de octobre 2005 à 2006 pour réhabiliter des quantités supérieures en plomb, mercure et cadmium nécessaires à certaines soudures, contacts électriques et lampes fluorescentes¹⁰⁵.

➔ *De façon générale, l'écobilan des produits n'est pas assez pris en compte par les constructeurs et leurs clients. Plus de transparence sur l'analyse de cycle de vie permettrait de prendre conscience des impacts environnementaux à tous les stades de vie d'un produit : fabrication, utilisation, recyclage et d'apporter un critère distinctif d'achat favorable à la protection de l'environnement.*

Fonctionnalités et architectures

Les dispositifs de compression et de déduplication peuvent avoir un impact significatif sur les performances et nécessitent aujourd'hui certaines précautions pour ne pas impacter les activités. Aussi selon les types de données et leur usage, les effets de diminution de volumétrie peuvent être très différents.

L'ILM, la gestion du cycle de vie des données, peut s'avérer être difficile à mettre en œuvre, notamment les critères de décision de déplacement des données. Une plus forte association de ces critères avec les applicatifs faciliterait son implémentation tout en améliorant son efficacité. L'ILM nécessite des précautions préalables afin que les déplacements de données générés (et supplémentaires) ne nuisent pas aux performances des espaces de stockages et donc des activités.

Mesures et métriques de rendement énergétique

Il est très difficile d'utiliser des métriques pour comparer différentes infrastructures de stockage et distinguer celle au meilleur rendement énergétique. En effet, un obstacle majeur subsiste actuellement : il faut s'accorder sur une activité, une charge de travail « type » appuyée par des standards qui restent à définir.

¹⁰⁵ http://smt.pennnet.com/Articles/Article_Display.cfm?ARTICLE_ID=266435&p=35&cat=HMST

CONCLUSION

Seront abordés...

- L'avenir du marché de l'informatique « verte », notamment en France ;
- la contribution des T.I.C. et du stockage pour le développement durable ;
- les principales voies en développement à approfondir ;
- les contributions possibles issues de la thèse :
 - Apporter de nouveaux contenus dans nos réponses à appels d'offres pour répondre aux attentes « vertes » des DSI et décideurs, et les sensibiliser à certains critères d'optimisation et de rationalisation de l'infrastructure de stockage ;
 - Proposer l'élaboration d'un outil d'évaluation de l'infrastructure de stockage de type « baromètre green storage », afin de permettre aux DSI d'évaluer rapidement le niveau de maturité de leur gouvernance et les axes d'amélioration possibles.

"Quand l'homme aura pêché le dernier poisson, tué le dernier animal, coupé le dernier arbre, pollué la dernière goutte d'eau, alors, il se rendra compte que l'argent n'est pas comestible"
proverbe amérindien iroquois mohawk

4.1 Un marché en plein développement

D'après Forrester, dans son étude¹⁰⁶ publiée en mars 2008 « The Dawn of Green IT Services », les services pour « l'informatique verte » de l'ordre de 500 millions de dollars en 2008, vont représenter un marché de 4,8 milliards de dollars d'ici 2013.

Le budget alloué au Green IT développera de nombreuses initiatives

Selon IDC, les « initiatives vertes » représentent en moyenne 5,9 % des budgets informatiques et atteindront 7,9 % d'ici 2010.

Les entreprises françaises commencent à réagir

Les entreprises françaises ont jusqu'à présent fait preuve de peu de motivation pour prendre en compte l'impact environnemental des T.I.C.. C'est sans doute lié au coût particulier de l'énergie en France, qui de par sa nature électronucléaire est bon marché comparé à d'autres pays dont les centrales fonctionnent au charbon ou au pétrole et dont les émissions de CO₂ sont particulièrement polluantes.

Les institutions françaises communiquent peu sur les meilleures pratiques existantes pour la mise en œuvre et l'utilisation « éco-responsable » des T.I.C..

Il suffit de se rendre sur le site internet du « Berkeley Lab »¹⁰⁷, laboratoire national américain géré par l'université de Berkeley et faisant partie du département de l'énergie (DOE), pour constater l'ampleur des ressources mises à disposition de tous depuis plusieurs années. Tout responsable de salle informatique et de systèmes d'informations peut accéder gratuitement à de nombreuses ressources :

- constatant et expliquant l'évolution de la densité énergétique et ses répercussions ;
- décrivant les recommandations améliorant significativement l'efficacité énergétique du data center et de ses systèmes ;
- démontrant le retour sur investissement des pratiques proposées par les économies énergétiques réalisées ;
- permettant l'accès : aux groupes de recherches, aux experts et à des outils : « meilleures pratiques », « auto-benchmark », « études de cas », « guide de design ».

➔ *La France doit combler son retard : en mars 2009, trop peu de documents sont disponibles pour appréhender les recommandations et meilleures pratiques pour rendre plus « verts » les systèmes d'information ainsi que leur capacité à réduire l'impact sur l'environnement de nombreux autres secteurs.*

¹⁰⁶ <http://www.forrester.com/research/Document/Excerpt/0,7211,43731,00.html>

¹⁰⁷ http://hightech.lbl.gov/data_centers.html

Création d'une commission Green IT

Le 22 décembre 2008, Christine Lagarde, ministre de l'économie a confié à Michel Petit, président de la section scientifique du « Conseil général des technologies de l'information » (CGTI), la mission de constituer un groupe de réflexion « green IT » pour des T.I.C. moins polluantes et aussi favoriser leur utilisation au service du développement éco-responsable des entreprises :

- « (...) le groupe devra élaborer des propositions concrètes sur les data centers afin qu'ils réduisent leur consommation énergétique et réutilisent mieux la chaleur produite et faciliter l'appropriation par les entreprises (...) des solutions ainsi développées »

Le gouvernement reconnaît dans son communiqué que les **T.I.C. constituent un important potentiel de croissance tout en satisfaisant les exigences de la réduction des émissions de gaz à effet de serre :**

- « (...) une meilleure exploitation des T.I.C. permettrait de gagner plus de un demi point de croissance » et « (...) Les nouvelles technologies, notamment au service du développement durable, sont une réelle opportunité économique pour l'Europe et un levier de croissance primordial dans le contexte actuel (...) »

→ *Le groupe de travail Green IT, constitué en janvier, intégrera des profils variés : professionnels du secteur, industriels, fédérations, DSI, DRH, collectivités territoriales et associations d'utilisateurs et rendra ses conclusions en mai 2009 afin d'encourager les initiatives vertes et soutenir les programmes associés.*

4.2 Le paradoxe des T.I.C. et de l'écologie

Les Technologies de l'information et de la communication permettent à quasiment tous les secteurs de notre vie quotidienne : commerce, transport, administration..., de s'améliorer sur le plan de leur efficacité tout en réduisant leur impact sur l'environnement.

Le développement des T.I.C. est un contributeur majeur pour optimiser et améliorer de nombreux secteurs de notre vie quotidienne si nous ne pouvons pas adopter un autre mode de vie moins « technologiquement carboné » et asservi aux transports, aux médias et à une consommation générale excessive.

La contribution « étendue » des systèmes de stockage

On attend des systèmes de stockage de données comme de tous les autres composants des systèmes d'informations, une contribution à l'innovation, pour limiter l'empreinte environnementale de l'entreprise.

Les progrès technologiques offrant une capacité de stockage et de traitement de données à grande échelle, **permettent de nombreuses avancées pour diminuer nos impacts sur l'environnement et contribuer à un développement plus durable**, par exemple pour les domaines suivants :

- une base de données **décisionnelle** enrichie d'outil de Data Mining peut ainsi apporter une analyse pertinente des transports liés aux approvisionnements et favoriser les fournisseurs les plus proches ou les moins polluants pour un rapport qualité / prix de leurs produits égal ;

- les techniques de **simulation** des systèmes de calcul haute performance (HPC¹⁰⁸) réduisent le nombre de tests en réel ; les systèmes HPC nécessitent de très importants moyens de stockages (et de sauvegardes), les quantités de données générées pouvant atteindre facilement plusieurs Téraoctets, voir plusieurs Péta-octets comme c'est le cas au C.E.A.¹⁰⁹ pour la simulation nucléaire ;
- de nouveaux modes de travail, consommateurs de stockage comme le **télétravail et la téléprésence**, peuvent réduire de façon importante nos déplacements ainsi que les nouveaux **outils collaboratifs** qui simplifient et permettent le travail en groupe à distance ;
- **les processus de « dématérialisation »** engagés dans de nombreuses entreprises, dans les collectivités et les ministères publics, améliorent significativement les services, l'accès et le traitement des informations, tout en réduisant ou éliminant des stocks papier, polluants et encombrants ; la dématérialisation est l'opportunité d'améliorer les processus et de rationaliser les usages.
- le développement des **télé-services** : E-Commerce, E-Santé, E-Service... ;
- **l'intelligence logicielle** comme la modélisation, l'anticipation et l'optimisation de la consommation ;
- **l'apprentissage et la formation** : sensibilisation, partage de connaissance, E-Learning et services en ligne ;

➔ *Les nouvelles technologies de l'information et des communications, leurs capacités de stockage et de traitement de données révolutionnent la productivité de multiples secteurs. Le mouvement général de développement durable favorise son développement, surtout lorsque son usage est rationalisé. Le développement durable devrait ainsi pérenniser le Green IT.*

4.3 Solutions : du vert clair au vert foncé

Il n'y a pas de remède miracle et unique pour rendre plus « vert » les systèmes d'information, de même pour les infrastructures de stockage. Il est parfois difficile de distinguer les bonnes solutions « vertes » des mauvaises, une plus ou moins grande efficacité permet de les classer, néanmoins les experts les séparent généralement en au moins deux catégories :

- vert foncé : la réduction énergétique est réelle, souvent obtenue en réduisant le nombre des applications, en retirant des serveurs et des données, mais assez difficile à appliquer si l'on veut permettre à l'entreprise de croître et d'élargir ses activités.
- vert clair : l'innovation technologique permet de « **faire plus avec moins** » en offrant une plus grande productivité, et soutenir la croissance des activités.

➔ Le marché « vert » des technologies de l'information est tellement dynamique qu'il semble ne pas avoir le temps d'éliminer les solutions qui sont uniquement du « marketing vert » : le greenwashing. Les solutions proposées peuvent être plus ou moins vertes.

¹⁰⁸ H.P.C. : High Performance Computing

¹⁰⁹ C.E.A. : Commissariat à l'Energie Atomique

4.4 Les principales voies de développement à approfondir

Les technologies suivantes sont en plein essor et leur développement devrait apporter des contributions « vertes » significatives aux infrastructures de stockage.

Diminuer le nombre de systèmes de refroidissement

Souvent devenus indépendants des serveurs avec leur propre cabinet et système de refroidissement, les systèmes de stockage pourraient revenir s'intégrer dans les armoires des serveurs. En effet, la performance de la technologie SAS¹¹⁰ combinée ou non avec la compacité des disques durs 2,5 pouces forme par exemple, une combinaison attractive pour les serveurs « lames ».

Intégrer les baies de disques durs avec leurs switches SAN dans le même châssis que les serveurs, peut contribuer à optimiser significativement l'espace et la climatisation dans le data center.

Les réseaux FCoE et 10 Gb Ethernet

La performance de ses réseaux a deux objectifs différents. Le FCoE permet d'unifier le réseau et réduit le nombre de cartes réseau nécessaires. Le 10 Gb Ethernet a une vitesse suffisante pour s'affranchir éventuellement du réseau SAN (4 et 8 Gbits/s) et de ses commutateurs, en utilisant les protocoles NAS (mode fichier) et I-SCSI (mode bloc).

Storage as a Service (SaaS)

De nouveaux services de stockage distants se développent, prenant par exemple en charge la sauvegarde des données ou le Plan de Reprise d'Activité de l'entreprise (PRA). Grâce à l'association du haut débit et des technologies dénommées « cloud computing », signifiant littéralement « nuage informatique », les services sont délocalisés, distants et omniprésents.

→ Le SaaS peut alléger l'entreprise d'infrastructures de stockages liées à la sauvegarde et au Plan de Reprise d'Activité., énergétivores et souvent « passives » (en attente de reprise d'activité). Les centres de données spécialisés dans le SaaS devraient atteindre des tailles respectables permettant une optimisation du taux d'utilisation de leurs systèmes et de leurs climatisations, tout en répondant aux contraintes de plus en plus fortes, des nouvelles réglementations et normes.

L'autopréservation (self healing)

Ces technologies ne sont pas nouvelles, mais le Green IT devrait les stimuler. Le self healing est par exemple la capacité d'un disque dur à détecter et anticiper ses défaillances (recopie en arrière plan, de secteurs défectueux vers des secteurs ou autres disques sains).

→ L'extension du self healing à d'autres composants des systèmes de stockage contribue à allonger la durée de vie et réduire les interventions de maintenance sur les équipements de stockage.

¹¹⁰ SAS : Serial Attached SCSI, technologie d'interface disque dur avec un débit de 3 Gbits/s (SAS-2 avec 6 Gbit/s annoncés) pour chaque périphérique (contrairement au SCSI qui partage sa bande passante entre ses périphériques)

La déduplication globale

La première étape de l'implémentation de la déduplication franchie, il s'agit alors de prendre en compte les nouveaux îlots de stockage dédupliqués, et d'utiliser des technologies permettant d'effectuer une déduplication globale.

D'autres progrès technologiques sont attendus, notamment pour des dispositifs permettant de soutenir la compression, décompression et la déduplication des volumes de données. L'objectif étant de diminuer l'impact sur les performances de l'utilisation de ces fonctionnalités.

Des technologies de juke-box de disques durs sont en développement, dont le principe est de prendre le disque dur, le mettre en fonctionnement, de lire ou d'écrire son contenu puis de le replacer et le déconnecter après usage. Ainsi le juke-box est aussi économe en énergie, lorsque leurs données ne sont pas accédées, que les bibliothèques de bande magnétique.

Etre plus vert en amont...

D'autres aspects concernant la qualité de développement applicative et la gestion d'un nouveau projet sont aussi en développement et nécessiteraient un approfondissement.

Développement logiciel

Optimiser le développement de l'application, son code, pour rationaliser le plus amont possible, les besoins en ressources de stockage et en création de données.

Méthode projet

Il s'agit de détecter les incohérences de développement, d'optimiser l'organisation des équipes et des phases, d'identifier les impacts possibles des fonctionnalités implémentées sur les infrastructures futures.

4.5 Suggestions et développements possibles issus de la thèse

Cette thèse professionnelle porte l'espoir d'être partagée avec le plus grand nombre au sein de l'entreprise. Des développements possibles issus de la thèse et des enseignements reçus tout au long du cycle du Mastère, sont développés dans ce chapitre sous deux formes.

4.5.1 Apport de nouveaux contenus aux réponses à appels d'offres

Les critères d'achats évoluent, notamment pour évaluer le rendement énergétique et la capacité des systèmes à optimiser leur utilisation.

Coût énergétique du cycle d'utilisation du produit

Désormais sensibles aux coûts d'exploitation, les décideurs ont l'obligation de connaître la consommation électrique ainsi que les besoins en climatisation des équipements. Il s'agit de fournir une évaluation précise, annuelle, des différentes solutions proposées.

Aussi, les DSI doivent avoir des évaluations de consommations les plus opérationnelles possible, pour associer le plus optimalement aux nouvelles infrastructures de stockage, les moyens en approvisionnement énergétique et les dispositifs de refroidissement nécessaires.

Mettre en évidence certaines caractéristiques et performances des produits

Des critères d'achat deviennent majeurs : capacité, performance transactionnelle, consommation en watt, espace occupé dans le data center, évolutivité. Ces critères peuvent

se combiner pour fournir des ratios. Aussi de nouvelles informations sont appréciées, comme la capacité de mise en veille totale ou partielle du système pour gérer l'infrastructure de stockage de façon dynamique, selon la charge des activités.

Adapter les réponses aux appels d'offres au nouveau code de bonne conduite européen du data center

Le guide des meilleures pratiques du nouveau code de bonne conduite du data center propose l'utilisation de critères, de métriques, de tableaux permettant l'évaluation de l'infrastructure de stockage et de la gestion de ses données. L'enjeu est d'adapter les réponses aux appels d'offres pour fournir un contenu adapté, « en ligne », avec les recommandations du code de bonne conduite et en rapport avec les attentes et contraintes des DSI qui ont adopté ce programme ou qui souhaitent le mettre en œuvre.

4.5.2 Elaboration d'un baromètre d'évaluation Green Storage

L'intérêt d'un outil d'évaluation de type baromètre est de réaliser par un entretien oral avec la DSI, un état général de l'infrastructure de stockage et de son usage. Il s'agit d'obtenir pour la DSI, une vision de l'état de la gouvernance de stockage à partir d'une représentation synthétique, indiquant les principaux points de forces et de faiblesses.

Les axes d'optimisations et de rationalisations sont ainsi mis en évidence, pour ensuite construire un plan d'amélioration priorisant les différentes actions à mettre en œuvre.

L'avantage est de distinguer immédiatement les actions possibles :

- à court terme et qui concernent les changements d'usages et de comportements de l'administration et de l'utilisation de l'infrastructure de stockage et nécessitant peu d'investissements matériels tout en apportant des bénéfices significatifs. Ces changements pourront concerner par exemple, les utilisateurs, les directions métiers, la DSI et l'éventuelle direction de la gouvernance des données ;
- à moyen terme et qui nécessitent aussi une remise en question de l'architecture globale de stockage incluant des études de faisabilité tout en respectant le plan de développement de la DSI et de l'entreprise ;
- à moyen et long terme et qui passent par des investissements en moyens matériels et humains modifiant significativement l'infrastructure en place, tout en démontrant de véritables retours sur investissement (ROI¹¹¹ et TCO).

La démarche d'application du baromètre est une première étape d'évaluation sensibilisant la DSI aux bénéfices de certains composants, architectures, fonctionnalités et services. Si elle adopte aussi le code de bonne conduite européen, ce dernier stimulera le déploiement de la plupart des critères exposés.

La DSI devrait pouvoir renseigner toutes les familles de critères, pour autant cela ne signifie pas qu'elle en porte la responsabilité, notamment pour certaines domaines qui sont de la responsabilité de la direction générale ou de la direction de la gouvernance des données.

¹¹¹ ROI : Return On Investment, TCO : Total Cost of Ownership

Exemple de Baromètre Green Storage

<u>Facteurs</u>	Note 1 à 5	%	<u>Domaines</u>	Note 1 à 5	%
<u>1. Datacenter et stockage</u>			<u>1. Organisationnel</u>		
Implémentations	2		Equipe opération	4	
Inventaire énergétique	3		infrastructure stockage		
Taux d'occupation cabinet	2		Equipe Gouvernance	1	
Taux d'utilisation système	3		des données		
Moyenne	2,50	50,00%	Moyenne	2,50	50,00%
<u>2. Architecture</u>			<u>2. Green IT-Stockage</u>		
Consolidation	4		Développement des	1	
Virtualisation	1		compétences		
Mutualisation	1		Traitement des déchets	1	
Moyenne	2,00	40,00%	recyclage		
<u>3. Technologie</u>			<u>3. Financier</u>		
SAN, NAS, FCOE	1		Sensibilisations	1	
Multiprotocoles	1		Moyenne	1,00	20,00%
Multidisques	3		<u>3. Financier</u>		
Disques Basses Cons.	1		Evaluation des risques	3	
VTL	0		Calcul impact financier	1	
Moyenne	1,20	24,00%	des pertes de données		
<u>4. Fonctionnalité</u>			<u>4. Développement durable</u>		
Compression	2		SLA et facturation	2	
Déduplication	1		Moyenne	2,00	40,00%
Thin Provisioning	2		<u>4. Développement durable</u>		
Snapshot - clone virtuel	3		Politique de sauvegarde	4	
Moyenne	2,00	40,00%	Politique d'archivage	1	
<u>5. Organisation des données</u>			<u>5. Transparence et</u>		
Espaces de stockage "Tiers"	4		<u>Communication</u>		
HSM énergétique	1		Catalogue direct.métiers	4	
ILM - Gestion du cycle de vie	2		Portail services stockage	3	
Moyenne	2,33	46,67%	Assistances utilisateurs	4	
<u>6. Gouvernance des données</u>			<u>5. Transparence et</u>		
Storage compliance	3		<u>Communication</u>		
Data compliance	4		(protection et suppression		
Data privacy	4		des données)		
			Moyenne	3,67	73,33%

Legacy data migration	1		
Data Management	3		
Moyenne	3,00	60,00%	
<u>7. Gouvernance du stockage</u>			
Indicateurs de consomt.	2		
Management actif de l'énerg.	1		
Usage de nouveaux métriq.	1		
Catalogue de stockage orienté niveaux de service	4		
Intégration critères Green dans les appels d'offres	4		
Audits énergétiques	2		
Programme de bonne conduit.	1		
Moyenne	2,14	42,86%	Bilan Global du Baromètre => 43,57%

Figure 38. Exemple de baromètre Green Storage

4.6 En conclusion...



« L'homme ne fait pas la terre, c'est la terre qui fait l'homme »
proverbe amérindien mohawk

Il existe un ensemble de solutions constituées de technologies distinctives, de choix d'architectures et de fonctionnalités, de recommandations et de mises en œuvre précises, pour réduire significativement la consommation énergétique de l'infrastructure de stockage, tout en maintenant le niveau de sa productivité.

L'ensemble de ces choix constitue une gouvernance plus « verte » des systèmes de stockage qui doit s'accompagner d'une autre gouvernance « verte » transversale dans toutes les activités de l'entreprise : celle des données, afin de rationaliser leur usage, et maîtriser leur formidable croissance.

L'invention du microprocesseur et l'avènement de nouveaux moyens de communication mobile et à haut débit, ont transformé profondément nos vies, désormais les Technologies de l'Information et des Communications (T.I.C.) se sont rendues indispensables à nos usages personnels et professionnels.

Internet, qui a bouleversé notre quotidien, est l'exemple le plus spectaculaire du mariage réussi entre l'informatique et les réseaux de communication.

Le développement des T.I.C. ne cesse de conquérir de nouveaux domaines et de façon spectaculaire. De constantes innovations légitiment leur rôle de premier plan dans toujours

plus de secteurs où l'efficacité de nouveaux outils (RFID, GPS, WIFI...) est sans conteste un réel progrès technologique en termes de performances et de commodités.

Pour l'entreprise, l'adoption continue des progrès technologiques est capitale si elle veut rester compétitive et rentable. Les systèmes d'informations sont aujourd'hui l'épine dorsale de ses activités.

Indéniablement, dans le même temps, notre consommation technologique effrénée a contribué à la dégradation de notre environnement naturel déjà largement saturé par l'utilisation excessive des énergies fossiles. Le constat du réchauffement climatique, de la disparition de nombreuses espèces animales, de la pollution de l'air et du sol s'impose à nous et toutes ces alarmes sonnent aujourd'hui en même temps.

L'ère de l'utilisation intempestive des T.I.C. dans une société « hautement carbonée »¹¹² doit laisser place à une ère où chaque usage des T.I.C. est justifié et efficient, s'intégrant dans une nouvelle économie « basement carbonée », fortement orientée vers le développement durable.

Les T.I.C. si prometteuses pour le bien-être de l'humanité et si contributrices aux progrès techniques d'autres secteurs en les rendant plus intelligents et productifs, doivent aussi démontrer aujourd'hui leurs « propres » efficiences.

Réussir ce défi passera par un changement comportemental plus conscient de l'impact de l'activité humaine sur l'environnement. Cela impliquera de profonds et nombreux changements à tous les niveaux de la société et dans tous les domaines technologiques.

Valoriser les usages, réfléchir à nos actes, adopter des comportements rationnels, doit s'appliquer autant aux constructeurs, intégrateurs, SSII qu'aux acheteurs et utilisateurs des T.I.C.. Ce changement d'attitude vers un comportement plus raisonnable et humble peut tout en apportant des économies substantielles, réduire significativement nos émissions polluantes et nos trop nombreux déchets.

Nous savons que respecter la nature en la préservant, c'est nous respecter nous-mêmes intergénérationnellement. Nous respecter, non seulement pour ce que nous réussirons à transmettre avec fierté aux prochaines générations, mais aussi parce que nous aurons su préserver ou contribuer à régénérer notre unique source de subsistance : la Terre.

Aurons-nous plus de sagesse dès aujourd'hui et demain pour changer ? Serons-nous assez intelligent et tenace pour rendre meilleur ce futur proche où il est question de notre qualité de vie, mais surtout de celle de nos enfants ?

¹¹²The Climate Group : « Smart 2020 : Enabling the low carbon economy in the information age »

BIBLIOGRAPHIE – SOURCES

Les références sont regroupées en deux catégories principales :

- les sources dites « académiques », écrites par des spécialistes et experts, à savoir les études des cabinets tels que Forrester ou Gartner, les livres et journaux du type Sloan Management Review ;
- les sources journalistiques issues de journaux papiers ou électroniques tels que CIO.com ou le journal du net pour lesquelles les liens internet mentionnés sont susceptibles de ne plus exister à moyen terme.

Les sources concernant le domaine de la « Data Governance » sont mises en évidence.

SOURCES ACADÉMIQUES :

The Climate Group – Global e-Sustainability Initiative (GeSI), *Smart 2020 : Enabling the low carbon economy in the information age*

Livre Blanc IDC (03/08), *The Diverse and Exploding Digital Universe Mars 2008*
http://education.emc.com/content/common/docs/ism/ISM_diverse-exploding-digital-universe.pdf

Rapports et synthèse du GIEC (2007), *Climate Change 2007 : Synthesis Report*
http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_spm.pdf
 Synthèse GIEC du gouvernement français (2007), *Groupe d'Experts Intergouvernemental sur l'Evolution du Climat*
http://www.effet-de-serre.gouv.fr/images/documents/DOC_20GIEC_20INTERNET%5B1%5D.pdf

IDC, (document n°212714, 06/08), *The Real Costs to Power and Cool All the World's External Storage*
<http://www.idc.com>

Forrester Research, Christopher Mines et Euan Davis (26/11/2007), *Topic overview : Green IT*
<http://www.forrester.com/Research/Document/0,7211,43494,00.html>

R.S. Kaplan et D.P. Norton, Harvard Business Review, Vol.70, Iss I, pp. 71-79, *The Balanced Scorecard : Measures That Drive Performance*

Dr Li-Hsing SHIH, Mr Kun-Shiang LIN, Mlle Wen-Yi HUANG ; « Resources management program », Department of Resources Eng. National Cheng Kung University, *Proactive Corporate Green Strategic Management Using Balanced Scorecard*

Livre blanc IDC, (réf. : #IDCW28Q), *Baromètre du Green It, les entreprises françaises face aux enjeux d'une stratégie informatique environnementale*
<http://www.idc.com>

Matt Stansberry, (site SearchDataCenter.com « The Green Data center, Energy efficient computing in the 21st Century » 13/03/08) *The Green Data Center 2.0*
http://searchdata.center.techtarget.com/generic/0,295582,sid80_gci1273283,00.html

Livre blanc du Green Grid, *Green Grid Metrics : describing Data Center Power Efficiency*
http://www.thegreengrid.org/gg_content/Green_Grid_Metrics_WP.pdf

Forrester Research, *Worldwide PC adoption Forecast to 2015*
<http://www.forrester.com/research/document/Excerpt/0,7211,42496,00.html>

Laboratoire Fédéral Suisse d'Essai des Matériaux et de Recherche, *Global Perspectives on E-Waste*
http://www.empa.ch/plugin/template/empa/*/51486/---/l=2

Eric Williams, Université des Nations Unis, *Computers and the Environment : Understanding and Managing Their Impacts*
http://update.unu.edu/archive/issue31_5.htm

Stéphane Krawczyk, Karim Bahloul, IDC, *L'optimisation du stockage : enjeux, difficultés et stratégies gagnantes des entreprises européennes*
<http://www.bull.com/fr/storage/wp.html>

Christopher Mines, Forrester, *The Dawn of Green IT Services*
<http://www.forrester.com/research/Document/Excerpt/0,7211,43731,00.html>

Uptime Institute / McKinsey, *Report on Data Center Energy Efficiency Released*
<http://uptimeinstitute.org/content/view/168/57>

James M.Kaplan, Rishi Roy et Rajesh Srinivasaraghavan, The McKinsey Quarterly, *Meeting the demand for data storage*
http://www.mckinseyquarterly.com/Meeting_the_demand_for_data_storage_2153

Uptime Institute, *Tier Classification Define Site Infrastructure Performance*
[http://uptimeinstitute.org/wp_pdf/\(TUI3026E\)TierClassificationsDefineSiteInfrastructure.pdf](http://uptimeinstitute.org/wp_pdf/(TUI3026E)TierClassificationsDefineSiteInfrastructure.pdf)

Uptime Institute, *Four Metrics Define Data Center Greenness*
[http://uptimeinstitute.org/wp_pdf/\(TUI3009E\)FourMetricsDefineData_center.pdf](http://uptimeinstitute.org/wp_pdf/(TUI3009E)FourMetricsDefineData_center.pdf)

J.C. Henderson et N.Venkatraman, (1993) *Strategic Alignment Leveraging Information Technology for Transforming Organizations*
<http://www.research.ibm.com/journal/sj/382/henderson.pdf>

Paul P. Tallon et Richard Scannell, ACM, (2007) *Information Life Cycle Management*

DATA GOVERNANCE :

Philip Russom , TDWI, *Data Governance Strategies : Helping your Organization Comply, Transform and Integrate*
<http://www.tdwi.org/research/display.aspx?ID=9033>

Sid Adelman, DM Review (01/08), *Strategy Components*
<http://www.dmreview.com>

Sid Adelman, DM Review (03/08) , *Without a Data Governance Strategy*
<http://www.dmreview.com>

Dan Power, DM Review (03/08), *The Politics of Master Data Management & Data Governance*
<http://www.dmreview.com>

Gwen Thomas, The Data Governance Institute (2007) *The DGI Data Governance Framework*
<http://www.DataGovernance.com>

Danette McGilvray, DM Review (12/06), *Data Governance : a Necessity in an Integrated Information World, part 1*

Danette McGilvray, DM Review (01/07), *Data Governance : a Necessity in an Integrated Information World, part 2*
<http://www.dmreview.com>

David Marco, DM Review (10/06), *Understanding Data Governance and Stewardship- Part 1, Part2, Part3*

<http://www.dmreview.com>

Livre blanc JEMM research et Informatica (01/08), *Des données de Qualité, Exploitez le capital de votre organisation*

<http://www.jemmresearch.com>

Livre blanc OCTO Technology (01/04), *Architecture de systèmes d'information, Gouvernance de la donnée*

<http://www.octo.com>

SOURCES JOURNALISTIQUES

Sciences et Vie (hors série 06/08) *Construire un monde durable*

DATA GOVERNANCE

Michael Grebb, BUZZ (09/05) *IBM Dominates Data Governance*

Colin Beasty, <http://www.destinationcrm.com> (01/08) *The Master Piece*

Robert Reeg, <http://www.cio.com> (10/07), *The mission-critical impact of Data Governance on your corporate assets*

Diann Daniel, <http://www.cio.com>: (01/08) *Dirty Data No More : Five Tips for Data Governance*

Paul Bach, Financial Executives International, <http://www.fei.org> (08/06) *360 Degree Approach to Data Governance*

AUTRES SOURCES INTERNET

Guide législatif et technique RoHS Farnell, *Guide législatif et technique RoHS*
http://fr.farnell.com/images/fr_FR/RoHS_Manual_FR.pdf

ANNEXES

Guide d'entretien : synthèse des questions posées pendant l'étape de collecte des informations

Contexte général :

- Votre activité est-elle contrainte ou sensibilisée aux enjeux environnementaux ?
- Quelles normes ou réglementations de type Green IT sont appliquées ?
- Quels sont pour vous les critères essentiels pour des systèmes d'informations plus écologiques ?

Contexte particulier de la DSI et du data center :

- Avez-vous mis en place des contrôles ou des indicateurs pour mesurer ces critères ?
- Par exemple, des indicateurs de consommations énergétiques, de performance ou d'efficacité orientés Green IT ?
- Avez-vous déjà réalisé un audit énergétique ou l'envisagez-vous à court ou moyen terme ?
- Connaissez-vous des technologies « Green IT » ? Les utilisez-vous ?
- Plus particulièrement pour les systèmes de stockage de données ?

Critères Green Storage : consommations énergétiques et espaces dans le data center :

- Connaissez-vous ou utilisez-vous des systèmes de stockage de type MAID ? des types de disque dur plus efficaces énergétiquement ? d'autres technologies efficaces en termes de consommations énergétiques ou d'espaces physiques ?

Critères Green Storage : fonctionnalités et architectures « Green Storage » :

- Connaissez-vous ou utilisez-vous des fonctionnalités de déduplication ? de « Thin Provisioning » ? d'autres fonctionnalités ou architectures de stockages contributrices Green ? Des exemples d'architectures virtualisées ou mutualisées ?

Critères Green Storage : « Data Governance » :

- Avez-vous une politique spécifique pour canaliser et manager la croissance de vos données ? Appliquez-vous ou envisagez-vous d'appliquer une « Data Governance » ? Quels sont pour vous les critères les plus efficaces « Green IT » en termes de D.G. ? Data Management ? Data Privacy ? Data Compliance ? Legacy Data Migration ? Data Protection ?

Autres critères Green Storage

- D'autres critères vous semblent-ils significatifs ? Pourquoi ?

Classement par tiers de I à IV des data centers – Uptime Institute :

Obligations standards par Tier selon UPTIME INSTITUTE	TIER I	TIER II	TIER III	TIER IV
Sources de puissance (GE, UPS, prod. de froid...)	<i>1 système</i>	<i>1 système</i>	<i>1 système mutualisé</i>	1 système + 1 système
Redondance des composants de chaque système	<i>N</i>	<i>N+1</i>	<i>N+1</i>	Minimum N+1
desserte ou chaîne de distribution	<i>1</i>	<i>1</i>	<i>1 normal + 1 alternatif en veille</i>	2 actifs en simultanée
Compartmentage des systèmes et sous-systèmes	<i>non</i>	<i>non</i>	<i>oui</i>	<i>oui</i>
Maintenance planifiée avec continuité de service	<i>non</i>	<i>non</i>	<i>oui</i>	<i>oui</i>
Tolérance aux pannes (<i>un seul événement</i>)	<i>non</i>	<i>non</i>	<i>non</i>	<i>oui</i>
Nécessite des matériels informatiques double alimentation	<i>non</i>	<i>non</i>	<i>oui</i>	<i>oui</i>
Taux de disponibilité	<i>99,67 %</i>	<i>99,75 %</i>	<i>99,98 %</i>	<i>99,99 % et +</i>
Durée annuelle d'interruption du système informatique (statistique)	<i>28,8 H</i>	<i>22 H</i>	<i>1,6 H</i>	<i>0,8 H</i>

ABREVIATIONS

ACV	Analyse de cycle de vie
CRM	Customer Relationship Management
CSR	Corporate Social Responsibility Responsabilité Sociale d'Entreprise
DC	Data center
DAS	Direct Attached Storage Attachement direct de stockage
DCE	Data center Efficiency
DCIE	Data center Infrastructure Efficiency
DCPE	Data center Performance Efficiency
Dfe	The Design for the Environment (programme environnemental de l'agence EPA américaine)
EMAS	Eco Management and Audit Scheme
FC	Fibre Channel
HSM	Hierarchical Storage Management – Stockage des données de façon hiérarchique
IEEE	Institute of Electronic & Electrical Engineering
ILM	Information Life Cycle Management Gestion du cycle de vie des informations
ISO	International Organisation for Standardization
IT	Information Technology. Technologie de l'information
ITIL	Information Technology Infrastructure Library (Bibliothèque pour l'infrastructure des technologies de l'information)

ISO	International Standards Organization
NAS	Network Attached Storage
PUE	Power Usage Effectiveness
RoHS	Restriction of the use of certain Hazardous Substances (in electrical and electronic equipment) Limitation de l'utilisation de certaines substances dangereuses
ROI	Return On Investment Retour sur investissement
RTO	Return Time Objective : objectif de temps de retour à la normale dans un plan de secours (durée maximale d'indisponibilité)
RPO	Return Point Objective : objectif de point de retour cible dans un plan de secours (définit un niveau de perte de données admissible)
RSE	Responsabilité Sociétale d'Entreprise (Corporate Social Responsibility)
SAN	Storage Area Network
SAS	Serial Attached SCSI
S-ATA	Serial Advanced Technology Attachment
SLA	Service Level Agreement (contrat de niveau de service)
SPEC	Standard Performance Evaluation Corporation (http://www.spec.org)
SSD	Solid State Disk – Disque à base de mémoire flash

TABLE DES ILLUSTRATIONS

<i>Nombre</i>	<i>Page</i>
Figure 1. Extrait du 4e rapport d'évaluation du climat, GIEC, 2007	24
Figure 2. Production d'électricité mondiale par type d'énergie	27
Figure 3. Gartner "A profound Economic Change Is Gathering Pace"	30
Figure 4. Evolution du data center	41
Figure 5. Evolution énergétique moyenne par rack, source IDC	43
Figure 6. Métriques du Green Grid	45
Figure 7. Substances toxiques	49
Figure 8. Composants toxiques	50
Figure 9. Sondage IDC octobre 2008	52
Figure 10. Evolution de la demande en stockage, IDC & McKinsey	53
Figure 11. Consommation de la volumétrie, SNIA	56
Figure 12. Les différentes efficacités Green IT, StorageIO Group	61
Figure 13. Illustration d'EMC de la R.S.E.	62
Figure 14. Principe alternance aile chaude / aile froide	67
Figure 15. Coupe illustrée d'un câble FCoE	69
Figure 16. Bénéfice du FCoE	69
Figure 17. Consommation des technologies disque et capacités, EMC	71
Figure 18. Disques durs : consommation énergétique et capacité, EMC	72
Figure 19. Indicateur radial du Green Grid	74
Figure 20. Thin Provisionning, EMC	78
Figure 21. Combinaison des fonctionnalités, SNIA	79
Figure 22. Bénéfice de la virtualisation, VMWare	82
Figure 23. Management hiérarchique et énergétique du stockage, F-Laura	85
Figure 24. Tiering dynamique, EMC	87
Figure 25. Répondre au changement de valeur des données, source ACM	88
Figure 26. Exemple de mutualisation avec la solution Calypso	89
Figure 27. Ex. développement produit, IDC & McKinsey	95
Figure 28. Niveaux de maturité des services de la DSI, ITMSF	96
Figure 29. Catalogue de stockage, 1ere partie, IDC & McKinsey	97
Figure 30. Catalogue de stockage, 2ème partie, IDC & McKinsey	98
Figure 31. Extrait de catalogue de service de stockage, IDC & McKinsey	99
Figure 32. Balanced ScoreCard Green, National Cheng Kung University	103
Figure 33. Balanced ScoreCard Green Storage, F-Laura	104
Figure 34. Balanced ScoreCard initiative Green Storage, F-Laura	106
Figure 35. Classe de service de stockage, CRIP	109
Figure 36. Usage sauvegarde et archivage, FedISA	110
Figure 37. Synthèse Green Storage, F-Laura	112
Figure 38. Exemple de baromètre Green Storage	128