



Livre Blanc

Virtualisation des copies de données La prochaine révolution des technologies informatiques ?



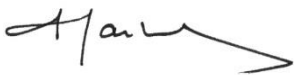
Lorsque j'occupais le poste de DSI de BNP-Paribas, j'étais entouré de prestataires et d'offres de solutions techniques d'un côté, et de prestataires de conseil et management de l'autre. Mais aucune de ces entreprises n'était capable de m'offrir un service d'inventaire, de gestion des assets au bon niveau technique et orienté business. C'est pourquoi j'ai créé Elée, dont la motivation est résumée par notre slogan : « Advanced business services for the CIO ».

Elée a été un accélérateur, pour le marché français du Software Asset Management. Il s'agit de gérer, pour le compte des DSI, la douloureuse question des licences, des audits des fournisseurs, et de se tourner vite vers une dépense choisie.

*A côté de ce besoin, mes discussions avec mes anciens collègues DSI m'ont indiqué **une nouvelle préoccupation : la gestion des copies de données**. Trop de copies, dangereuses, pénibles à manipuler, chères en stockage. Or le monde anglo-saxon, parfois dubitatif vis-à-vis de notre IT-QI, est déjà lancé dans une grande transformation. Si l'on transpose en France les résultats obtenus là-bas par les grands acteurs, nous avons sous la main une économie pour les entreprises comprise entre 3 et 5 milliards d'Euros. De quoi financer la transformation digitale !*

C'est pourquoi Elée travaille sur ce sujet intensément depuis 2015. Mais c'est encore une problématique mal connue et mal comprise. Nous avons décidé de rédiger et de publier, dans les prochains mois, une série de documents et de livres blancs. En voici le tout premier. La virtualisation des copies de données offre un axe de travail simple et puissant. Je souhaite que ce document explique pourquoi ... et que vous puissiez vous saisir de ce levier.

Bonne lecture !



*Alain Marbach
Président d'Elée*

LES REVOLUTIONS MAJEURES DE L'INFORMATIQUE

La croissance que nos civilisations ont connue depuis quelques décennies a été, pour une large part, la conséquence de la formidable explosion des technologies de l'informatique. Or, si l'on se retourne vers le passé, il n'y a eu depuis 70 ans qu'un petit nombre de vraies révolutions dans ce domaine. Si l'on exclut les technologies des disques, des puces et des réseaux pour se concentrer sur les aspects d'architectures et de logiciels, elles tiennent sur les doigts d'une main. Chacun construira sa propre liste, en fonction de ses goûts et de ses centres d'intérêt. En voici une :

1. Les langages de programmation évolués, qui ont remplacé le langage machine et l'assembleur. Le premier grand langage de programmation, Fortran, date de 1954. Les avantages de ces langages étaient tels qu'ils se sont généralisés en quelques années seulement. Leur summum est la promesse « write once, run everywhere ».
2. Internet est évidemment une innovation majeure. Internet a remplacé les infrastructures de transport de données dédiées (le téléphone, les réseaux militaires, les réseaux universitaires).
3. Une des révolutions les plus visibles, celle qui a peut-être les plus grandes conséquences y

compris au plan de notre civilisation, est que tout le savoir du monde est disponible facilement, partout et à presque tous. A qui devons ceci ? A Internet bien sûr, mais aussi aux grands moteurs de recherche, qui permettent de trouver les documents les plus pertinents sur pratiquement n'importe quelle question, où qu'ils soient et quel que soit leur format.

4. La virtualisation des machines est également une innovation importante. Elle aura mis très longtemps à s'imposer, entre la première machine virtuelle d'IBM (le VM/CMS, 1972) et la généralisation dans les centres de données des produits de VMWare ou d'Hyper V. Aujourd'hui, ce mouvement se poursuit sous la forme du Cloud.

A l'exception d'Internet, ces révolutions se sont traduites par l'apparition d'un tout petit nombre de leaders mondiaux incontestables. Pour les utilisateurs, elles se sont traduites par des modifications fondamentales de la manière de travailler, par des gains de coût d'un facteur d'au moins 10, et par l'apparition de nouveaux modèles d'affaires et de produits ou de service.

D'où l'importance de se demander quelle sera la prochaine révolution, et ce qu'elle apportera.

UN POINT COMMUN : L'ABSTRACTION, LA VIRTUALISATION

Dans chacun des cas cités plus haut, les technologies ont suivi un mouvement de balancier.

Au début, les performances du matériel (ce qui supporte la technologie) sont limitées. Il faut donc des solutions « proches de la machine ». C'est à cette condition que l'on peut avancer. Mais il en résulte, inévitablement, des silos, des technologies incompatibles, des rigidités.

Dans un second temps, la couche physique franchit un palier, et permet la construction d'une couche d'abstraction entre le matériel et les fonctions que

l'on veut fournir. Pour les langages, ce sont les compilateurs. Pour internet, ce sont les protocoles TCP/IP, HTTP, FTP...

Les avantages sont toujours les mêmes. Prenons deux exemples que sont les langages évolués et les machines virtuelles :

1. Les utilisateurs gagnent en simplicité et en universalité de leurs développements et de leurs applications. Un programme écrit en Java ou en C++ pourra tourner, sans changement, sur des machines très nombreuses. On n'aura

pas besoin de former les développeurs à de nouveaux langages pour chaque nouvelle machine. Pour les serveurs virtuels, on peut déplacer les applications d'une machine ou d'un centre à un autre, leur allouer plus ou moins de ressources...

2. Les utilisateurs gagnent aussi en agilité. Les développements en langage évolué sont beaucoup plus rapides, comme l'est l'installation d'une machine virtuelle.
3. Les coûts sont, au passage, écrasés, souvent d'un facteur 10 ou plus.

« Demandons-nous maintenant quelle pourrait être la prochaine rupture technologique »

Elle doit porter sur l'un des composants des systèmes d'information qui serait devenu à la fois source de valeur et source de complexité.

LA PROCHAINE REVOLUTION ?

Quel composant est, aujourd'hui, plus important que les données, dont tous les gourous nous décrivent l'importance stratégique, et qui sont le cœur de la mode du Big Data ? Mais en quoi sont-elles devenues si complexes ?

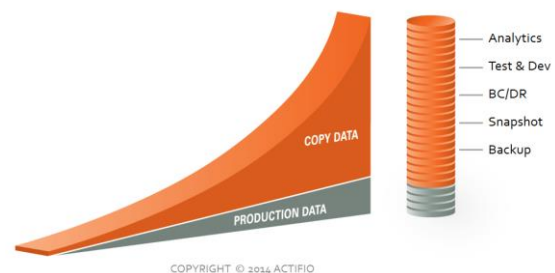
« Tout praticien de l'informatique sait que les données de production, qui sont le cœur de l'entreprise, sont copiées d'innombrables fois »

Imaginons une entreprise avec une seule base de données de production pesant 1 TB. Tous les jours, une copie de cette base est envoyée aux services de contrôle et de conformité, qui en gardent une par mois. Dans le même temps, le marketing, les ventes et le contrôle de gestion reçoivent des copies pour leurs activités de pilotage. Régulièrement, le service de développement requiert plusieurs copies pour faire ses tests. Enfin, une copie de sauvegarde est faite tous les jours, sur deux sites.

Notre base de production a donc fait des petits : une vingtaine de copies hors production, plus les sauvegardes quotidiennes. Mais il y a encore plus grave : les copies de sauvegardes sont complètes, et l'entreprise en conserve une quinzaine de versions. Il y a donc au total 35 copies de notre base ! Et cela coûte 100 k€ par an, juste pour les disques, et encore d'avantage en licences, manipulation et administration.

Et toutes ces copies ne sont pas identiques. Supposons, dans notre exemple fictif, un taux de changement de la base de 10% par jour. On peut alors calculer qu'environ 90% de ces données hors production sont des doublons. Si l'on ajoute que les bases de données croissent en taille, typiquement de 25% par an, la gabegie saute aux yeux.

Tout ceci pèse, en coût de stockage, en temps d'administration, en complexité, en rigidité. Un terrain évident pour la virtualisation ! Il est temps que le balancier de la technologie nous débarrasse de cette complexité.

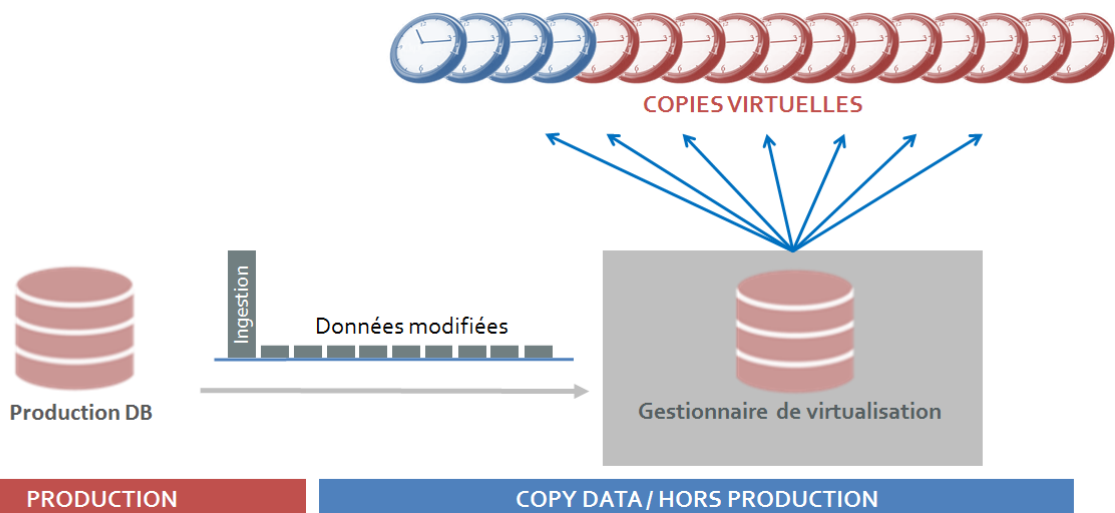


LA VIRTUALISATION DES COPIES DE DONNEES

Un mot de vocabulaire : en anglais, **on désigne toutes ces données hors production par l'expression « Copy data »**. Nous parlerons donc de Copy Data Management, ou de Copy Data Virtualization.

Le concept est développé depuis 10 ans par la société Actifio, maintenant suivie par Delphix. Ces deux sociétés proposent aujourd'hui des outils industrialisés, qui couvrent un spectre assez large de technologies. Dans la suite, nous utiliserons les termes retenus par Actifio.

Une fois les données dans la *Golden Copy*, la création d'une copie virtuelle est instantanée (typiquement 1 seconde, quel que soit le volume de données). Dans la technologie d'Actifio, les données sont en effet au format natif. Il est inutile de reconstruire la base, comme le ferait une application de sauvegarde-restauration. Au contraire, la création d'une copie virtuelle est fonctionnellement un *Mount*, la simple mise à disposition d'un pointeur, sans le moindre mouvement des données.



La virtualisation des copies consiste à gérer une seule copie (la *Golden Copy* selon les termes d'Actifio), et à la présenter sous la forme d'un nombre quelconque de copies virtuelles aux applications qui en ont besoin.

A partir de la capture initiale d'une copie de la base de production, la mise à jour de la *Golden Copy* ne requiert plus que le transport des changements incrémentaux depuis la production. L'intrusion sur la production et le besoin en bande passante sont ainsi réduits au strict minimum. Ces mises à jour sont datées (notion de Point In Time, PIT).

La *Golden Copy* gère tous ces incréments, elle peut proposer une copie virtuelle image de la base de production à tout PIT.

C'est l'un des intérêts fondamentaux d'une telle technologie, tout à fait analogue à la restauration d'une machine virtuelle à partir de son image mémoire.

Le nombre de copies virtuelles est pratiquement illimité. Leur empreinte stockage est très faible, le gestionnaire de virtualisation ne stockant que les modifications. Donc il n'est plus nécessaire que la création d'une copie virtuelle passe par un DBA : c'est un nouveau gain de temps et de flexibilité, et cela décharge les DBA d'une tâche particulièrement peu gratifiante.

L'IMPACT DE LA VIRTUALISATION DES COPIES DE DONNEES

On retrouve en matière de copies de données les grands bénéfices de la virtualisation : simplicité, agilité, robustesse et coûts.

LA SIMPLICITE

La virtualisation remplace tout un ensemble d'applications. Voici l'exemple d'un utilisateur récent de la solution Actifio :

Avant		Affected Hardware				Affected Software						
Current State	Requirement	EMC VMAX Storage	Net Backup Server	Replication Manager Server	Data Domain Storage	IBM Tivoli Storage Manager	EMC Timefinder	Net backup	Replication Manager	Data Domain	SQL Replication	SDF Replication
User Acceptance (UAT)	Performance	x	x	x	x	X	x	x	x	x	x	x
On Line Transactions	Performance	x	x	x	x	X	x	x	x	x	x	x
Train(Prod & UAT)	Functional	x										
Test (Prod & UAT each)	Functional	x										
Test/Train Refresh	Functional	x	x	x	x	x	x	x	x	x	x	x
Quality Assurance	Functional	x	x	x	x	x	x	x	x	x	x	x
Application 1	Performance	x	x	x	x	x	x	x	x	x	x	x
Application 2	Performance	x	x	x	x	x	x	x	x	x	x	x
Application 3	Performance	x	x	x	x	x	x	x	x	x	x	x
Disaster Recovery	Performance	x	x	x	x	x	x	x	x	x	x	x

LA ROBUSTESSE

Les premières solutions de virtualisation de copies de données ont été construites dans la perspective de protection (back-up et recovery). Dans ce domaine aussi, les résultats sont spectaculaires. Relancer une application de production sur une copie propre se fait en quelques minutes, au lieu de plusieurs heures ! Et, comme la mise à jour est réellement incrémentale, des RTO courts ne coûtent pratiquement rien en trafic.

Après		Affected Hardware				Affected Software						
Current State	Requirement	EMC VMAX Storage	Net Backup Server	Replication Manager Server	Data Domain Storage	IBM Tivoli Storage Manager	EMC Timefinder	Net backup	Replication Manager	Data Domain	SQL Replication	SDF Replication
User Acceptance (UAT)	Performance	x										
On Line Transactions	Performance	x										
Train(Prod & UAT)	Functional											
Test (Prod & UAT each)	Functional											
Test/Train Refresh	Functional											
Quality Assurance	Functional	x										
Application 1	Performance	x										
Application 2	Performance	x										
Application 3	Performance	x										
Disaster Recovery	Performance	x										

L'AGILITE

La virtualisation permet la création de copies en libre-service. Le gain est particulièrement visible pour le développement des applications. La mise à disposition d'une copie de données pour les tests prend aujourd'hui facilement quelques semaines. Comment concilier cette durée avec les pratiques agiles, pour lesquelles on livre une nouvelle version toutes les trois semaines ? Seule la virtualisation des copies de données résout cette quadrature, en permettant la fourniture instantanée de copies virtuelles en libre-service.

On peut calculer l'impact financier de la virtualisation sur les activités de développement d'applications. Les impacts sont spectaculaires : des dizaines de pourcent d'efficacité de la R&D et des mois de Time To Market.

LES COUTS

En plus des gains consécutifs à la simplification, à l'amélioration des développements et à la robustesse, la virtualisation permet d'évidentes économies de stockage. Nous venons de faire une estimation avec un de nos clients, et nous aboutissons, dans cet exemple, à un gain de volume en Tier 1 de 70% !

Pour tous ces progrès, nous donnerons des exemples plus précis, et des méthodes de calcul des bénéfices, dans les mois à venir.

CONCLUSION

La virtualisation des copies de données apparaît comme une des grandes ruptures technologiques en cours dans le monde de l'informatique. Elle n'est encore pas très connue, en particulier en France, car elle a moins de 10 ans, et n'a atteint sa maturité que depuis 2013-2014. Mais signe de son importance, elle apporte déjà des gains à deux grands chiffres*¹ pour toutes les fonctions qui gravitent autour de la production :

- Le développement, en permettant l'agilité
- La protection, par ses avantages de RTO, de RPO et de simplicité
- Le stockage et l'infrastructure, par ses gains de volumes de données et de besoins de trafic
- La sécurité, en permettant une anonymisation efficace

L'histoire indique que les grandes technologies de virtualisation et d'abstraction mettent une dizaine d'années à se généraliser.

Mais la France de 2016 est plus technique, plus agile, davantage décidée à livrer bataille. Parions alors que les acteurs leaders de notre pays, grandes entreprises françaises, acteurs de la nouvelle économie, belles ETI et acteurs au service du public, les auront largement déployées dans leur infrastructure avant 2020... pour leur succès et notre succès à tous. C'est ce que nous faisons. Pour vous.

¹ Plus de 20% de gains en temps et charge de développement, plus de 30% en stockage, division par plus de dix des temps de mise à disposition des données, claires ou anonymisées, et des temps de reprise... mais ceci est l'histoire de nos prochains documents



contact@elee.fr
01 84 20 12 02