

# Conférence SNIA stockage (4): la protection continue des données

Dans le stockage des données, la limitation est directement liée à la nature séquentielle du média et au nombre réduit de périphériques interdisant tout empilement d'écritures multiples et quasi-simultanées sur l'unité de stockage.

Le disque lève toutes ces limitations et offre désormais de nouvelles perspectives à la protection de données. Ces atouts justifient ce nouveau concept de **CDP** (*continuous data protection* ou PCD, **protection continue des données**), les offres fleurissent chaque semaine. On recense plus de 30 offres sur le marché, toutes se réclamant de ce nouvel acronyme à la mode dans l'industrie du stockage -le fameux CDP. Est-ce du Canada Dry? Ou comme disait un industriel connu de tous «*J'en ai rêvé, l'industrie l'a fait*»? Car oui, la promesse du CDP est bien là: il doit permettre d'éliminer les fenêtres de sauvegarde et de supprimer les fameux **RPO** et **RTO**, respectivement **objectifs de points** et de **temps de recouvrement**. Le premier se situe avant la défaillance et signifie pour le CDP l'élimination de la perte de données. Le second renvoie à la capacité instantanée de remettre à disposition l'information. La différence entre ces deux moments, c'est la somme ajoutée de leur valeur correspondant au temps d'indisponibilité total. Pour certains, ce temps est encore trop long malgré les progrès proposés par les méthodes de 'snapshot' (ou « clichés » intermédiaires). Les premières adoptions du CDP devraient se faire donc par les établissements financiers et les opérateurs de télécommunications qui sont demandeurs de telles solutions car chaque minute compte. Tous ces utilisateurs potentiels du CDP sont déjà clients de solutions de protection de données généralement de type entreprise. Néanmoins, l'offre CDP jusqu'à aujourd'hui est essentiellement représentée par une poignée de 'startups' et de quelques autres sociétés établies sans de réelle disponibilité chez les grands fournisseurs. La réactivité au marché, en réponse à la demande des utilisateurs en forte croissance, impose le mouvement d'acquisitions et les accords OEM qui commencent à se mettre en place. Nous assistons au démarrage de l'offre chez les grands acteurs ce qui devrait faciliter l'adoption de cette technologie sur 2006 et surtout 2007 par les grandes entreprises. Essayons d'expliquer les **mécanismes sous-jacents** du CDP. Le premier élément de positionnement de la technologie CDP réside dans son indépendance avec les outils traditionnels de sauvegarde. Le CDP n'est donc pas nécessairement synonyme de 'backup'. D'ailleurs sur le marché, on parle aujourd'hui de protection de données et non de sauvegarde de données, terme relayé aux oubliettes. Le CDP est disponible sous forme logiciel ou matériel comme une offre à part entière ou fonctionnalité d'un outil existant de type option de produit de sauvegarde par exemple. Une méthode simple utilisée par plusieurs offres est visiblement l'ajout d'un chemin miroir aux données couplé à une 'appliance' offrant l'intelligence CDP qui à chaque bloc passant dans cette appliance, donc pour chaque écriture, ajoute et stocke le temps correspondant. Ainsi, est constitué un espace de stockage composé des blocs avec leur date d'écriture permettant de « remonter » le temps et d'offrir une vue à un instant donné. On appelle cette approche **TAS, pour Time Addressable Storage**, l'unité de stockage étant adressable par le temps et non plus par la convention de nommage habituelle. Par analogie aux fonctions « avance et retour rapide » de tous les appareils audio ou vidéo que nous possédons tous, on demande au système de fournir l'image des données à un moment précis avant la défaillance par exemple et

ainsi le système CDP offre le volume de données qui correspond au point de recouvrement souhaité. Cette considération est primordiale pour différencier les approches '*snapshots*' ou clichés de données largement disponibles et utilisées sur le marché. En effet, en utilisant les 'snapshots', seules sont disponibles les images constituant les clichés de données établies à des fréquences généralement programmées par intervalle de temps. Alors qu'avec le CDP tous les instants sont disponibles sans que l'image demandée du stockage corresponde à un 'snapshot'. Ainsi, la terminologie CDP évoque deux notions importantes : la **notion de Single Point in Time** image régie par la méthode 'snapshot' et celle de **Any Point in Time** image intégrée à un vrai système CDP. En termes d'implémentation, trois niveaux distincts sont proposés : une approche bloc, systèmes de fichiers ou applications et on retrouve de façon assez analogue les architectures introduites il y a quelques années avec la virtualisation du stockage disque : les approches dans le chemin de données et hors du chemin de données. Pour faire le tri dans ce dédale d'offres où chacun essaie de se frayer un chemin, retenons l'importance du terme Continu dans l'acronyme CDP et il suffit de demander au fournisseur s'il est basé sur un vrai ou faux mode continu. Steve Duplessie de l'Enterprise Strategy Group baptise lui-même cette fausse catégorie de K-CDP, « Kinda » CDP, c'est-à-dire de sorte de CDP -souvenons-nous du Canada Dry! **NB:** Pour plus d'informations sur cette technologie, le comité SNIA France organise un séminaire (gratuit) à Paris le 8 novembre prochain ([www.snia-france.org](http://www.snia-france.org)). Certains détails sont également disponibles sur le blog [continuousdataprotection.blogspot.com](http://continuousdataprotection.blogspot.com). (\*) Président, fondateur de l'association SNIA France