

Le CAS, ou lettres de noblesse de l'archivage actif

Ces besoins prolongent les contraintes métiers, internes ou contractuelles déjà existantes renforçant ainsi la pression exercée sur les services informatiques, juridiques et la direction générale de l'entreprise.

Souvenons-nous des scandales Enron, Worldcom ou Parmalat en Europe. L'important, on l'aura bien compris, est de pouvoir retrouver l'information dans un temps raisonnable en s'assurant que celle-ci n'a pas été modifiée au cours du temps surtout quand l'information a valeur probante. Les produits de l'industrie doivent intégrer cette caractéristique de données de référence ou fixe au cours de temps. Les solutions disponibles ont été au cours du temps les bandes, les disques optiques et plus récemment les disques et bandes de type WORM (Write Once Read Many) c'est-à-dire à écriture unique pour plusieurs lectures possibles mais elles présentaient toutes le désavantage de n'assurer que la préservation des données sur le support sans offrir des mécanismes associés de recherche, de garantie d'intégrité, d'évolutivité et de réduction de l'espace occupé par les données. L'industrie s'est donc penchée sur ce défi et a, depuis quelques années, répondu par l'approche de **stockage adressable par contenu** ou CAS pour Content Addressable ou Aware Storage. Un des pionniers fut Filepool, société belgo-hollandaise acquise pour 50M\$ seulement par EMC en avril 2001. Depuis son introduction sur le marché en 2002 par EMC, l'offre FilePool devenue entre temps Centera, a été un véritable succès, plusieurs Peta-Octets sont installés sur la planète, sans être véritablement inquiétés par la concurrence. La réaction s'est fait attendre et les quelques startups sur ce segment mettent sur le marché leur produit depuis seulement quelque temps. Parmi les plus actives mentionnons Archivas qui a signé un accord OEM avec HDS récemment, Bycast, Caringo, Nexsan Technologies qui s'est offert EverTrust et Permabit. HP est lui aussi présent avec l'offre RISS issu du rachat de Persist Technologies, SUN avec l'offre StorageTek Intellistor et IBM avec la famille DR550. NetApp propose lui NearStore couplé à sa gamme d'outils snaplock, lockvault? mais ce n'est pas du CAS. Caringo est intéressant car la société dédiée au CAS a été fondée par Paul Carpentier, l'un des fondateurs de FilePool, et pour la petite histoire, le nom retenue pour la société correspondant au début des noms des 3 fondateurs : Carpentier, Ring et Goros. Revenons sur l'architecture CAS qu'il convient d'approfondir afin de mieux comprendre pourquoi elle est adaptée aux besoins d'archivage. Le modèle retenu est appelé RAIN (Redundant Array of Independant Nodes) et la plupart des solutions d'aujourd'hui reposent sur ce modèle. Cette architecture a l'immense avantage d'être fortement évolutive et donc au cours du temps de pouvoir permettre de stocker de forts volumes de données sur du matériel standard enrichi de fonctionnalités avancées. On parle ici d'architecture scale-out ou horizontale pour indiquer que le traitement est réparti sur plusieurs systèmes en parallèle. Plusieurs générations sont apparues, la première correspond à Centera, le seconde plutôt aux autres offres sauf Caringo qui est plutôt de 3ème et dernière génération. Pour mémoire, l'architecture RAIN est utilisée aussi par 2 offres de protection de données : Avamar Axion et Symantec NetBackup PureDisk Remote Office Edition. Une bonne solution CAS doit répondre à **plusieurs critères** : · Indépendance matérielle et surtout pérennité des composants · Capacité massive d'évolutivité pour soutenir les volumes importants stockés au cours du temps · Performance en entrées/sorties · Garantie de

l'intégrité de données, généralement réalisé grâce aux algorithmes de cryptographie MD5 et la famille SHA · Auto-configuration surtout dans les phases de reconfiguration ou d'ajout d'éléments · Auto-administration car le système risque de vivre plus longtemps que la durée de l'affectation de l'administrateur · Auto-réparation en cas de défaillance, le système s'auto-protège et se reconfigure · Et interface simple avec l'environnement au travers d'APIs ouvertes ou de protocoles standards (http, NFS ou CIFS) Côté standardisation, la SNIA est très active sur le sujet avec **3 axes essentiels** : · L'archivage centenaire où une vraie réflexion est menée pour développer notamment tout une série de meilleures pratiques, · Le SDDF (Self-Describing Data Format) qui permet définir un format de données qui embarque avec lui les meta-data c'est-à-dire les données de gestion et les attributs associés et · XAM (eXtensible Access Method) pour permettre l'échange, l'import-export des données entre solution d'archivage et notamment les solutions de type CAS. On entend déjà les analystes et certains fournisseurs dire que le CAS, la génération actuelle et montante, sera le « prochain gros truc » du monde du stockage. A suivre? _____ (*) président et fondateur de l'association SNIA France