

IT Press Tour : Sonian, archivage des mails dans le cloud

Newton, Massachusetts – 10.000 clients, 5 milliards d'objets, 2 Po en gestion, 150 millions de documents archivés par jour, les chiffres laissent rêveur pour une start-up créée en 2007 ! **Sonian** a été la première société à proposer un service d'archivage des e-mails basé sur le cloud, jusqu'aux fichiers attachés.



À quelques encablures de Boston, Sonian entend profiter des avantages qu'apporte le cloud public en s'appuyant sur les offres et architectures cloud des grands acteurs : Amazon, IBM, Rackspace, et prochainement HP. Cette stratégie lui permet de bénéficier des meilleurs coûts en fonction des offres de ses fournisseurs, dont des réductions pour les volumes importants. Mais pas seulement, puisque Sonian propose également ses offres en marque blanche, reprises par ses partenaires à leur compte.

Le marché de l'archivage des e-mails

Le marché présente un énorme potentiel, 247 milliards d'e-mails sont échangés quotidiennement – en augmentation de 20 % tous les ans, et 35 % pour les fichiers attachés – dont une partie doit être conservée, parfois plusieurs années. À ce titre, les lois, réglementations, régulations, gouvernances, etc., imposent conservation, redondance, voire rétention de l'information, sur des périodes légales qui peuvent s'étaler sur 5 à 10 ans.

Selon Sonian, en 2014 le marché de l'archivage des e-mails devrait représenter 3,342 milliards de dollars sur le legacy (stockage dans l'entreprise) et 1,536 milliard de dollars en mode SaaS.

Une offre granulaire et en marque blanche

Pour archiver les messages, la solution – qui reproduit les interfaces des principales messageries – intercepte les e-mails via SMTP et en dirige une copie vers le cloud. Les données transitent cryptées. En créant sa propre architecture de gestion des mails, l'éditeur s'est concentré principalement sur son moteur de recherche et son moteur d'indexation.

Jeff Dickerson, CEO de Sonian, nous décrit la composante économique de sa solution : « *Le stockage représente traditionnellement 20 % du coût, la recherche 20 % (search) et l'indexation 60 %* ».

L'optimisation opportuniste du stockage au travers des offres des hébergeurs aurait permis de rapporter singulièrement son coût à 2 %. Idem pour l'indexation, qui en priorisant les métadonnées – origine, type, taille et le temporel – aurait également ramené ce coût à 2 %. Reste le 'search', plus difficilement compressible, dont le coût a cependant été ramené à 24 %.

« Nous sommes disruptifs par le prix, déclare **Jeff Dickerson**. Alors pourquoi stocker sur Dropbox, Box, etc., alors qu'archiver est moins cher ? C'est pour notre approche pragmatique et nos solutions qu'Amazon a investi dans Sonian... »

Et de rappeler qu'IBM propose également la solution de l'éditeur, mais en marque blanche, qui en revanche tourne sur son propre cloud. « Nous sommes granulaires sur le cloud, nous prenons un document, le stockons au plus économique ou sur le cloud de notre partenaire, et nous assurons la recherche.

»

Une architecture dédiée

À la question d'adopter également le cloud Windows Azure, **Greg Arnette**, cofondateur à l'origine de l'infrastructure déployée par Sonian, nous répond sans ambages : « Microsoft a mis trop de choses spécifiques dans son offre cloud. Nous sommes dans l'attente de son adoption des standards. L'avenir est dans le cloud hybride avec Openstack. »

Le CTO de Sonian se penche ensuite sur l'infrastructure de l'offre, qui permet une forte granularité de la recherche, jusqu'aux documents attachés. La solution propose deux indexations : l'Indexing on demand pour les fichiers, et l'indexation de tout le document avec la possibilité d'intervenir manuellement.

Il est également possible de tagger les messages dans le 'search'. Le résultat affiche le contenu texte du mail et un extrait du (des) document(s) attaché(s). Sonian supporte ainsi plus de 400 types de fichiers différents.

Une interrogation émerge lorsque nous découvrons la complexité de l'architecture de Sonian : ne serait-elle pas très proche des architectures de big data ? « Le cloud nous amène au social, qui nous amène à la mobilité, donc au big data par les algorithmes, confirme **Greg Arnette**. Avec 5 milliards d'objets, un pétaoctets de données et 3700 unités de traitement, le multi-cloud de Sonian correspond bien à une architecture cloud big data. »

Voir aussi

[Quiz Silicon.fr – Le vocabulaire du cloud](#)