

[Communiqués de presse](#)

## **Le Stockage cognitif : enseigner aux ordinateurs ce qu'ils doivent apprendre et ce qu'ils doivent oublier**

**Paris - 05 avr. 2016:** Fermez les yeux et souvenez-vous de vos dernières vacances.

Les souvenirs dont vous vous rappelez ont été capturés car votre cerveau accorde une grande importance aux expériences significatives, telles qu'un magnifique coucher de soleil ou un merveilleux dîner.

Dans le même temps, votre cerveau accorde automatiquement une faible importance ou oublie les choses non pertinentes comme attendre à un feu ou vous enregistrer pour un vol. Grâce au stockage cognitif, les ordinateurs peuvent faire de même.

On peut leur apprendre à faire la différence entre des données à forte valeur ajoutée et celles à faible valeur ajoutée à *savoir des souvenirs ou des informations*, et cette différenciation peut être utilisée afin de déterminer ce qui est stocké, à quel endroit et pour combien de temps.

Avec l'augmentation des coûts de l'énergie et l'explosion du Big Data, en particulier celui relatif à l'Internet des objets, il s'agit d'un enjeu crucial dans la mesure où cela pourrait conduire à d'importantes économies en matière de capacité de stockage, ce qui impliquerait une réduction des coûts des supports et de la consommation d'énergie.

### **Comment fonctionne le stockage cognitif ?**

Dans un nouveau document publié hier dans la revue IEEE journal *Computer*, IBM stockage et les data scientists [Giovanni Cherubini](#), [Jens Jelitto](#) et [Vinodh Venkatesan](#) dévoilent le concept de stockage cognitif et même s'il n'est pas encore disponible, il pourrait l'être très prochainement.

Le concept repose sur une métrique qu'ils nomment « la valeur de la donnée » et qui est analogue à celle utilisée pour déterminer la valeur d'une œuvre d'art : plus la demande est élevée et plus l'œuvre est rare, plus elle aura une valeur élevée, ce qui nécessitera une sécurité renforcée.

Par exemple, si 1000 employés accèdent aux mêmes fichiers chaque jour, la valeur des données qu'ils contiennent devrait être très importante, comme un Van Gogh inestimable. Un système de stockage cognitif apprendrait ceci et stockerait ces fichiers sur des supports rapides de type flash. De plus, le système ferait des sauvegardes multiples de ces fichiers. Enfin, ces fichiers pourraient mériter une sécurité supplémentaire afin qu'ils ne soient pas accessibles sans autorisation.

Bien sûr, le contraire existe aussi. Des données rarement consultées, telles que des fichiers PDF de documents fiscaux vieux de 20 ans, devraient être sauvegardés sur des supports froids comme des bandes et disponibles uniquement sur demande. Un système de stockage cognitif saurait également que les dossiers fiscaux doivent

être conservés au moins 7 ans et qu'ils peuvent être supprimés après cette période.

Dans de nombreux cas, la valeur de la donnée peut changer au cours du temps et un système de stockage cognitif peut aussi s'adapter.

Une façon de déterminer sa valeur est de suivre les modèles d'accès d'une donnée ou la fréquence à laquelle elle est utilisée. Des personnes peuvent également ajouter des balises de métadonnées aux données pour aider le système à se former, en fonction du contexte dans lequel la donnée est utilisée. Par exemple, un astronome peut étiqueter un ensemble de données provenant de la galaxie d'Andromède comme hautement important ou moins important.

Comme détaillé dans le document, les scientifiques d'IBM ont testé le stockage cognitif en utilisant 1,77 million de fichiers partagés par sept utilisateurs. Ils ont utilisé un classement simple de catégorie 1, 2 et 3 basé sur les métadonnées et comprenant l'identifiant de l'utilisateur, l'identifiant du groupe, la taille du fichier, les autorisations d'accès au fichier, la date et l'heure de création, l'extension du fichier et les répertoires contenus dans le chemin. Ils ont ensuite divisé les données du serveur en données par utilisateur, dans la mesure où chaque utilisateur peut définir différentes catégories de fichiers qu'il juge importants.

Le résultat : une précision de la prédiction de la valeur des données de près de 100% pour la plus petite catégorie.

###

## **Cognitive Storage: Teaching Computers What to Learn and What to Forget**

Close your eyes and think back to your last vacation.

The memories you are recalling were captured because your brain automatically puts a high value on significant experiences, such as a beautiful sunset or an amazing dinner.

Simultaneously, your brain also automatically puts a low value or forgets irrelevant things like waiting at a traffic light or checking in for your flight. With cognitive storage, computers can do the same.

Computers can be taught to learn the difference between high value and low value data *i.e. memories or information*, and this differentiation can be used to determine what is stored, where it is stored and for how

long.

With rising costs in energy and the explosion in Big Data, particularly from the Internet of Things, this is a critical challenge as it could lead to huge savings in storage capacity, which in means less media costs and less energy consumption.

### **How does cognitive storage work?**

In a new paper appearing today in the IEEE journal *Computer*, IBM storage and data scientists [Giovanni Cherubini](#), [Jens Jelitto](#) and [Vinodh Venkatesan](#) unveil the concept of cognitive storage and while it isn't available yet, it could be very soon.

The idea is based on a metric they call data value, which is analogous to determining the value of a piece of art — the higher the demand and the rarer the piece typically means it will have a higher value, requiring tight security.

For example, if 1,000 employees are accessing the same files every day, the value of that data set should be very high, just like a priceless Van Gogh. A cognitive storage system would learn this and store those files on fast media like flash. In addition, the system would automatically backed up these files multiple times. Lastly, the files may want to have extra security so they cannot be accessed without authorization.

Of course there is also the opposite. A data set, which is rarely accessed, like PDF files of 20 year-old tax documents, should be stored on cold media like tape and only available upon request. A cognitive storage system would also know that tax records need to be kept for at least 7 years and that they can be deleted after that period.

In many situations, data value can also change over time and a cognitive storage system can also adapt.

One way to determine its value is to track the access patterns of the data or the frequency it is used. Individuals can also add metadata tags to the data to help train the system, depending on the context in which the data is used. For example, an astronomer may tag a data set coming from the Andromeda galaxy as highly important or less important.

As detailed in the paper, IBM scientists have tested cognitive storage using 1.77 million files across seven users. Using a simple ranking of class 1, 2 and 3 based on metadata including user ID, group ID, file size, file permissions, date and time of creation, file extension, and directories in the path. They then split the server data into data per user, as each user could define different classes of files they deem important.

The result, a data value prediction accuracy of nearly 100% for the smaller class set.

---