

Caractéristiques attendues d'un service de stockage pérenne

C. Huc – groupe Pérennisation des Informations Numériques (<http://pin.cnes.fr>)

Contexte :

Ce texte ne constitue pas une spécification mais il contient une esquisse de réflexion pour l'élaboration d'une véritable spécification de besoin. **Il essaie de définir ce qui peut être attendu d'un service de stockage sans pour autant être exhaustif sur ces besoins ni directif sur les solutions techniques ou organisationnelles permettant de répondre à ce besoin.**

C'est un texte préparatoire à la réunion du groupe PIN du 28 juin dédiée au stockage pérenne des données numériques.

Nous nous plaçons ici dans le cadre de la pérennisation de l'information numérique. Le Modèle de Référence OAIS (1) (ISO14721 :2003) (voir aussi la version française en cours de normalisation (2)) montre à quel point la question de la pérennisation de l'information sous forme numérique est complexe.

Nous nous intéressons ici plus spécifiquement à ce que nous appelons le stockage pérenne qui correspond, au plan fonctionnel à l'entité « stockage » du modèle OAIS. Cependant, pour marquer la différence entre l'entité « stockage » du Modèle OAIS (qui n'est pas un modèle d'implémentation) et notre réflexion sur sa mise en œuvre, nous parlerons ici de service de stockage.

Notre analyse du Modèle OAIS et notre retour d'expérience sur plusieurs dizaines d'années de préservation de données sous forme numérique nous conduit à concevoir l'organisation d'une archive numérique long terme basée sur 3 services :

- un service des 'Entrées'
- un service de stockage
- un service de gestion des données et communication

(cf. le cours du groupe PIN sur la 'stratégie globale et organisation de l'archivage numérique')

Le besoin

Dans la vision proposée plus haut, un service de stockage doit donc prendre en charge les données à préserver, assurer la préservation de ces données quelle que soit la durée, et restituer ces données à la demande.

Schématiquement, nous analyserons les caractéristiques attendues de ce service sous les angles suivants :

- les interfaces avec les clients du service et les fonctionnalités offertes. C'est ce point qui sera le plus développé ici.
- la garantie de préservation des données sans altération aucune : c'est la garantie d'intégrité
- la capacité à s'adapter aux évolutions de la technologie sans impact sur la préservation des fichiers et sans modification de l'interface programmatique avec les clients.
- La sécurité et la confidentialité

Les questions relatives à l'authenticité des données ou des documents archivés ne sont pas abordées ici

LES INTERFACES AVEC LES CLIENTS

Dans le schéma d'organisation, il apparaît que les autres services de l'archive sont des clients du service de stockage. Ces services peuvent être locaux ou distants.

Dans une vision plus large (cf. Modèle OAIS chapitre 6), on peut aussi penser qu'un service de stockage puisse être partagé par plusieurs clients distincts et indépendants.

Le service de stockage met à disposition de ses clients une interface programmatique permettant de mettre en œuvre les différentes fonctions

Les fonctions de base offertes aux clients :

- versement d'un ou d'un ensemble de fichier (au stade actuel, la donnée versée est supposée prendre la forme d'un ou d'un ensemble de fichiers)
 - affecter chaque fichier à une classe de service en matière de préservation (permet d'affecter à chaque fichier un niveau de criticité en matière de préservation, suivant la classe de service choisie pour un fichier, le service de stockage sera amené à dupliquer ou non le fichier, à stocker les médias en plusieurs sites séparés, à utiliser tel ou tel type de média...)
 - affecter chaque fichier à une classe de service en matière de restitution (suivant la classe, le temps de restitution pourra être de quelques secondes, quelques minutes, ...voire un jour ouvrable)
 - gérer ses noms de fichier en suivant les règles du système d'exploitation qu'il utilise et sans contrainte imposée par le service de stockage,
 - suppression ou renommage d'un fichier stocké
- demander la restitution d'un ou d'un ensemble de fichiers (la restitution d'une portion de fichier dans des conditions à définir paraît aussi une fonction de base intéressante dans le contexte de la technologie XML)
- créer, modifier, supprimer une arborescence de stockage virtuelle afin de permettre au client de disposer d'une vue organisée de ses données, quels que soient les supports physiques utilisés
- disposer d'une vue globale détaillée permanente de tout ou partie de l'espace de stockage
- avoir accès aux informations statistiques sur les données entrantes ou sortantes du service de stockage

Il importe de souligner que ces fonctionnalités doivent être disponibles quels que soient les moyens matériels et logiciels des clients. En d'autres termes, le service de stockage n'induit aucune contrainte sur ses clients en matière de choix des matériels et logiciels utilisés.

Dans cette approche, le service de stockage voit tout fichier à archiver comme un train de bits dont il doit conserver l'intégrité. Il n'est pas concerné par l'information dont ce train de bits est porteur. La capacité à restituer, pour l'utilisateur final, le contenu intelligible du train de bit n'est pas de la responsabilité du service de stockage.

Pour un même client, il peut y avoir plusieurs 'utilisateurs' distincts du service de stockage. Nous définissons ici l'utilisateur comme étant la personne physique ou le programme qui active les fonctions d'accès au service d'accès. Tous les utilisateurs d'un même client n'ont pas le même profil. Il s'en suit un besoin en terme de gestion de droits d'accès, cette gestion pouvant être en partie 'déléguée' au client lui-même.

La Stabilité de l'interface programmatique au cours du temps est une contrainte essentielle. Toute modification de cette interface induit immédiatement des conséquences sur les logiciels des divers services clients.

LA GARANTIE D'INTEGRITE

La garantie d'intégrité des données par le service de stockage constitue probablement sa caractéristique première.

Cette garantie d'intégrité implique deux aspects distincts :

1^{er} aspect : la mise en œuvre par le service de stockage d'un ensemble de procédures incluant une politique de choix des supports, une surveillance de l'état des supports (état des CRC, nombres d'opérations de lectures...), la production et l'analyse des statistiques d'activité du service, des renouvellements ou duplications systématiques périodiques des supports, la gestion des données sur plusieurs lieux de stockage distincts, l'usage d'empreintes numériques....

2^{ème} aspect : la capacité à apporter, vis-à-vis des clients, la démonstration de la capacité du service de stockage à véritablement garantir l'intégrité des données stockées avec un degré de confiance satisfaisant. Cet aspect implique une certaine transparence sur tous ces aspects, la possibilité de subir des audits ou des inspections, l'éventuelle 'certification' du service au regard d'un standard de certification à venir, etc.

LA PRISE EN COMPTE DU TEMPS

Le contexte est celui de l'archivage long terme. Le service de stockage est donc un service de longue durée qui doit être capable :

- de maîtriser la migration des données sur les médias physiques appropriés en fonction de l'évolution des technologies et de l'obsolescence de certaines d'entre elles,
- de gérer en parallèle plusieurs types de média physiques
- de faire évoluer certains de ses propres composants matériels et logiciels sans interruption du service vis-à-vis des clients et sans modification de l'interface programmatique.

Le Modèle O AIS établit une classification des supports. Pour ce qui concerne le service de stockage, les catégories identifiées sont : le rafraîchissement de support, la duplication, et le ré empaquetage). Le client du service de stockage est donc déchargé de toute gestion des supports physiques de données quelle que soit la durée de stockage et quelles que soient la taille des fichiers.

En ce qui concerne le second point, on peut considérer que le service de stockage doit être « multi fournisseurs ». Le standard d'interopérabilité (SNIA Storage Management Initiative Specification) permet l'écriture d'interfaces d'accès compatibles entre elles pour des moyens de stockage distincts va dans ce sens. Pour plus d'information sur cette norme : <http://www.snia.org>.

SNIA : Storage Networking Industry Association (Storagetek et IBM notamment en sont membres)

LA CONFIDENTIALITE DES DONNEES

En fonction du contexte, il y aura un besoin plus ou moins élevé d'assurer la confidentialité des données. Il y a donc lieu de pouvoir hiérarchiser les besoins et les technologies permettant d'y répondre.

REFERENCE

- (1) Reference Model for an Open Archival Information System, (ISO14721 :2003), version PDF disponible sur <http://www.ccsds.org/CCSDS/recommandreports.html#interchange>
- (2) Version française du Modèle OAIS en cours de normalisation, accessible en ligne sur <http://pin.cnes.fr> rubrique 'actualités'