

Solution Big Data pour projet collaboratif

Contexte

Les projets collaboratifs et expérimentaux comme le projet SAM génèrent un nombre considérable d'information (données) qu'il est nécessaire de produire, stocker, manipuler, partager et conserver. Anticiper les besoins, unifier les visions et simplifier les processus sont autant de sujets qui doivent rapidement être adressés pour garantir le succès de ces projets.

Définition de données

L'intervention de plusieurs acteurs lors d'un projet nécessite de mettre en place une **définition commune** pour chaque information (**taxonomie du projet**). Ces définitions accompagnées d'un ensemble de propriétés constituent le **langage commun** autour duquel s'articule l'ensemble de travaux (acquisitions des données, traitements, analyses, etc.). Ce langage commun permet ainsi à chaque acteur d'anticiper et de préparer l'ensemble des processus associés à la manipulation de ces informations.

Format de données

La **taxonomie du projet** doit être accompagnée du format de données (structure de l'information).

Cette structure peut être définie grâce à des bases de données ou plus généralement de fichiers informatiques. Cette dernière solution offre le plus de flexibilité et une meilleure adaptabilité à tous les types de projets envisageables. Il est recommandé d'utiliser des **formats simples** (ex : fichier texte) car ils permettent à tout type d'utilisateurs de les exploiter.

Partage de données

Trois modèles sont possibles :

- **Décentralisé** : chaque partenaire souhaite héberger ses propres données,
- **Centralisé** : l'ensemble des données est mis en commun dans une solution unique. Ce modèle homogénéise l'accès aux données et simplifie le partage
- **Hybride** : les deux modèles précédents sont utilisés

Solution technique

De tels projets doivent reposer sur un **socle technologique** (ici, ensemble de logiciels) **Open-Source*** autour desquels il est possible de capitaliser les efforts fournis, bénéficier du soutien de la communauté et permettre la contribution de nouveaux partenaires et/ou l'ouverture des travaux. Pour ce faire, il existe déjà **des standards** sur lesquels s'appuyer pour construire ces solutions qui garantissent une **indépendance technologique** vis-à-vis des solutions logicielles propriétaires des acteurs internationaux des GAFAMs (Google Apple Facebook Amazon Microsoft).

Suivant le socle technologique choisit, il est également possible d'associer des outils permettant une visualisation des traitements définis pour simplifier le travail des utilisateurs et le traitement des données en temps-réels.

L'**interopérabilité** (connexion) de ce type de solutions avec les **écosystèmes des partenaires** est la clé pour réduire les efforts en lien avec la manipulation de l'information et libérer la charge de travail au profit des aspects métiers des projets.

*Open-Source : licence de logiciel qui autorise les utilisateurs d'utiliser, d'étudier, de modifier et de distribuer le logiciel et son code source à toute personne et à toutes fins, aussi dit libre de droits

Conclusion

Disposer d'un **langage** commun entre partenaires et d'une solution technologique libre de droits, **interopérable** et facilement utilisables par tout type de profil métier sont la clé pour mener à bien des projets ambitieux dont la donnée est au centre des problématiques traitées.

Il est nécessaire de se **concerter** et de **s'accorder** en début de projet, garantissant ainsi la production de résultats de qualité et minimisant les aléas relatifs à tout projet collaboratif.