

RÉFÉRENTIEL HAYSTACK

Juillet 2018

CONTEXTE

Le Big Data est le nouvel or noir du XXI^{ème} siècle. Les décisions qui seront prises autour des données dans les prochaines années impacteront de manière majeure la société dans les décennies à venir.

Aujourd'hui, si les données commencent à se développer dans nos sociétés modernes, celles-ci font face à un triple challenge, appelé « **challenge des 3 V** » :

- Challenge de **Volume** → il existe un défi de traitement du volume des données pour les entreprises.
- Challenge de **Vitesse** → certaines entreprises se basent parfois sur des données datant parfois de plusieurs mois, ce qui les rend moins compétitives. Aujourd'hui, les entreprises pour mieux répondre aux attentes clients ont besoin de données en temps réel ou presque.
- Challenge de **Variété** → la variété des types de données est un challenge supplémentaire à prendre en compte car le croisement de données différentes permet parfois d'expliquer certains phénomènes et de mener des actions préventives.

LES DATA : UNE MINE D'INFORMATION A STRUCTURER ET DECRIRE

Aujourd'hui, la multiplication des outils d'IOT permet de fournir une importante quantité de données. Cependant, les data issues de ces objets connectés manquent de structuration, de précision ; ce qui ne permet pas de rendre celles-ci opérationnelles et utilisables par des process logiciels.

En effet, si une donnée permet bien d'obtenir une information, quel que soit son type (chaîne de caractères, nombre, tableau...), celle-ci n'est exploitable et intéressante à utiliser que si elle dispose de **métadonnées**, ou **données sémantiques**, apportant un contenu et une « identité » à cette donnée. Les métadonnées d'une donnée peuvent par exemple donner les renseignements suivants :

- ◆ **L'horodatage** ;
- ◆ **La nature** et l'unité de la donnée : ce que représente la donnée (pression, vitesse, consommation...) ;
- ◆ Toute information complémentaire susceptible de **caractériser de manière unique** la donnée.

Exemple :

La valeur est de 20

- Le capteur fournit la valeur,

La Température est de 20

- La valeur du capteur représente un type de donnée,

La Température est de 20 °C

- La valeur mesurée par le capteur est exprimée dans une unité,

La Température de l'air au 4^{ème} étage est de 20 °C

- Des informations complémentaires sur la localisation de la donnée peuvent être apportées,

A 12 h 00 le 12/07/2018 la Température de l'air au 4^{ème} étage est de 20 °C

- L'horodatage permet de suivre l'évolution dans le temps de la donnée.

LE PROJET HAYSTACK

Le projet Haystack est une **initiative open source** qui a pour but de développer et standardiser les données sémantiques apparaissant sous forme de tags. Ces métadonnées sont définies par consensus de la part de l'ensemble de la communauté. Le but de ce projet est de donner du sens à la donnée pour qu'elle puisse être utilisable tant par un humain que par une machine. Haystack est tourné vers l'ensemble des usages du bâtiment.

Le projet Haystack permet ainsi une automatisation plus grande des systèmes connectés du bâtiment grâce à :

- ◆ Une **interopérabilité** forte, c'est-à-dire la capacité de systèmes à interagir entre eux avec différents protocoles,
- ◆ Une **normalisation** des données et des standards qui permet de rendre les data identifiables,
- ◆ Une **structuration** de ces données,
- ◆ L'**indépendance** des modèles,
- ◆ L'utilisation opérationnelle de la **sémantique** des données, c'est-à-dire de la signification intrinsèque d'une donnée.

Aujourd'hui utilisée sur des milliers de bâtiments par plus d'un million d'utilisateurs, cette solution est adoptée par les constructeurs d'équipements nouvelle génération et a déjà prouvé sa fiabilité. Cet outil peut être implémenté dans de petits systèmes, des contrôleurs de réseaux ainsi que des bases de données.