

# LE GLOSSAIRE DATA - VOLUME II



**Data Cleansing :** Ce terme signifie “nettoyage des données”. Cette étape consiste à supprimer des données incohérentes, corriger des erreurs comme des doublons ou des données mal saisies.



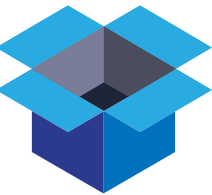
**Chief Data Officer :** Le CDO assure un contrôle globale de la gouvernance des données. Il est garant de la qualité des données collectées et de la conformité des métadonnées.



**Machine learning :** Grâce à l’intelligence artificielle, le Machine Learning (apprentissage automatique) se traduit par un apprentissage spécifique d’algorithmes qui se développent grâce à des données. Ce développement est automatique.



**IoT :** Ce terme désigne “l’internet des objets”. De plus en plus d’objets peuvent se connecter les uns aux autres (véhicules connectés, montres connectées, meubles connectés, etc.), formant un réseau de connexion, l’internet des objets.



**OpenData :** La “donnée ouverte” est une donnée numérique dont l’accès et l’utilisation sont en libre accès aux usagers. Cette donnée peut soit venir de sources publiques ou privées, que ce soit par une collectivité, un service public ou une entreprise.



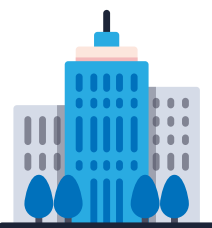
**Dark Data :** Ces “données sombres” correspondent à toutes les données recueillies par les entreprises mais qui ne sont pas utilisées par la suite dans un but précis. Ces dernières ne seront sûrement jamais analysées.



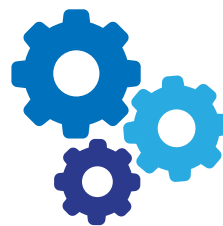
**Qualité des données :** Afin de ne pas se reposer sur des données erronées, il est nécessaire de s’appuyer sur des données fiables. Voilà pourquoi la mise en place d’une stratégie de gouvernance des données est conseillée.



**Data Scientist :** Ce statisticien expert est capable de manipuler des outils spécifiques du Big Data pour comprendre des enjeux business et de les analyser. Le Data Scientist peut ainsi faire le lien entre les aspects techniques et les besoins métiers.



**Data Center :** Le centre de données est une plateforme permettant le stockage et la sécurisation de traitements relatifs aux données. La plupart du temps, ces Data Center sont mis en place par des prestataires externes spécialisés en gestion des données.



**3V :** Les trois “v” représentent Volume, Vitesse et Variété, les trois critères d’une plateforme Big Data.