

# Où placer ses données et pour quels usages ?

	Base relationnelle	Entrepôt de données	Data Lake	Data Mart	Base opérationnelle
Types de données	Structurées, alphanumériques, texte et dates organisées selon le modèle relationnel	Relationnelles issues de systèmes transactionnels, des bases opérationnelles et des applications	Structurées et non structurées, issues de capteurs, des sites Web, d'applications métiers, d'apps mobiles, ...	Relationnelles, sous-ensemble de données spécifiques	Transactionnelles, issues de nombreuses sources
A quoi ça sert ?	Traitements de transactions	Les données sont stockées pour les outils de BI, le reporting en mode batch et la visualisation de données	Analyse des Big Data, Machine Learning, analyse prédictive et découverte des données (data discovery)	Les données sont exploitées par une communauté d'utilisateurs spécifiques à des fins analytiques	Ingérer, intégrer, stocker et préparer les données pour l'analytique ou l'opérationnel. Ces données alimentent souvent un entrepôt de données
Sources de données	Les données sont issues d'une seule source, comme par exemple un système transactionnel	Multiples, en provenance de systèmes relationnels	Multiples sources qui renferment des formats variés de données	Les données proviennent généralement de l'entrepôt de données, mais peuvent aussi être extraites de systèmes opérationnels ou externes	Multiples sources, issues des applications de l'entreprise
Normalisation des données	Les schémas sont statiques, les usages normalisés	Les schémas sont dénormalisés ; schema-on-write	Dénormalisés, Schema-on-read	Normalisés ou dénormalisés	Dénormalisés
Avantages	Garantit la cohérence des données aux applications métier	Les données historiques et leurs sources sont stockées dans un même emplacement; les données sont classées pour en faciliter l'accès	Les données sont stockées dans leur format natif pour augmenter la flexibilité donnée aux data scientists dans l'analyse et le développement de modèles	L'accès aux bonnes données, à destination des bonnes applications et bons utilisateurs, est simple et rapide	Les requêtes sont accélérées car effectuées sur de petits ensembles de données en temps réel (ou presque temps réel). Pour des usages analytiques ou opérationnels
Qualité des données	Les données sont organisées et cohérentes	Les données sont sélectionnées et triées, centralisées et prêt à l'usage à des fins de BI et d'analyse	Les données sont brutes et peuvent ne pas avoir été triées	Les données sont sélectionnées et triées	Les données sont nettoyées et mises en conformité, mais peuvent ne pas être aussi cohérentes que dans un entrepôt de données