



DATA LAKE – CONSTRUIRE SA PLATEFORME DE DONNEES EN IAAS

Le Big Data, depuis plusieurs dizaines d'années déjà, est une tendance forte dans le monde de l'IT. Les entreprises tentent de construire leur data lake ou de le faire évoluer en une plateforme de données complète. Le stockage et le traitement de gros volumes de données n'est plus un enjeu aujourd'hui, les technologies existent et ont été éprouvées. Cependant l'écosystème est vaste et il peut être difficile de trouver la technologie adaptée à son contexte pour la déployer en mode IaaS.

Objectif du stage

L'objectif de ce stage est de cartographier les technologies disponibles pour chaque fonctionnalité d'un data lake à travers un cas d'usage comprenant toutes les briques techniques d'ingestion, de stockage (système de fichiers distribué, NoSQL), de traitement par batch et d'analyse des données. Ces technologies doivent être utilisables dans les systèmes d'information des entreprises « on premise » (les solutions managées du cloud ne sont pas éligibles). Dans le cadre de ce stage, les technologies seront installées en IaaS.

Travaux à réaliser

Intégré pour la durée de votre stage à la Taskforce Data, vous serez amené, sous la supervision d'un consultant, à réaliser les travaux suivants :

- Rédiger une note de cadrage du sujet de stage, comprenant :
 - Votre compréhension du stage et ses objectifs
 - La définition du cas d'usage implémenté via la plateforme
 - Un planning prévisionnel et ses jalons clés
- Identifier les solutions à mettre en place pour chaque brique technique
- Mettre en place les outils de l'architecture sur une infrastructure cloud (IaaS)
- Implémenter le cas d'usage défini à l'aide des outils installés
- Etudier les méthodes d'exposition des données pour une utilisation dans les applications transactionnelles
- Identifier les forces et faiblesses des solutions
- Créer un white paper / séminaire sur la mise en place d'une plateforme de données en IaaS

En parallèle, vous participerez à des missions de conseil, en intégrant une équipe de consultants Nexworld, placée sous la responsabilité d'un directeur de mission.

Outils et technologies (non exhaustif)

- ✓ Hadoop, MapReduce, Spark, Hive, HBase, MongoDB, Cassandra, Redis, Neo4J, Zookeeper, Scality, Ceph, Redis
- ✓ Java, Python