

# BigData : Informatica V10

Référence : PYCB081

Durée : 3 jours

Certification : **Aucune**

## CONNAISSANCES PREALABLES

- Connaissances générales sur les manipulations de données BigData (NoSQL).

## PROFIL DES STAGIAIRES

- Chefs de projet, développeurs, architectes et toute personne souhaitant utiliser Informatica dans le cadre d'applications BigData.

## OBJECTIFS

- Savoir concevoir et mettre en œuvre des traitements de données avec Informatica BigData Management dans un environnement hadoop, et traiter des données NoSQ.

## CERTIFICATION PREPAREE

Aucune

## METHODES PEDAGOGIQUES

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émergence

## FORMATEUR

Consultant-Formateur expert Bigdata

## METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation de fin de stage adressée avec la facture

## CONTENU DU COURS

### Introduction

- Présentation de Informatica Developer Tool 10.1 BigData Management
- Architecture de Big Data Management, présentation de l'interface
- Création d'objets, de connexions vers des bases de données relationnelles, vers des fichiers, Visualisation de données, mise en œuvre des fonctions de mapping et transformations.

### Modélisation

- Présentation des outils : Business Modeler, JobDesigner
- Mise en œuvre des principales connexions.
- Intégration de fichiers XML et CSV
- Etude des composants de transformation
- Analyse du code et exécution des jobs.

### Optimisation des jobs

- Utilisation des métadonnées, import/export, propagation sur les jobs, configuration de connexions réutilisables
- Stockage des variables de contexte dans les fichiers properties et .ini

### Liens avec les bases de données

- Présentation des bases de données supportées
- Opérations sur les tables, Connexion à un schéma de bases de données
- Gestion des transactions
- Utilisation de SQLBuilder pour créer des requêtes

### Big Data

- Concepts Hadoop, architecture
- Présentation HDFS, YARN, Traitements distribués avec MAP Reduce

### Injection et export de données

- Mise en œuvre de PowerCenter

- Utilisation de SQOOP pour migrer des données vers Hadoop
- Partitionnement et parallélisme

### **Architecture de Big Data Management**

- Caractéristiques BigData, couche d'abstraction Informatica

### **Différents liens vers Hadoop**

- Fonctionnalités des interfaces Hive MR/Tez, Blaze, Spark, Smart Executor

### **Supervision et diagnostic**

- Mise en œuvre des outils de supervision et diagnostic : Plans d'exécution, monitoring mapping, et troubleshooting

### **Qualité des données**

- Description du process de nettoyage des données
- Gestion de la qualité des données

### **Parsing de fichiers complexes**

- Mise en œuvre avec Avro, Parquet, JSON

### **Accès à des bases de données NoSQL**

- Présentation du théorème de CAP
- Exemples avec HBase, MongoDB, Cassandra