

## BigData, informatica v10

Durée: 3 jours

1950 €

12 au 14 mars  
18 au 20 juin

17 au 19 septembre  
10 au 12 décembre

### Public:

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

### Objectifs:

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

### Connaissances préalables nécessaires:

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

### Programme:

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 oeuvre des fonctions de mapping et tranformations.

Modélisation : Présentation des outils : Business Modeler, JobDesigner  
Mise en oeuvre 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

## BigData, informatica v10

- 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 oeuvre 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 oeuvre 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 oeuvre avec Avro, Parquet, JSON
- Acces à des bases de données NoSQL** : Présentation du théorème de CAP  
Exemples avec HBase, MongoDB, Cassandra