

KAİNA-COM TRAINING CATALOGUE

Le Big Data



KDS009 – Le Big Data

Référence KDS009

Niveau

☒ Débutant
☐ Intermédiaire
☐ Expert

Nombre de jours Programme de formation :

- 16 heures (4H/jour)

Lieu de la formation

☐ I: e-learning, Formation individuelle (Formation en ligne)
☒ V: v-learning, classe virtuelle
☐ C: c-learning, cours présentiel

KAÏNA-COM

LE CARRÉ HAUSSMANN II,
6 Allée de la Connaissance
77127 Lieusaint - France

Prérequis Un niveau d'anglais business moyen est requise car la formation sera dispensée en anglais.

Public Chefs de projet, chefs de produit et Managers, Développeurs et Architectes qui veulent connaître le Big Data.

Continued on next page



KDS009 – Le Big Data, continued

Objectifs

Aujourd'hui, les entreprises ont la capacité de collecter une grande quantité de données. Le traitement d'une grande quantité de données nécessite de nouvelles technologies capables de collecter, de nettoyer, de traiter et de stocker efficacement une quantité importante d'informations.

De nombreuses entreprises sont parvenues à la conclusion que le fait de ne pas utiliser ces données collectées entraîne en réalité une perte importante d'argent. Le marché du Big Data devrait dépasser les 200 milliards de dollars cette année.

Ce cours fournit la base de l'environnement, de l'architecture, du processus et des outils disponibles Big Data et NoSQL DB. Le cours présentera également les méthodologies Big Data et les recommandations de déploiement.

Continued on next page



KDS009 – Le Big Data, continued

Course Contents

Course Contents :

Table 1: KDS009 - Course Contents

Chapter	Description
Introduction	• Definition: Big Data, NoSQL • The need for Big Data technology • Tradition technologies Vs Big Data technologies • Big Data project requirements • Big Data Project workflow
Big Data Architecture	• Big Data project definitions • Data sources & development resources • Big Data technologies evaluation – The need for POC
Data Collection & Ingestion	• Streaming Concept – Rest API • Apache Kafka – AWS Kinesis, Azure Event Hub • Apache Flume • ELK package – Logz.io
Hadoop – Introduction	• What is Hadoop? • Hadoop Architecture • Hadoop File System (HDFS) – Architecture – NameNode & DataNode • Hadoop MapReduce • Apache YARN • Apache Oozie, ZooKeeper • Project non-functional support – Sentry, Tez, Ambri, Knox, Falcon

Continued on next page



KDS009 – Le Big Data, continued

Course Contents, continued

Chapter	Description
Project decision – Hadoop deployment	• Hadoop Distribution – Examples: Cludera, Hortonworks • Hadoop as a service • Can Big Data project switch environments? • Hadoop deployment requirements • Hadoop Performance Best Practices
Hadoop project POC	• POC environment • Using Apache Pig! & Apache Sqoop for POC
Large-scale data processing framework	• Apache Storm • Apache Spark – Concept & Architecture – Programming with Spark – Spark Streaming – Spark SQL – MLlib – GraphX
Project development cycle & deployment – Spark	• Big Data – Development methodologies – Waterfall Vs Agile • ETL development cycle & deployment • Tests Cycle

Continued on next page



KDS009 – Le Big Data, continued

Course Contents, continued

Chapter	Description
Big Data DB types	• Key-Values Stores – Redis • Column Family Stores (Wide Column Stores) – Apache HBase – Apache Cassandra • Document Databases – MongoDB • Graph Databases – Mathematical Graph as a DB
Big Data & RDBMS	• Product logic • Apache Hive – Architecture – Batch Processing • Apache Impala – Massively Parallel Processing (MPP)
Big Data Northbound Interfaces	• Big Data to OLAP • BI Visualization • Scaling BI over Big Data
Big Data – system ATP	• Big Data – system ATP
The End	• Trends & Conclusions • Q&A • Course's Evaluation

