

# KAÏNA-COM TRAINING CATALOGUE

## Big Data: modélisation, technologie et architecture

---



## KDS008 – Big Data: Modeling, Technology and Architecture

**Référence** KDS008

**Niveau**

- Débutant
- Intermédiaire
- Expert

**Nombre de jours** Programme de formation :

- 24 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** Développeurs, architectes, chefs de produit et responsables souhaitant en savoir davantage sur le Big Data.

*Continued on next page*



## **KDS008 – Big Data: Modeling, Technology and Architecture,** 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*



## KDS008 – Big Data: Modeling, Technology and Architecture, Continued

### Course Contents

#### Course Contents :

Table 1: KDS008 - Course Contents

Chapter	Description
<b>Introduction</b>	<ul style="list-style-type: none"> <li>• Definition: Big Data, NoSQL</li> <li>• The need for Big Data technology</li> <li>• Tradition technologies Vs Big Data technologies</li> </ul>
<b>Big Data Architecture</b>	<ul style="list-style-type: none"> <li>• Big Data Architecture</li> </ul>
<b>Data Collection &amp; Ingestion</b>	<ul style="list-style-type: none"> <li>• Streaming Concept               <ul style="list-style-type: none"> <li>– Rest API</li> </ul> </li> <li>• Apache Kafka               <ul style="list-style-type: none"> <li>– AWS Kinesis, Azure Event Hub</li> </ul> </li> <li>• Apache Flume</li> <li>• Log Stash</li> <li>• Commercial solutions – Splunk, Logz.io</li> </ul>
<b>Hadoop</b>	<ul style="list-style-type: none"> <li>• What is Hadoop?</li> <li>• Hadoop Architecture</li> <li>• Hadoop File System (HDFS)               <ul style="list-style-type: none"> <li>– Architecture</li> <li>– NameNode &amp; DataNode</li> </ul> </li> <li>• Hadoop MapReduce</li> <li>• Apache YARN</li> <li>• Apache Oozie, Sentry, Tez, HCatalog, ZooKeeper, Ambri, Knox, Falcon</li> <li>• Hadoop Distribution               <ul style="list-style-type: none"> <li>– Examples: Cloudera, Hortonworks</li> </ul> </li> <li>• Hadoop Performance Best Practices</li> </ul>
<b>Apache Pig</b>	<ul style="list-style-type: none"> <li>• Apache Pig</li> </ul>

*Continued on next page*



## KDS008 – Big Data: Modeling, Technology and Architecture, Continued

### Course Contents, continued

Chapter	Description
<b>Apache Storm</b>	<ul style="list-style-type: none"> <li>• Apache Storm</li> </ul>
<b>Apache Spark</b>	<ul style="list-style-type: none"> <li>• Concept &amp; Architecture</li> <li>• Programming with Spark</li> <li>• Spark Streaming</li> <li>• Spark SQL, Datasets, and DataFrames</li> <li>• MLlib</li> <li>• GraphX</li> </ul>
<b>Big Data DB types</b>	<ul style="list-style-type: none"> <li>• Big Data DB types</li> </ul>
<b>Key-Values Stores</b>	<ul style="list-style-type: none"> <li>• Redis</li> </ul>
<b>Column Family Stores (Wide Column Stores)</b>	<ul style="list-style-type: none"> <li>• Apache HBase</li> <li>• Apache Cassandra</li> </ul>
<b>Document Databases</b>	<ul style="list-style-type: none"> <li>• MongoDB                             <ul style="list-style-type: none"> <li>– Architecture &amp; Data Model</li> <li>– JSON query</li> <li>– Performance Best Practices</li> </ul> </li> </ul>
<b>Graph Databases</b>	<ul style="list-style-type: none"> <li>• Mathematical Graph as a DB</li> <li>• Architecture and components</li> </ul>
<b>'SQL' over Hadoop</b>	<ul style="list-style-type: none"> <li>• Apache Pig!</li> <li>• Apache Sqoop</li> <li>• Apache Hive                             <ul style="list-style-type: none"> <li>– Architecture – Batch Processing</li> <li>– Apache Impala</li> <li>– Massively Parallel Processing (MPP)</li> </ul> </li> </ul>

*Continued on next page*



## KDS008 – Big Data: Modeling, Technology and Architecture, continued

---

### Course Contents, continued

Chapter	Description
<b>Big Data Deployment</b>	<ul style="list-style-type: none"><li>• Local Data Center</li><li>• Hosting Services<ul style="list-style-type: none"><li>– AWS, Azure, Google</li></ul></li><li>• Pros and Cons</li></ul>
<b>Big Data Northbound Interfaces</b>	<ul style="list-style-type: none"><li>• Big Data to OLAP</li><li>• BI Visualization</li><li>• Scaling BI over Big Data</li></ul>
<b>The End</b>	<ul style="list-style-type: none"><li>• Trends &amp; Conclusions</li><li>• Q&amp;A</li><li>• Course's Evaluation</li></ul>

---

