

# KAÏNA-COM TRAINING CATALOGUE

## Big Data

---



**Nos locaux**  
KAÏNA-COM France  
LE CARRÉ HAUSSMANN II  
6 Allée de la Connaissance  
77 127 Lieusaint



**Contact**  
+33(0)9 50 20 91 64



**E-mail**  
info@kaina-com.fr



**Site Internet**  
www.kaina-com.fr

## KDS009 – Big Data

---

**Reference** KDS009

---

**Experience**  Beginner  
 Intermediate  
 Advanced

---

**Duration** Training Program:  
• 2 days

---

**Training Method**  I: i-learning, individual training (web-based training)  
 V: v-learning, virtual class  
 C: c-learning, classroom training  
**KAÏNA-COM**  
LE CARRÉ HAUSSMANN II,  
6 Allée de la Connaissance  
77127 Lieusaint - France

---

**Price** 1.390,50 € HT

---

**Prerequisite** None

---

**Audience** Project Managers, Product people and Managers, Developers and Architects who wants to know about Big Data.

---

*Continued on next page*



## **KDS009 – Big Data, Continued**

### **Objective**

Today companies have the capability to collect large amount of data. Handling large amount of data requires new technologies that are able to collect, cleanse, process and store effectively significant amount of information.

Many companies reached the conclusion that not using this collected data is actually losing large amount of money. Big Data market is estimated to surpass \$200 billion this year.

There is a tremendous business in Big Data and with the right methodologies and tools this row Data can be available for use. This course provide the basis for Big Data and NoSQL DB environment, architecture, process and available tools. The course will also present Big Data methodologies and deployment recommendations

---

*Continued on next page*



## KDS009 – Big Data, Continued

### Course Contents

#### Course Contents :

**Table 1: KDS009 - 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> <li>• Big Data project requirements</li> <li>• Big Data Project workflow</li> </ul>
<b>Big Data Architecture</b>	<ul style="list-style-type: none"> <li>• Big Data project definitions</li> <li>• Data sources &amp; development resources</li> <li>• Big Data technologies evaluation               <ul style="list-style-type: none"> <li>– The need for POC</li> </ul> </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>• ELK package               <ul style="list-style-type: none"> <li>– Logz.io</li> </ul> </li> </ul>
<b>Hadoop – Introduction</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, ZooKeeper</li> <li>• Project non-functional support               <ul style="list-style-type: none"> <li>– Sentry, Tez, Ambri, Knox, Falcon</li> </ul> </li> </ul>

*Continued on next page*



## KDS009 – Big Data, Continued

### Course Contents, continued

Chapter	Description
<b>Project decision – Hadoop deployment</b>	<ul style="list-style-type: none"><li>• Hadoop Distribution<ul style="list-style-type: none"><li>– Examples: Cloudera, Hortonworks</li></ul></li><li>• Hadoop as a service</li><li>• Can Big Data project switch environments?</li><li>• Hadoop deployment requirements</li><li>• Hadoop Performance Best Practices</li></ul>
<b>Hadoop project POC</b>	<ul style="list-style-type: none"><li>• POC environment</li><li>• Using Apache Pig! &amp; Apache Sqoop for POC</li></ul>
<b>Large-scale data processing framework</b>	<ul style="list-style-type: none"><li>• Apache Storm</li><li>• Apache Spark<ul style="list-style-type: none"><li>– Concept &amp; Architecture</li><li>– Programming with Spark</li><li>– Spark Streaming</li><li>– Spark SQL</li><li>– MLlib</li><li>– GraphX</li></ul></li></ul>
<b>Project development cycle &amp; deployment – Spark</b>	<ul style="list-style-type: none"><li>• Big Data – Development methodologies<ul style="list-style-type: none"><li>– Waterfall Vs Agile</li></ul></li><li>• ETL development cycle &amp; deployment</li><li>• Tests Cycle</li></ul>

*Continued on next page*



## KDS009 – Big Data, continued

### Course Contents, continued

Chapter	Description
<b>Big Data DB types</b>	<ul style="list-style-type: none"> <li>• Key-Values Stores               <ul style="list-style-type: none"> <li>– Redis</li> </ul> </li> <li>• Column Family Stores (Wide Column Stores)               <ul style="list-style-type: none"> <li>– Apache HBase</li> <li>– Apache Cassandra</li> </ul> </li> <li>• Document Databases               <ul style="list-style-type: none"> <li>– MongoDB</li> </ul> </li> <li>• Graph Databases               <ul style="list-style-type: none"> <li>– Mathematical Graph as a DB</li> </ul> </li> </ul>
<b>Big Data &amp; RDBMS</b>	<ul style="list-style-type: none"> <li>• Product logic</li> <li>• Apache Hive               <ul style="list-style-type: none"> <li>– Architecture – Batch Processing</li> </ul> </li> <li>• Apache Impala               <ul style="list-style-type: none"> <li>– Massively Parallel Processing (MPP)</li> </ul> </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>Big Data – system ATP</b>	<ul style="list-style-type: none"> <li>• Big Data – system ATP</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>

