

## Hadoop voor Big Data

### Doelgroep Cursus Hadoop voor Big Data

De cursus [Hadoop voor Big Data](#) is bedoeld voor developers, data analisten en anderen die willen leren met hoe je data kunt verwerken met Hadoop.

### Voorkennis training Hadoop voor Big Data

Om aan deze cursus deel te nemen is kennis van [programmeren in Java](#) en databases bevorderlijk voor de begripsvorming. Voorafgaande kennis van Java of Hadoop is niet nodig.

### Uitvoering cursus Hadoop voor Big Data

De theorie wordt behandeld aan de hand van presentaties. Illustratieve demo's worden gebruikt om de behandelde concepten te verduidelijken. Er is voldoende gelegenheid om te oefenen en afwisseling van theorie en praktijk. De cursustijden zijn van 9.30 tot 16.30.

### Certificering cursus Hadoop voor Big Data

De deelnemers krijgen na het goed doorlopen van de cursus een officieel certificaat Hadoop voor Big Data.

## Inhoud Cursus Hadoop voor Big Data

In de cursus Hadoop voor Big Data leren de deelnemers Apache Hadoop te gebruiken voor de opslag en verwerking van grote hoeveelheden data.

### Hadoop Architectuur

In de cursus Hadoop voor Big Data Hadoop komt de architectuur van Hadoop uitgebreid aan de orde. Hadoop gebruikt een eenvoudig programmeer model in een gedistribueerde omgeving over een cluster van computers.

### HDFS

Binnen een Hadoop cluster wordt het Hadoop Distributed File System (HDFS) gebruikt als bestandssysteem. In de cursus Hadoop voor Big Data Hadoop wordt HDFS uitgebreid besproken. HDFS is een horizontaal schaalbaar bestandssysteem dat opgeslagen staat op een cluster van servers. De data wordt gedistribueerd opgeslagen, en het bestandssysteem zorgt automatisch voor replicatie van data over het cluster.

### MapReduce

Een belangrijk algoritme voor het verwerken van Data is het MapReduce algorithm en hier wordt uitgebreid aandacht aan besteed.

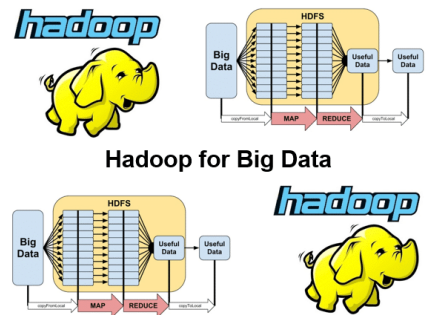
### Utilities

Tenslotte is er in de cursus Hadoop voor Big Data Hadoop aandacht voor tools en utilities die vaak in combinatie met Hadoop worden gebruikt zoals Zookeeper, Scoop, Ozie en Pig.

Duur: 3 dagen

Prijs: € 1999

[Open Rooster](#)



## Modules Cursus Hadoop voor Big Data

Module 1 : Hadoop Intro	Module 2 : Java API	Module 3 : HDFS
Big Data Handling No SQL Comparison to Relational DB Hadoop Eco-System Hadoop Distributions Pseudo-Distributed Installation Namenode Safemode Namenode High Availability Secondary Namenode Hadoop Filesystem Shell	Create via Put method Read via Get method Update via Put method Delete via Delete method Create Table Drop Table Scan API Scan Caching Scan Batching Filters	Hadoop Environment Hadoop Stack Hadoop Yarn Distributed File System HDFS Architecture Parallel Operations Working with Partitions RDD Partitions HDFS Data Locality DAG (Direct Acyclic Graph)
Module 4 : Hbase Key Design	Module 5 : MapReduce	Module 6 : Submitting Jobs
Storage Model Querying Granularity Table Design Tall-Narrow Tables Flat-Wide Tables Column Family Column Qualifier Storage Unit Querying Data by Timestamp Querying Data by Row-ID Types of Keys and Values SQL Access	MapReduce Model MapReduce Theory YARN and MapReduce 2.0 Daemons MapReduce on YARN single node MapReduce framework Tool and ToolRunner GenericOptionsParser Running MapReduce Locally Running MapReduce on Cluster Packaging MapReduce Jobs MapReduce CLASSPATH Decomposing into MapReduce	MapReduce Job Using JobControl class Joining data-sets User Defined Functions Logs and Web UI Input and Output Formats Anatomy of Mappers Reducers and Combiners Partitioners and Counters Speculative Execution Distributed Cache YARN Components
Module 7 : Hadoop Streaming	Module 8 : Utilities	Module 9 : Hive
Implement a Streaming Job Contrast with Java Code Create counts in Streaming App Text Processing Use Case Key Value Pairs \$yarn command Using Pipes	ZooKeeper Scoop Introduce Oozie Deploy and Run Oozie Workflow Pig Overview Execution Modes Developing Pig Script	Hive Concepts Hive Clients Table Creation and Deletion Loading Data into Hive Partitioning Bucketing Joins