

# Opleiding : Talend Open Studio, data-integratie voor big data

Gemeenschapsversie (gratis en open source)

*Praktijkcursus - 3d - 21u00 - Ref. IDB*

*Prijs : 2010 € V.B.*

Talend's data-integratieplatform breidt zijn mogelijkheden uit naar big data technologieën zoals Hadoop (HDFS, HBase, HCatalog, Hive en Pig) en de NoSQL databases Cassandra en MongoDB. Deze cursus biedt je de basis voor het gebruik van de Talend componenten die gemaakt zijn om te communiceren met big data systemen. Deze cursus gaat uitsluitend over Talend Open Studio (community, gratis en open source versie). De commerciële versie van Talend Studio, onder betaalde licentie, geïntegreerd in het Qlik-Talend Cloud portaal, wordt niet behandeld.

## Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✔ Talend beheersen in een big data-omgeving
- ✔ Talend gebruiken als koppeling tussen bestanden, toepassingen en databases
- ✔ Leer de filosofie van het gereedschap
- ✔ Beste praktijken toepassen en flexibele, robuuste informatiesystemen ontwerpen
- ✔ Je taken kunnen uitvoeren
- ✔ Gegevens lezen en schrijven naar HDFS en NoSQL databases met Talend jobs
- ✔ Transformatietaken maken met behulp van Pig en Hive
- ✔ Datakwaliteit beheren met Talend
- ✔ Sqoop gebruiken om de migratie van relationele databases naar Hadoop te vergemakkelijken
- ✔ Het gebruik van de componentenbibliotheek onder de knie krijgen
- ✔ Eenvoudige en complexe end-to-end ETL-processen (Extract, Transform en Load) uitvoeren

## Doelgroep

BI-consultants, architecten, projectmanagers, gegevensbeheerders of iedereen die gegevensstromen moet beheren.

### DEELNEMERS

BI-consultants, architecten, projectmanagers, gegevensbeheerders of iedereen die gegevensstromen moet beheren.

### VOORAFGAANDE VEREISTEN

Kennis van Hadoop, Spark en Kafka.

### VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

### BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ...  
De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

## Voorafgaande vereisten

Kennis van Hadoop, Spark en Kafka.

## Praktische modaliteiten

Opeenvolging van miniprojecten die leiden tot het ontwerp van Talend big data jobs van toenemende moeilijkheidsgraad.

## Opleidingsprogramma

### 1 Presentatie Talend Open Studio

- Gegevensintegratie. ETL-oplossingen.
- Big data. Ongestructureerde gegevens. NoSQL databases.
- Het Hadoop-ecosysteem (HDFS, MapReduce, HBase, Hive, Pig, enz.).
- TOS voor gegevensintegratie: gegevensintegratie.
- TOS voor gegevenskwaliteit: beheer van gegevenskwaliteit.
- TOS voor big data.
- Productfilosofie.

#### Praktisch werk

Installatie/configuratie van TOS voor big data. Aan de slag.

### 2 Banen ontwerpen

- Inleiding tot bedrijfsmodellering en functieontwerp.
- Eenvoudige transformatiecomponenten.
- Gegeneerde code bekijken, een taak uitvoeren.
- Taken instellen.
- Creëer en beheer je eigen variabelen.
- Goede ontwerppraktijken.

#### Praktisch werk

Ontwikkeling van een taak die verbinding maakt met een gegevensbron, filtering, transformatie en opslag van het resultaat in een bestand.

### 3 Gegevensintegratie in een cluster en NoSQL-databases

- Definitie van metagegevens over Hadoop-clusterverbindingen.
- Maak verbinding met een MongoDB, Neo4j, Cassandra of Hbase database en exporteer gegevens.
- Eenvoudige integratie van gegevens met een Hadoop-cluster.
- Presentatie van uitbreidingscomponenten.
- Utilisation du composant d'extension : capture de tweets et importation directe dans HDFS.

#### Praktisch werk

Lees tweets en sla ze op als bestanden in HDFS, analyseer de frequentie van de besproken thema's en sla de resultaten op in HBase.

## PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

## TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

## TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.

#### 4 Importeren/exporteren met Sqoop

- Gebruik Sqoop om gegevens te importeren, exporteren en bij te werken tussen RDBMS- en HDFS-systemen.
- Gedeeltelijke, incrementele import/export van tabellen.
- Een SQL-database importeren/exporteren van en naar HDFS.
- Big data opslagformaten (AVRO, Parquet, ORC, enz.).

##### Praktisch werk

Relationele tabellen migreren naar HDFS en omgekeerd.

#### 5 Gegevens manipuleren

- Presentatie van de Pig brick en de bijbehorende PigLatin taal.
- Belangrijkste Talend Pig componenten, Pig flow ontwerp.
- Ontwikkeling van UDF-routines.

##### Praktisch werk

Trends in het gebruik van een website identificeren door de logbestanden te analyseren.

#### 6 Architectuur en best practices in een Hadoop-cluster

- Efficiënte opslag in Hadoop ontwerpen.
- Data lake versus data warehouse: moet je kiezen?
- Hadoop en het Disaster Recovery Plan (DRP) in het geval van een groot incident.
- Automatiseer je workflows.

##### Praktisch werk

Creëer je data lake en automatiseer de werking ervan.

#### 7 Analyseer en sla uw gegevens op met Hive

- Hive-verbinding en schema metagegevens.
- De taal HiveQL.
- Ontwerp van Hive-stromen, uitvoeren van query's.
- De ELT-componenten van Hive implementeren.

##### Praktisch werk

Sla de trend van de aandelenkoersen op in HBase en consolideer deze stroom met Hive om de trend per uur voor een bepaalde dag weer te geven.

## Data en plaats

**KLAS OP AFSTAND**  
2026 : 29 juni, 18 nov.

**PARIS LA DÉFENSE**  
2026 : 22 juni, 4 nov.