Designing and Building Big Data Applications

Category: AI, Data & Analytics | **Vendor:** Various

Duration: 32.00 hours (4 days) **26.0 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Classroom

Course Overview

Cloudera University's four-day course for designing and building Big Data applications prepares you to analyze and solve real-world problems using Apache Hadoop and associated tools in the enterprise data hub (EDH).

About This Course

Cloudera University's four-day course for designing and building Big Data applications prepares you to analyze and solve real-world problems using Apache Hadoop and associated tools in the enterprise data hub (EDH).

Who Should Attend

This course is best suited to developers, engineers, and architects who want to use use Hadoop and related tools to solve real-world problems.

Learning Outcomes

Upon successful completion of this course, participants will be able to:

Skills learned in this course include:Creating a data set with Kite SDKDeveloping custom Flume components for data ingestionManaging a multi-stage workflow with OozieAnalyzing data with CrunchWriting user-defined functions for Hive and ImpalaWriting user-defined functions for Hive and ImpalaIndexing data with Cloudera Search

Detailed Course Outline

IntroductionApplication Architecture

- Scenario Explanation
- Understanding the Development Environment
- Identifying and Collecting Input Data
- Selecting Tools for Data Processing and Analysis
- Presenting Results to the Use

Defining & Using Datasets

- Metadata Management
- What is Apache Avro
- Avro Schemas
- Avro Schema Evolution
- Selecting a File Format
- Performance Considerations

Using the Kite SDK Data Module

- What is the Kite SDK
- Fundamental Data Module Concepts
- Creating New Data Sets Using the Kite SDK
- Loading, Accessing, and Deleting a Data Set

Importing Relational Data with Apache Sqoop

- What is Apache Sqoop
- Basic Imports
- Limiting Results
- Improving Sqoops Performance
- Sqoop 2

Capturing Data with Apache Flume

- What is Apache Flume
- Basic Flume Architecture
- Flume Sources
- Flume Sinks
- Flume Configuration
- Logging Application Events to Hadoop

Developing Custom Flume Components

- Flume Data Flow and Common Extension Points
- Custom Flume Sources
- Developing a Flume Pollable Source
- Developing a Flume Event-Driven Source
- Custom Flume Interceptors
- Developing a Header-Modifying Flume Interceptor
- Developing a Filtering Flume Interceptor
- Writing Avro Objects with a Custom Flume Interceptor

Managing Workflows with Apache Oozie

- The Need for Workflow Management
- What is Apache Oozie
- Defining an Oozie Workflow
- Validation, Packaging, and Deployment
- Running and Tracking Workflows Using the CLI
- Hue UI for Oozie

Processing Data Pipelines with Apache Crunch

- What is Apache Crunch
- Understanding the Crunch Pipeline
- Comparing Crunch to Java MapReduce
- Working with Crunch Projects
- Reading and Writing Data in Crunch
- Data Collection API Functions
- Utility Classes in the Crunch API

Working with Tables in Apache Hive

- What is Apache Hive
- Accessing Hive
- Basic Query Syntax
- Creating and Populating Hive Tables
- How Hive Reads Data
- Using the RegexSerDe in Hive

Developing User-Defined Functions

- What are User-Defined Functions
- Implementing a User-Defined Function
- Deploying Custom Libraries in Hive
- Registering a User-Defined Function in Hive

Executing Interactive Queries with Impala

- What is Impala
- Comparing Hive to Impala
- Running Queries in Impala
- Support for User-Defined Functions
- Data and Metadata Management

Understanding Cloudera Search

- What is Cloudera Search
- Search Architecture
- Supported Document Formats

Indexing Data with Cloudera Search

- Collection and Schema Management
- Morphlines
- Indexing Data in Batch Mode
- Indexing Data in Near Real Time

Presenting Results to Users

- Solr Query Syntax
- Building a Search UI with Hue
- Accessing Impala through JDBC
- Powering a Custom Web Application with Impala and Search

Additional Course Details

Nexus Humans Designing and Building Big Data Applications training program is a workshop that presents an invigorating mix of sessions, lessons, and masterclasses meticulously crafted to propel your learning expedition forward.

This immersive bootcamp-style experience boasts interactive lectures, hands-on labs, and collaborative hackathons, all strategically designed to fortify fundamental concepts.

Guided by seasoned coaches, each session offers priceless insights and practical skills crucial for honing your expertise. Whether you're stepping into the realm of professional skills or a seasoned professional, this comprehensive course ensures you're equipped with the knowledge and prowess necessary for success.

While we feel this is the best course for the Designing and Building Big Data Applications course and one of our Top 10 we encourage you to read the course outline to make sure it is the right content for you.

Additionally, private sessions, closed classes or dedicated events are available both live online and at our training centres in Dublin and London, as well as at your offices anywhere in the UK, Ireland or across EMEA.

Frequently Asked Questions

Q: What delivery options are available for Designing and Building Big Data Applications?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 4-day Designing and Building Big Data Applications course provides up to 26.0 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Designing and Building Big Data Applications training?

The training takes place over 4 day(s), with each day lasting approximately 32.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Designing and Building Big Data Applications?

Yes, we provide corporate training, dedicated training, and closed classes for Designing and Building Big Data Applications. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: Why choose Nexus Human for Designing and Building Big Data Applications?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Designing and Building Big Data Applications training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 14/11/2026.