

Training: Cloudera
Cloudera Search Training


FORM OF TRAINING	MATERIALS	PRICE	DURATION
Traditional	Digital materials	1780 EUR	3 days
Traditional	CTAB Tablet	1930 EUR	3 days
Distance learning	Digital materials	1780 EUR	3 days
Distance learning	CTAB Tablet	1780 EUR	3 days

LOCATIONS

Krakow - 5 Tatarska Street, II floor, hours: 9:00 am - 4:00 pm

Warsaw - 17 Bielska Street, hours: 9:00 am - 4:00 pm

TRAINING GOALS:

Cloudera University's three-day **Search training** course is for developers and data engineers who want to index data in Hadoop for more powerful real-time queries. Participants will learn to get more value from their data by integrating Cloudera Search with external applications. Cloudera Search brings full-text, interactive search and scalable, flexible indexing to Hadoop and an enterprise data hub. Powered by Apache Solr, Search delivers scale and reliability for a new generation of integrated, multi-workload queries.

Through instructor-led discussion and interactive, hands-on exercises, participants will learn Apache Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem, learning topics such as:

- Perform batch indexing of data stored in HDFS and HBase
- Perform indexing of streaming data in near-real-time with Flume
- Index content in multiple languages and file formats
- Process and transform incoming data with Morphlines
- Create a user interface for your index using Hue
- Integrate Cloudera Search with external applications
- Improve the Search experience using features such as faceting, highlighting, spelling Correction

CONSPECT:

- Introduction
- Overview of Cloudera Search

- What is Cloudera Search?
- Helpful Features
- Use Cases
- Basic Architecture
- Performing Basic Queries
 - Executing a Query in the Admin UI
 - Basic Syntax
 - Techniques for Approximate Matching
 - Controlling Output
- Writing More Powerful Queries
 - Relevancy and Filters
 - Query Parsers
 - Functions
 - Geospatial Search
 - Faceting
- Preparing to Index Documents
 - Overview of the Indexing Process
 - Understanding Morphlines
 - Generating Configuration Files
 - Schema Design
 - Collection Management
- Batch Indexing HDFS Data with MapReduce
 - Overview of the HDFS Batch Indexing Process
 - Using the MapReduce Indexing Tool
 - Testing and Troubleshooting
- Near-Real-Time Indexing with Flume
 - Overview of the Near-Real-Time Indexing Process
 - Introduction to Apache Flume
 - How to Perform Near-Real-Time Indexing with Flume
 - Testing and Troubleshooting
- Indexing HBase Data with Lily
 - What is Apache HBase?
 - Batch Indexing for HBase
 - Indexing HBase Tables in Near-Real-Time
- Indexing Data in Other Languages and Formats
 - Field Types and Analyzer Chains

- Word Stemming, Character Mapping, and Language Support
- Schema and Analysis Support in the Admin UI
- Metadata and Content Extraction with Apache Tika
- Indexing Binary File Types with SolrCell
- Improving Search Quality and Performance
 - Delivering Relevant Results
 - Helping Users Find Information
 - Query Performance and Troubleshooting
- Building User Interfaces for Search
 - Search UI Overview
 - Building a User Interface with Hue
 - Integrating Search into Custom Applications
- Considerations for Deployment
 - Planning for Deployment
 - Determining Hardware Needs
 - Security Overview
 - Collection Aliasing
- Conclusion

REQUIREMENTS:

This course is intended for developers and data engineers with at least basic familiarity with Hadoop and experience programming in a general-purpose language such as [Java](#), C, C++, Perl, or [Python](#). Participants should be comfortable with the Linux command line and should be able to perform basic tasks such as creating and removing directories, viewing and changing file permissions, executing scripts, and examining file output. No prior experience with Apache Solr or Cloudera Search is required, nor is any experience with HBase or SQL.

Difficulty level



CERTIFICATE:

The participants will obtain certificates signed by Cloudera (training completion).

TRAINER:

Certified Cloudera Instructor.