# Cloudera - Cloudera Search Training

| | |
|---|---|
| **Code:** | CLD-SEARCH |
| **Length:** | 3 days |
| **URL:** | View Online |

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.

## Skills Gained

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

## Prerequisites

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.

# Course Details

## Outline

- Introduction Overview of Cloudera Search

- Performing Basic Queries

- Writing More Powerful Queries

- Preparing to Index Documents

- Batch Indexing HDFS Data with MapReduce

- Near-Real-Time Indexing with Flume

- Indexing HBase Data with Lily

- Indexing Data in Other Languages and Formats

- Improving Search Quality and Performance

- Building User Interfaces for Search

- Considerations for Deployment

- Conclusion

---