

## About Us

### About Dilişim

Dilişim was founded in 2009 by Dr. Özgür Yılmazel who has a PhD in natural language processing and information extraction. Dilişim has expertise in Big Data Systems, Natural Language Processing and Search. Dilişim's vision and goal is to support its clients and create measurable value to its customers by utilizing data at their hand. Dilişim is Cloudera's first and only training partner in Turkey and also the only silver-level integrator partner in Turkey since 2012. Dilişim deployed first commercial Hadoop Cluster in Turkey, and it now runs the largest Hadoop Cluster in Turkey.

### About Cloudera

Founded in 2008, Cloudera was the first, and is currently, the leading provider and supporter of Apache Hadoop for the enterprise. Cloudera also offers software for business critical data challenges including storage, access, management, analysis, security, and search. Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data: The Enterprise Data Hub.

### What are Cloudera trainings?

Dilişim offers the following Cloudera trainings:

- › Cloudera Developer Training for Spark and Hadoop (4 days)
- › Cloudera Developer Training for Apache Spark (3 days)
- › Cloudera Administrator Training for Apache Hadoop (4 days)
- › Cloudera Data Analyst Training: Using Pig, Hive and Impala with Hadoop (4 days)
- › Cloudera Training for Apache HBase (3 days)

The trainings in Turkey are delivered by Dilişim by being the only training partner of Cloudera in Turkey.

### Why Cloudera Training?

Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem and experience the following:

- › Most comprehensive suite of courses to address the Hadoop objectives of every data professional: developers, administrators, and data analysts.
- › The industry's only truly dynamic and up-to-date Hadoop training curriculum
- › Delivered by full-time technical and Cloudera certified instructors
- › Industry leader in Hadoop with over 100.000 participants
- › Video tutorials and e-learning services





# Cloudera Training for Apache HBase

## Take your knowledge to the next level with Cloudera's Apache Hadoop Training and Certification

This three-day training course for Apache HBase enables participants to store and access massive quantities of multi-structured data and perform hundreds of thousands of operations per second.

### Advance Your Ecosystem Expertise

Apache HBase is a distributed, scalable, NoSQL database built on Apache Hadoop. HBase can store data in massive tables consisting of billions of rows and millions of columns, serve data to many users and applications in real time, and provide fast, random read/write access to users and applications.

### Hands-On Hadoop

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

- › The use cases and usage occasions for HBase, Hadoop, and RDBMS
- › Using the HBase shell to directly manipulate HBase tables
- › Designing optimal HBase schemas for efficient data storage and recovery
- › How to connect to HBase using the Java API to insert and retrieve data in real time
- › Best practices for identifying and resolving performance bottlenecks

### Audience & Prerequisites

This course is appropriate for developers and administrators who intend to use HBase. Prior experience with databases and data modeling is helpful, but not required. Knowledge of Java is assumed. Prior knowledge of Hadoop is not required, but Cloudera Developer Training for Apache Hadoop provides an excellent foundation for this course.

### HBase Certification

Upon completion of the course, attendees are encouraged to continue their study and register for the Cloudera Certified Specialist in Apache HBase (CCSHB) exam. Certification is a great differentiator; it helps establish you as a leader in the field, providing employers and customers with tangible evidence of your expertise.





## Cloudera Training for Apache HBase

---

### Introduction

#### Introduction to Hadoop and HBase

- › What Is Big Data?
- › Introducing Hadoop
- › Hadoop Components
- › What Is HBase?
- › Why Use HBase?
- › Strengths of HBase
- › HBase in Production
- › Weaknesses of HBase

#### HBase Tables

- › HBase Concepts
- › HBase Table Fundamentals
- › Thinking About Table Design

#### The HBase Shell

- › Creating Tables with the HBase Shell
- › Working with Tables
- › Working with Table Data

#### HBase Architecture

##### Fundamentals

- › HBase Regions
- › HBase Cluster Architecture
- › HBase and HDFS Data Locality

##### HBase Schema Design

- › General Design Considerations
- › Application-Centric Design
- › Designing HBase Row Keys
- › Other HBase Table Features

#### Basic Data Access with the HBase API

- › Options to Access HBase Data
- › Creating and Deleting HBase Tables
- › Retrieving Data with Get
- › Retrieving Data with Scan
- › Inserting and Updating Data
- › Deleting Data

#### More Advanced HBase API

##### Features

- › Filtering Scans
- › Best Practices
- › HBase Coprocessors

#### HBase on the Cluster

- › How HBase Uses HDFS
- › Compactions and Splits

#### HBase Reads and Writes

- › How HBase Writes Data
- › How HBase Reads Data
- › Block Caches for Reading

#### HBase Performance Tuning

- › Column Family Considerations
- › Schema Design Considerations
- › Configuring for Caching
- › Dealing with Time Series and Sequential Data
- › Pre-Splitting Regions

#### HBase Administration and Cluster Management

- › HBase Daemons
- › ZooKeeper Considerations
- › HBase High Availability
- › Using the HBase Balancer
- › Fixing Tables with hbck
- › HBase Security

#### HBase Replication and Backup

- › HBase Replication
- › HBase Backup
- › MapReduce and HBase Clusters

#### Using Hive and Impala with HBase

- › Using Hive and Impala with HBase

#### Conclusion

#### Appendix A: Accessing Data with Python and Thrift

- › Thrift Usage
- › Working with Tables
- › Getting and Putting Data
- › Scanning Data
- › Deleting Data
- › Counters
- › Filters

#### Appendix B: OpenTSDB



# Bigdata References

**AKBANK**

**aselsan**

**ASSISTT**

**avea**

**Azercell**

**COMODO**  
Creating Trust Online®

**Garanti**

**HAVELSAN**

**(IBTECH)**

**ihs telekom**

**innova**

**KG TEKNOLOJİ HİZMETLERİ**

**NETAS**

**ORACLE®**

**Sabancı  
Üniversitesi**

**simternet**

**Sistek**

**STM**

**TTNET**

**TURKCELL**

**KUZEY KIBRIS  
TURKCELL**

**TURKCELL  
GLOBAL BİLGİ**

**TÜİK**  
TÜRKİYE İSTATİSTİK KURUMU

**Tüpraş**

**TÜRKİYE  
BANKASI**

**vodafone**

**YapıKredi**