

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 Developer Training for Spark and Hadoop

Learn how to import data into your Apache Hadoop cluster and process it with Spark, Hive, Flume, Sqoop, Impala, and other Hadoop ecosystem tools

This four-day hands-on training course delivers the key concepts and expertise developers need to develop high-performance parallel applications with Apache Spark 2. Participants will learn how to use Spark SQL to query structure data and Spark Streaming to perform real-time processing on streaming data from a variety of sources. Developers will also practice writing applications that use core Spark to perform ETL processing and iterative algorithms. The course covers how to work with large datasets stored in distributed file system, and execute Spark applications on a Hadoop cluster. After taking this course, participants will be prepared to face real-world challenges and build applications to execute faster decisions, better decisions and interactive analysis, applied to a wide variety of use cases, architectures and industries. With this course update, we streamlined the agenda to help you quickly become productive with the most important technologies including Spark 2.

Hands-On Hadoop

Hands-on exercises take place on a live cluster, running in the cloud. A private cluster will be built for each student to use during the class. Through instructor-led discussion and interactive, hands-on exercises, participants will learn Apache Spark and how it integrates with the entire Hadoop ecosystem, learning:

- › Distribute, store and process data in a Hadoop cluster
- › Write, configure, and deploy Apache Spark applications on a Hadoop cluster
- › Use the Spark shell for interactive data analysis
- › Process and query structured data using Spark SQL
- › Use Spark Streaming to process a live data stream

Audience and Prerequisites

This course is designed for developers and engineers who have programming experience. Apache Spark examples and hands-on exercises are presented in Scala and Python, so the ability to program in one of those languages is required. Basic familiarity with the Linux command line is assumed. Basic knowledge of SQL is helpful. Prior knowledge of Hadoop is not required.

CCA Spark & Hadoop Developer

This course is an excellent place to start for people working towards the CCA Spark & Hadoop Developer certification. It covers many of the subjects tested in the certification exam.





Cloudera Developer Training for Spark and Hadoop

Introduction to Hadoop and the Hadoop Ecosystem

- › Apache Hadoop Overview
- › Data Ingestion and Storage
- › Data Processing
- › Data Analysis and Exploration
- › Other Ecosystem Tools
- › Introduction to the Hands-on Exercises

Apache Hadoop File Storage

- › Apache Hadoop Cluster Components
- › HDFS Architecture
- › Using HDFS

Distributed Processing on an Apache Hadoop Cluster

- › YARN Architecture
- › Working with YARN

Apache Spark Basics

- › What is Apache Spark?
- › Starting the Spark Shell
- › Using the Spark Shell
- › Getting Started with Datasets and DataFrames
- › DataFrame Operations

Working with DataFrames and Schemas

- › Creating DataFrames from DataSources
- › Saving DataFrames to Data Sources
- › DataFrames Schemas
- › Eager and Lazy Execution

Analyzing Data with DataFrame Queries

- › Querying DataFrames Using Column Expressions
- › Grouping and Aggregation Queries
- › Joining DataFrames

RDD Overview

- › RDD Overview
- › RDD Data Sources
- › Creating and Saving RDDs
- › RDD Operations

Transforming Data with RDDs

- › Writing and Passing Transformation Functions
- › Transforming Execution
- › Converting Between RDDs and DataFrames

Aggregating Data with Pair RDDs

- › Key-Value Pair RDDs
- › Map-Reduce
- › Other Pair RDD Operations

Querying Tables and Views with Apache Spark SQL

- › Querying Tables in Spark Using SQL
- › Querying Files and Views
- › The Catalog API
- › Comparing Spark SQL, Apache Impala and Apache Hive-on-Spark

Working with Datasets in Scala

- › Datasets and DataFrames
- › Creating Datasets
- › Loading and Saving Datasets
- › Dataset Operations

Writing Configuring and Running Apache Spark Applications

- › Writing a Spark Application
- › Building and Running an Application
- › Application Deployment Mode
- › The Spark and Application Web UI
- › Configuring Application Properties

Distributed Processing

- › Review: Apache Spark on a Cluster
- › RDD Partitions
- › Example: Partitioning in Queries
- › Stages and Tasks
- › Job Execution Planning
- › Example: Catalyst Execution Plan
- › Example: RDD Execution Plan

Distributed Data Persistence

- › DataFrame and Dataset Persistence
- › Persistence Storage Levels
- › Viewing Persisted RDDs

Common Patterns in Apache Spark Data Processing

- › Common Apache Spark Use Cases
- › Iterative Algorithms in Apache Spark
- › Machine Learning
- › Example: k-means

Apache Spark Streaming: Introduction to DStreams

- › Apache Spark Streaming Overview
- › Example: Streaming Request Count
- › DStreams
- › Developing Streaming Applications

Apache Spark Streaming: Processing Multiple Batches

- › Multi-Batch Operations
- › Time Slicing
- › State Operations
- › Sliding Window Operations
- › Preview: Structured Streaming

Apache Spark Streaming: Data Sources

- › Streaming Data Sources Overview
- › Apache Flume and Apache Kafka Data Sources
- › Example: Using a Kafka Direct Data Source



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

**KUZAY KIBRIS
TURKCELL**

**TURKCELL
GLOBAL BİLGİ**

TÜİK
TÜRKİYE İSTATİSTİK KURUMU

Tüpraş

**TÜRKİYE
BANKASI**

vodafone

YapıKredi