

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 Data Analyst Training: Using Pig, Hive and Impala with Hadoop

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

This four-day data analyst training course focusing on Apache Pig and Hive and Cloudera Impala will teach you to apply traditional data analytics and business intelligence skills to big data. Cloudera presents the tools data professionals need to access, manipulate, transform, and analyze complex data sets using SQL and familiar scripting languages.

Advance Your Ecosystem Expertise

Apache Hive makes multi-structured data accessible to analysts, database administrators, and others without Java programming expertise. Apache Pig applies the fundamentals of familiar scripting languages to the Hadoop cluster. Cloudera Impala enables real-time interactive analysis of the data stored in Hadoop via a native SQL environment.

Hands-On Hadoop

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

- › The features that Pig, Hive, and Impala offer for data acquisition, storage, and analysis
- › The fundamentals of Apache Hadoop and data ETL (extract, transform, load), ingestion, and processing with Hadoop tools
- › How Pig, Hive, and Impala improve productivity for typical analysis tasks
- › Joining diverse datasets to gain valuable business insight
- › Performing real-time, complex queries on datasets

Audience & Prerequisites

This course is designed for data analysts, business intelligence specialists, developers, system architects, and database administrators. Knowledge of SQL is assumed, as is basic Linux command-line familiarity. Knowledge of at least one scripting language (e.g., Bash scripting, Perl, Python, Ruby) would be helpful but is not essential. Prior knowledge of Apache Hadoop is not required.





Cloudera Data Analyst Training: Using Pig, Hive, and Impala with Hadoop

Introduction

Hadoop Fundamentals

- › The Motivation for Hadoop
- › Hadoop Overview
- › Data Storage: HDFS
- › Distributed Data Processing: YARN, MapReduce, and Spark
- › Data Processing and Analysis: Pig, Hive, and Impala
- › Data Integration: Sqoop
- › Other Hadoop Data Tools
- › Exercise Scenarios Explanation

Introduction to Pig

- › What Is Pig?
- › Pig's Features
- › Pig Use Cases
- › Interacting with Pig

Basic Data Analysis with Pig

- › Pig Latin Syntax
- › Loading Data
- › Simple Data Types
- › Field Definitions
- › Data Output
- › Viewing the Schema
- › Filtering and Sorting Data
- › Commonly-Used Functions

Processing Complex Data with Pig

- › Storage Formats
- › Complex/Nested Data Types
- › Grouping
- › Built-In Functions for Complex Data
- › Iterating Grouped Data

Multi-Dataset Operations with Pig

- › Techniques for Combining Data Sets
- › Joining Data Sets in Pig
- › Set Operations
- › Splitting Data Sets

Pig Troubleshooting and Optimization

- › Troubleshooting Pig
- › Logging
- › Using Hadoop's Web UI
- › Data Sampling and Debugging
- › Performance Overview
- › Understanding the Execution Plan
- › Tips for Improving the Performance of Your Pig Jobs

Introduction to Hive and Impala

- › What Is Hive?
- › What Is Impala?
- › Schema and Data Storage
- › Comparing Hive to Traditional Databases
- › Hive Use Cases

Querying with Hive and Impala

- › Databases and Tables
- › Basic Hive and Impala Query Language Syntax
- › Data Types
- › Differences Between Hive and Impala Query Syntax
- › Using Hue to Execute Queries
- › Using the Impala Shell

Data Management

- › Data Storage
- › Creating Databases and Tables
- › Loading Data
- › Altering Databases and Tables
- › Simplifying Queries with Views
- › Storing Query Results

Data Storage and Performance

- › Partitioning Tables
- › Choosing a File Format
- › Managing Metadata
- › Controlling Access to Data

Relational Data Analysis with Hive and Impala

- › Joining Datasets
- › Common Built-In Functions
- › Aggregation and Windowing

Working with Impala

- › How Impala Executes Queries
- › Extending Impala with User-Defined Functions
- › Improving Impala Performance

Analyzing Text and Complex Data with Hive

- › Complex Values in Hive
- › Using Regular Expressions in Hive
- › Sentiment Analysis and N-Grams
- › Conclusion

Hive Optimization

- › Understanding Query Performance
- › Controlling Job Execution Plan
- › Bucketing
- › Indexing Data

Extending Hive

- › SerDes
- › Data Transformation with Custom Scripts
- › User-Defined Functions
- › Parameterized Queries

Choosing the Best Tool for the Job

- › Comparing MapReduce, Pig, Hive, Impala, and Relational Databases
- › Which to Choose?

Conclusion



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