Training: Cloudera
# ARCH-492 Architecting Cloudera Edge to AI

CLOUDERA Training Partner

## TRAINING GOALS:

Designing Edge to AI Applications is a 4-day learning event that addresses advanced big data architecture topics for building edge to AI applications to cover streaming, operational data processing, analytics, and machine learning.  The workshop brings together technical contributors into a group setting to design and architect solutions to a challenging business problem. The workshop addresses big data architecture problems in general, and then applies them to the design of a challenging system.

Throughout the highly interactive workshop, participants apply concepts to real-world examples resulting in detailed synergistic discussions. The workshop is conducive for participants to learn techniques for architecting big data systems, not only from Cloudera's experience but also from the experiences of fellow participants.

More specifically, this workshop addresses advanced big data architecture topics, including, data formats, transformation, transactions, real-time, batch and machine learning processing, scalability, fault tolerance, security, and privacy, minimizing the risk of an unsound architecture and technology selection.

What you'll learn

- Cloudera Data Platform
- Big Data Architecture
- Building Scalable applications
- Building Fault Tolerant Solutions
- Security and Privacy Deployment on Public, Private, and Hybrid Cloud

What to expect

Participants should mainly be architects, developer team leads, big data developers, data engineers, senior analysts, dev ops admins and machine learning developers who are working on big data or streaming applications and have an interest in how to design and develop such applications on CDP. To gain the most from the workshop, participants should have working knowledge of popular Big Data and streaming technologies such as HDFS, Spark, Kafka, Hive/Impala, Data Formats, and relational database management systems. Detailed API level knowledge is not needed, as there will not be any programming activities and instead the focus will be on architecture design.



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

## CONSPECT:

- Introduction
  - Team activity: Team Introductions
- Technology Review
  - HDFS
  - HBase
  - Kudu
  - Map Reduce
  - Spark, including SparkSQL and SparkML
  - Hive
  - Impala
  - Relational Database Management Systems
  - Spark streaming
  - Apache Flume
  - Apache NiFi
  - Apache Kafka
- Workshop Application Use Cases
  - Oz Metropolitan
  - Architectural questions
  - Team activity: Review Metroz Use Cases and Logical Architecture
- Application Vertical Slice
  - Definition
  - Minimizing risk of an unsound architecture
  - Selecting a vertical slice
  - Team activity: Metroz Vertical Slice
- Application Processing
  - Real time, near real time processing
  - Batch processing
  - Data access patterns
  - Delivery and processing guarantees
  - Data consistency and ACID transactions
  - Stream processing guarantees
  - Machine Learning pipelines
  - Team activity: Metroz Processing
- Application Data



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- Three V's of Big Data
- Data Lifecycle
- Data Formats
- Transforming Data
- Team activity: Metroz Data Requirements

- Scalable Applications
  - Scale up, scale out, scale to X
  - Determining if an application will scale
  - Poll: scalable airport terminal designs
  - Spark scalability and parallel processing
  - Scalable storage engines: HDFS, Ozone, Kafka and Kudu
  - Team activity: Scaling Metroz
- Fault-Tolerant Distributed Systems
  - Principles
  - Transparency
  - Hardware vs. Software redundancy
  - Tolerating disasters
  - Stateless functional fault tolerance
  - Stateful fault tolerance
  - Replication and group consistency
  - Application tolerance for failures
  - Team activity: Failures in Metroz
- Security and Privacy
  - Principles
  - Security Architecture
  - Knox Security Architecture
  - Ranger Security Architecture
  - Setting security policies with Ranger
  - Threat Analysis
  - Team activity: Securing Metroz
- Deployment
  - Cluster sizing and evolution
  - On-premise vs. Cloud
  - Edge computing
  - Team activity: Deploying Metroz
- Software Architecture



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- Architecture artifacts
- Team activity: Metroz Physical Architecture

- Potential CDP Solutions
  - Review of Uber and Lyft Big data platforms
  - Review of Metroz CDP solution architectures
- Wrap Up

## REQUIREMENTS:

Participants should mainly be architects, developer team leads, big data developers, data engineers, senior analysts, dev ops admins and machine learning developers who are working on big data or streaming applications and have an interest in how to design and develop such applications on CDP. To gain the most from the workshop, participants should have working knowledge of popular Big Data and streaming technologies such as HDFS, Spark, Kafka, Hive/Impala, Data Formats, and relational database management systems. Detailed API level knowledge is not needed, as there will not be any programming activities and instead the focus will be on architecture design.

## Difficulty level



## CERTIFICATE:

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

Upon completion of the course, attendees are encouraged to continue their study and register for the Cloudera Certified Administrator (CCA) exam
https://www.cloudera.com/about/training/certification/cdhhdp-certification/cca-admin.html

Certification is a great differentiator. It helps establish you as a leader in the field, providing employers and customers with tangible evidence of your skills and expertise.

## TRAINER:

Certified Cloudera Instructor



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007