



Big Data Analytics

MOOC course presentation

under guidance of Prof. Amit Ganatra

Hadoop Platform and Application Framework

Platform : [Coursera](#)

Offered by: UCSanDeigo

Authors: Natasha Balac, Paul Rodriguez, Andrea Zonca

Time taken to complete course - 26 hours

Subdivided into 5 weeks

Total Quizzes: 11 (Achieved 94% overall)

Coding Assignments: 4

Week-1

Hadoop Basics

What is Hadoop?

Big Data Hadoop Stack

Hadoop Stack Basics

Hadoop Distributed File System

Hadoop 'Zoo'

Hadoop Ecosystem Major Components

Hands on Exploration of the Cloudera VM

Week-2

Hadoop Stack

More on Hadoop Stack

Hadoop Distributed File System

MapReduce Framework and Yarn

Hadoop Execution Environment

Overview of Tez and Spark

Hadoop Based Application and Services

Intro to Hbase

Intro to Hive

Intro to Pig

Commands used for various application

Week-3

Hadoop Distributed

File System

HDFS Architecture and Configuration
Performance Envelop
Read/Write Process in HDFS
HDFS Performance and Tuning
Tuning Parameters
Performance and Robustness
HDFS Access and Commands
HDFS API and Application

Week-4 Map/Reduce

Introduction to Map/Reduce

Map/Reduce Framework

Programming Assignment

Running WordCount with Hadoop Streaming (.py)

Example and Principle of Map/Reduce

Joining Data

Vector Multiplication

Computational Cost

Programming Assignment

Joining Data with Hadoop Streaming (.py code)

Week-5

Apache Spark

- Introduction to Apache Spark
- Architecture of Spark
- Setup of PySpark on Cloudera VM
- Resilient Distributed Dataset and Transformations
- Actions in Spark
- Memory caching in Spark
- Broadcast variables and Accumulators
- Programming Assignment
- Advance Join in Spark

"Intelligence plus character that is the goal of the true education"

-Dr. Martin Luther King Jr.



Apr 20, 2022

Purvam Tejendrakumar Sheth

has successfully completed

Hadoop Platform and Application Framework

an online non-credit course authorized by University of California San Diego and offered through Coursera

Natasha Balac, Paul Rodriguez, Andrea Zonca, Mahidhar Tatineni
San Diego Supercomputer Center

COURSE CERTIFICATE



Verify at:
<https://coursera.org/verify/6NKAW42VNPGQ>

Coursera has confirmed the identity of this individual and their participation in the course.